



LLM Reasoners

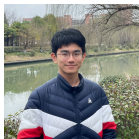
New Evaluation, Library and Analysis
of step-by-step reasoning with LLMs



Shibo
Hao*



Yi Gu*



Haotian
Luo*



Tianyang
Liu



Xiyan
Shan



Xinyuan
Wang



Shuhua
Xie



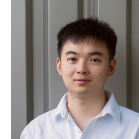
Haodi
Ma



Adithya
Samavedhi



Zhen
Wang



Zhiting
Hu

Large Language Model Reasoning

≡ Google The Keyword

In this story



Research

GPT-4 surpasses
advan

Our new benchmark approach to MMLU enables Gemini to use its reasoning capabilities to think more carefully before answering difficult questions, leading to significant improvements over just using its first impression.

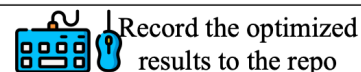
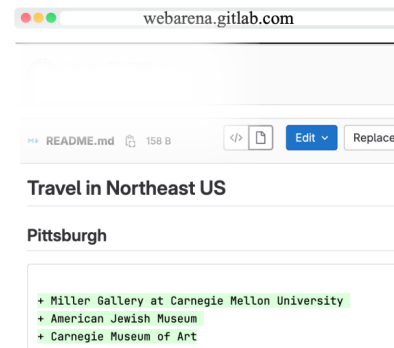
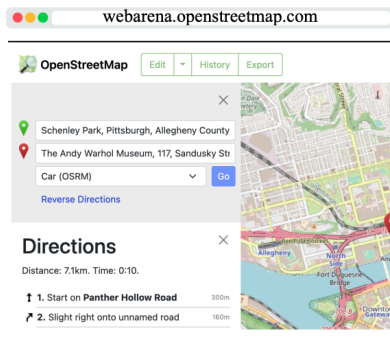
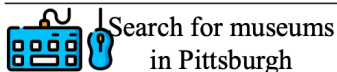
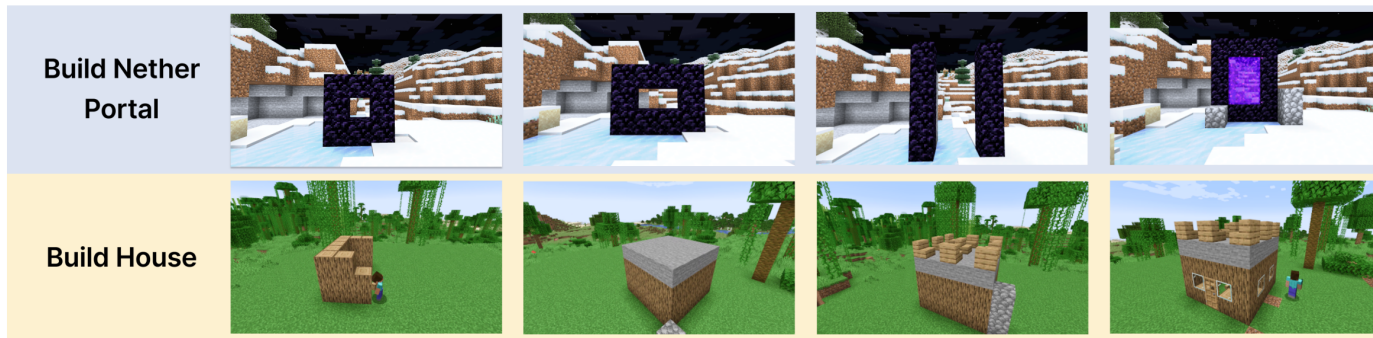


Anthropic
@AnthropicAI

Today, we're announcing Claude 3, our next generation of AI models.

The three state-of-the-art models—Claude 3 Opus, Claude 3 Sonnet, and Claude 3 Haiku—set new industry benchmarks across reasoning, math, coding, multilingual understanding, and vision.

Large Language Model Reasoning



Outline

Reasoning with LLMs:

- Algorithms
- Library
- Evaluation
- Analysis

Outline

Reasoning with LLMs:

- Algorithms
- Library
- Evaluation
- Analysis

Large Language Model Reasoning

Model Input

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

Model Output

A: The answer is 27.

Large Language Model **Step-by-step** Reasoning

Standard Prompting

Model Input

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

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

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?

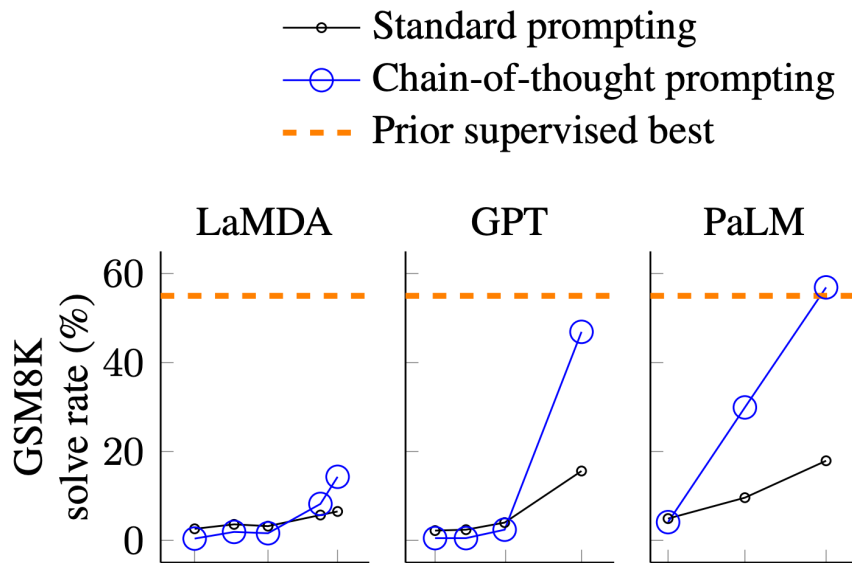
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

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

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

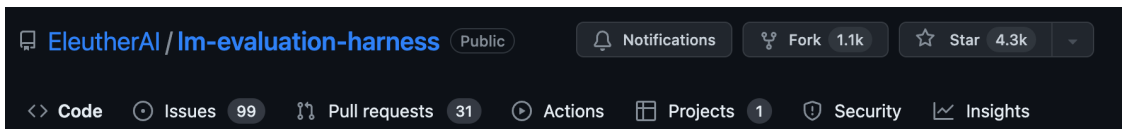
Large Language Model **Step-by-step** Reasoning



Large Language Model **Step-by-step** Reasoning

Becomes the **Default Choice**

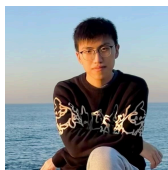
 **Open LLM Leaderboard**



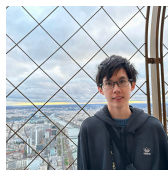
	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku
Undergraduate level knowledge <i>MMLU</i>	86.8% 5 shot	79.0% 5-shot	75.2% 5-shot
Graduate level reasoning <i>GPQA, Diamond</i>	50.4% 0-shot CoT	40.4% 0-shot CoT	33.3% 0-shot CoT
Grade school math <i>GSM8K</i>	95.0% 0-shot CoT	92.3% 0-shot CoT	88.9% 0-shot CoT
Math problem-solving <i>MATH</i>	60.1% 0-shot CoT	43.1% 0-shot CoT	38.9% 0-shot CoT

Can we design algorithms to generate better reasoning chains with LLMs?

Reasoning with Language Model is Planning with World Model



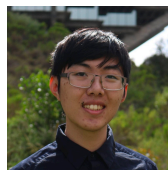
Shibo Hao*



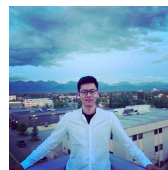
Yi Gu*



Haodi Ma



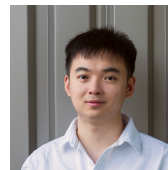
Joshua Hong



Zhen Wang



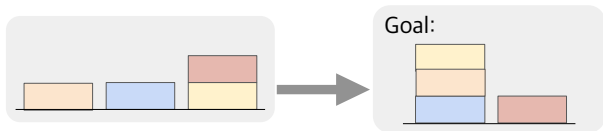
Daisy Wang



Zhiting Hu

Chain-of-thoughts vs Human reasoning

Blocksworld: How to move the blocks to the goal state?



A: Chain-of-Thoughts Prompting (CoT) with LLM

- Autoregressive decoding

System 1

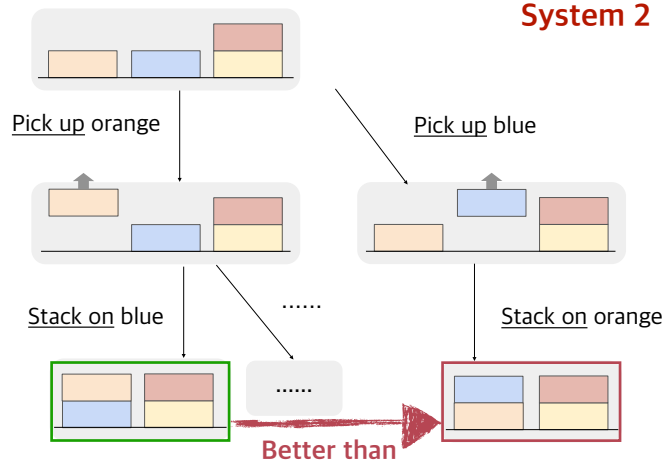
Invalid Action!
The yellow block is still
under the red one.

1. Pick up the orange block.
2. Stack it on the blue block.
3. Pick up the yellow block. ✗
4. Stack it on the orange block.
5. Pick up the red block.
6. Put it on the table.

B: Human Reasoning

- Internal **world model** to track **states**
- **Explore** alternative reasoning paths
- **Assess outcomes** by looking ahead

System 2



On the planning abilities of large language models (a critical investigation with a proposed benchmark) [Valmeekam et al, 2023]

Chain-of-thought prompting elicits reasoning in large language models [Wei et al., 2022]

Mental models: Towards a cognitive science of language, inference, and consciousness [Johnson-Laird, 1983]

From System 1 Deep Learning to System 2 Deep Learning [Bengio, 2019]

Reasoning-via-Planning (RAP 🎵)

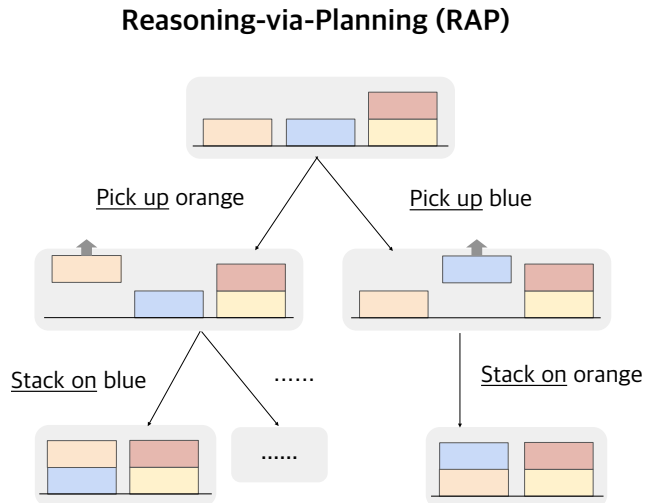
Human Reasoning

- Internal **world model** to track **states**
- **Explore** alternative reasoning paths
- **Assess outcomes** by looking ahead

How to enable LLMs to reason close to humans?

Reasoning-via-Planning: RAP 🎵

- Repurpose LLM as **world model**
- Principled **planning** algorithm
- **Rewards** to estimate outcomes



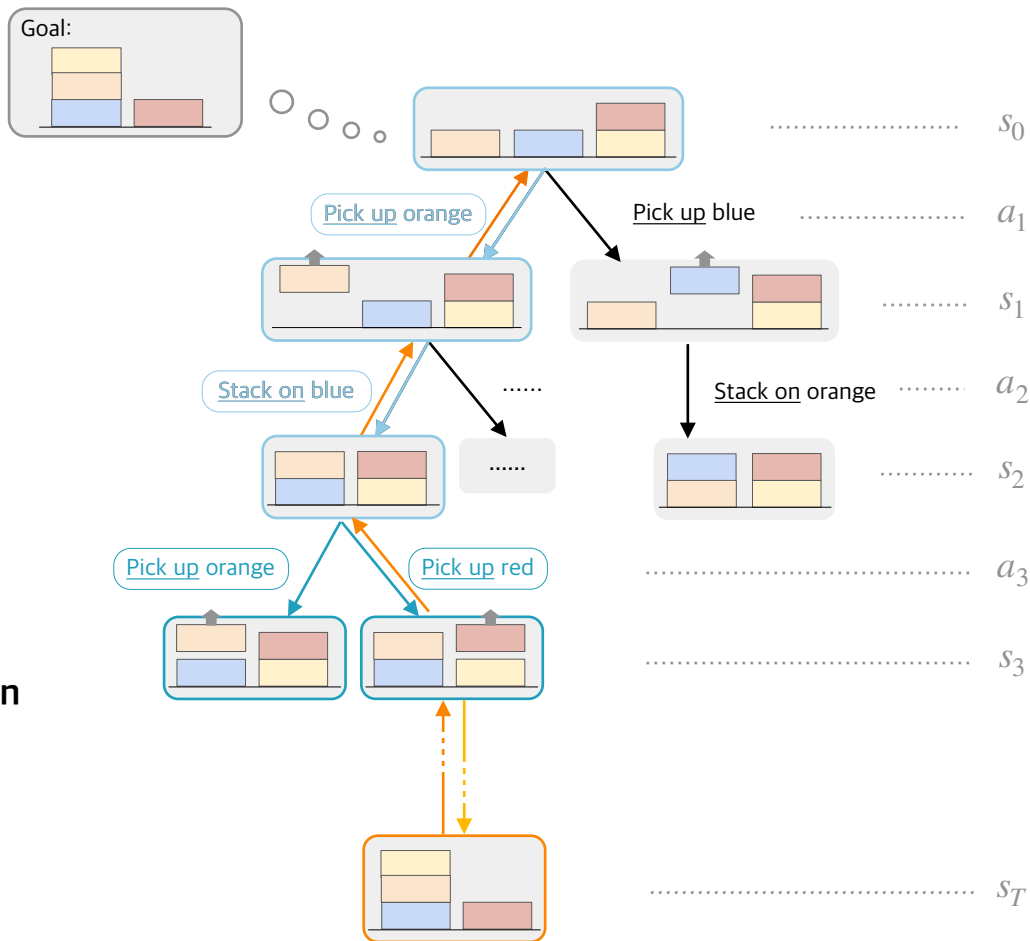
Planning Algorithm

Monte Carlo Tree Search (MCTS):

Iteratively build reasoning tree

1. Selection
2. Expansion
3. Simulation
4. Back-propagation

Balanced exploration and exploitation



Rewards in RAP

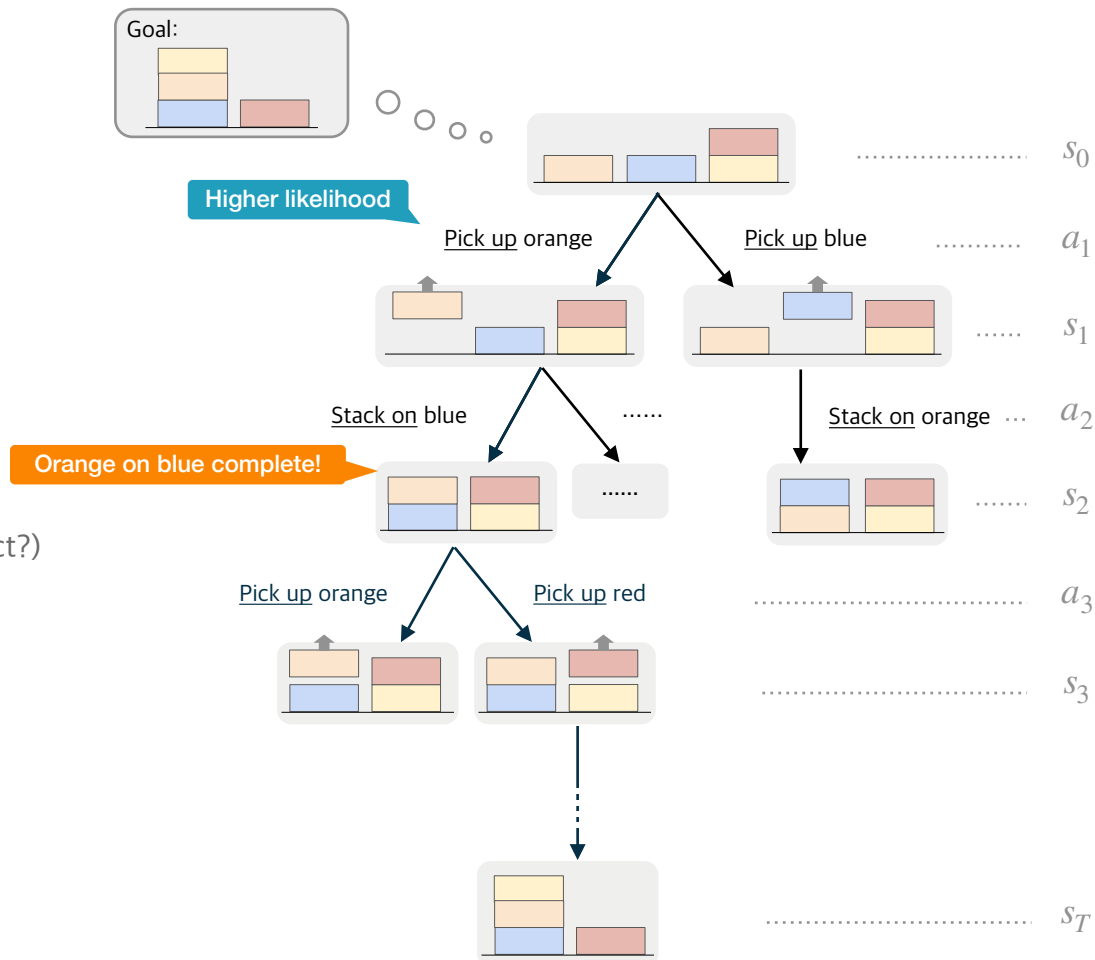
Reward design is **flexible**

In Blocksworld:

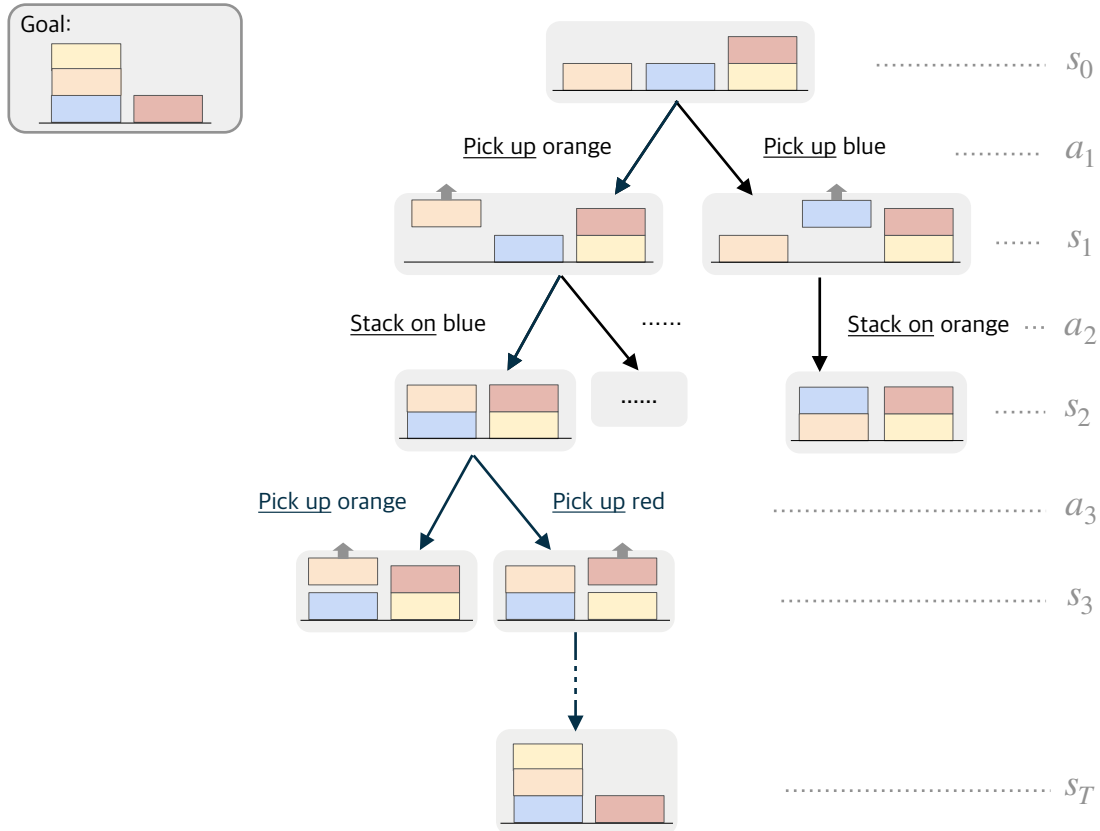
- **Likelihood** of actions
- **Task-heuristic** (# of subgoals)

Other possible rewards:

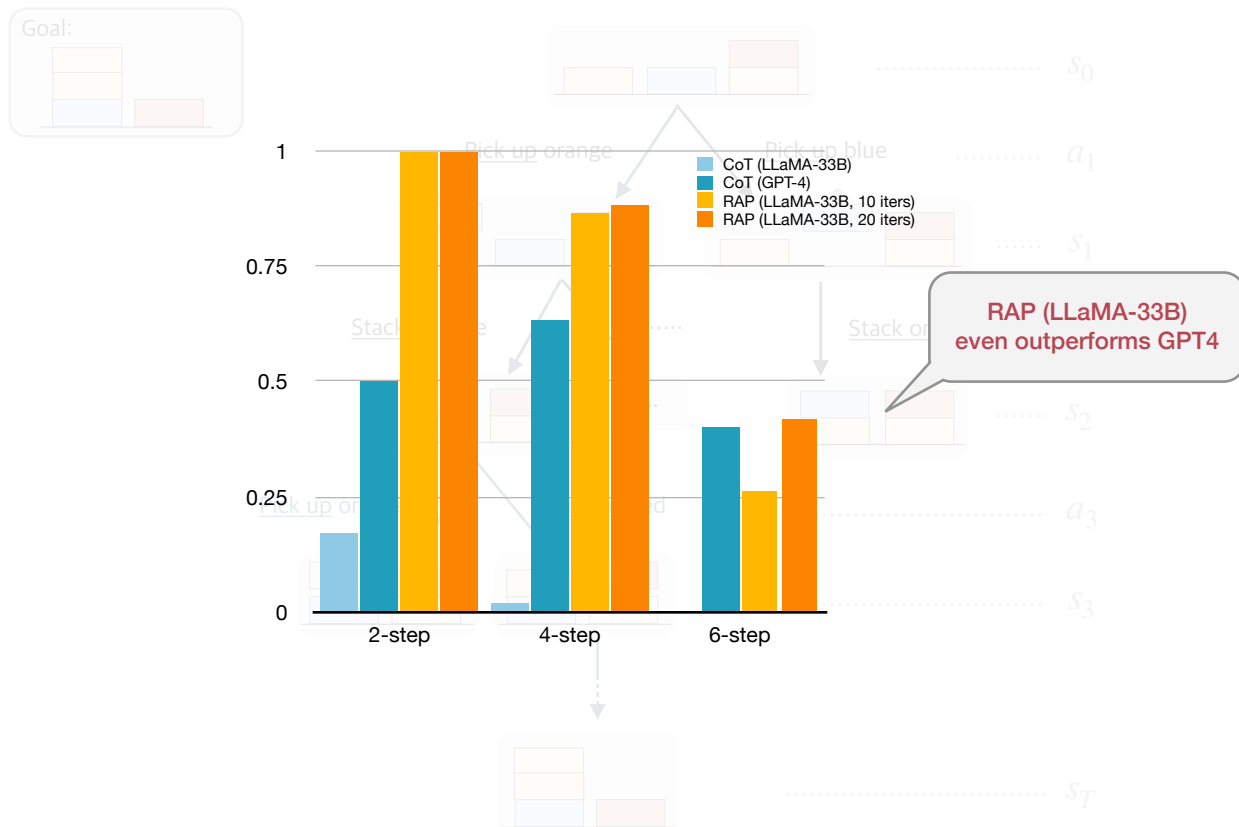
- **Self-evaluation** by LLM (e.g. useful? correct?)
- **Confidence** of next state
-



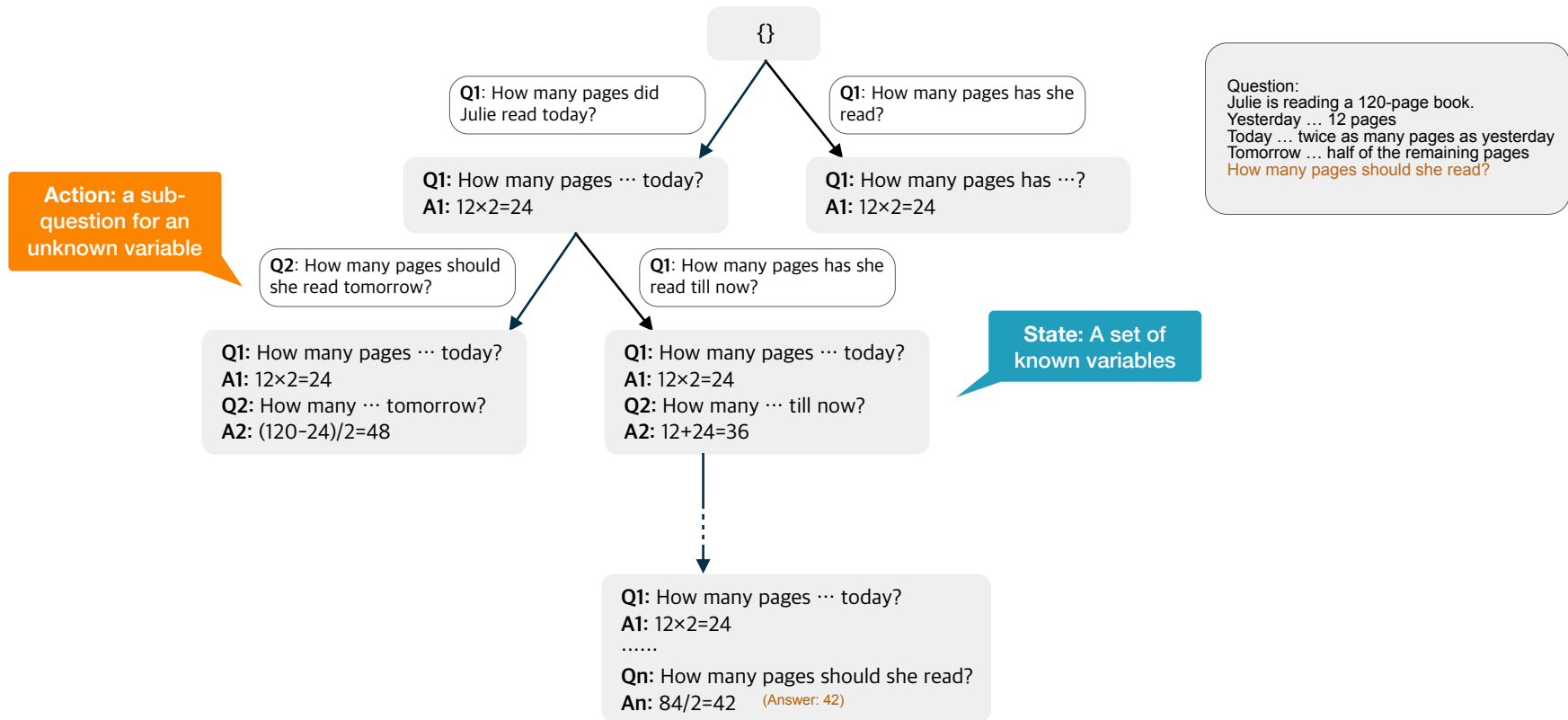
RAP on Plan Generation (Blocksworld)



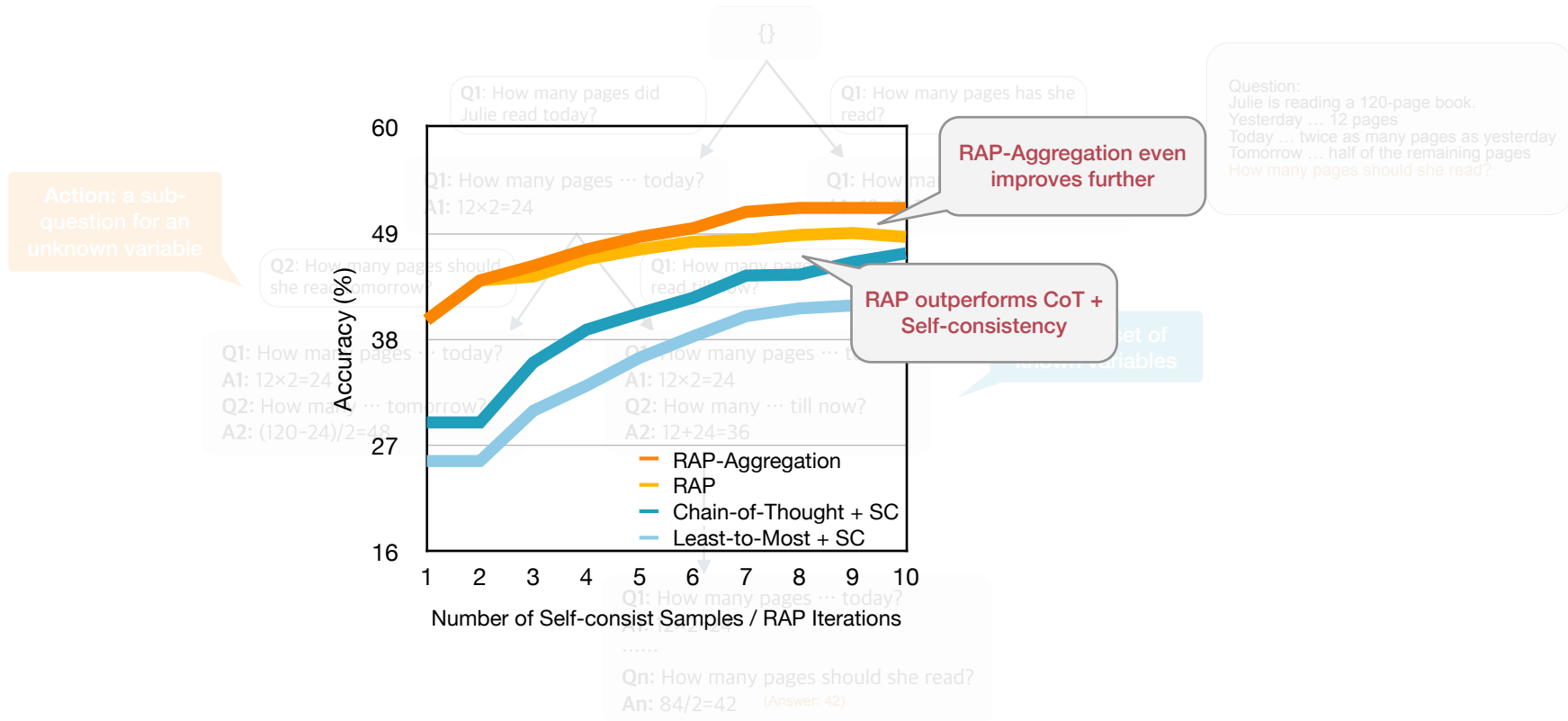
RAP on Plan Generation (Blocksworld)



RAP on Mathematical Reasoning (GSM8k)

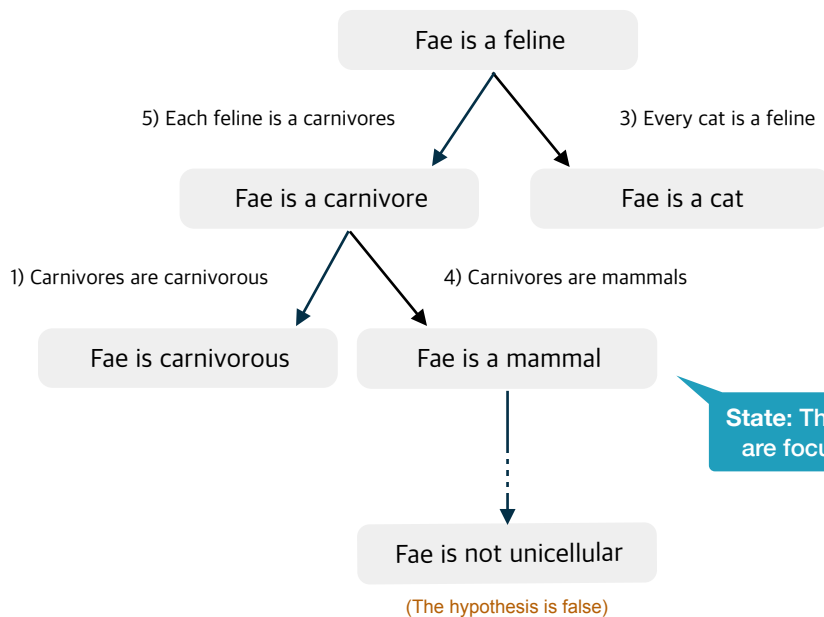


RAP on Mathematical Reasoning (GSM8k)



RAP on Logical Reasoning (PrOntoQA)

Action: selecting
a rule from the
rule set



Rules:

- (1) Carnivores are carnivorous
- (2) Animals are not unicellular
- (3) Every cat is a feline
- (4) ...

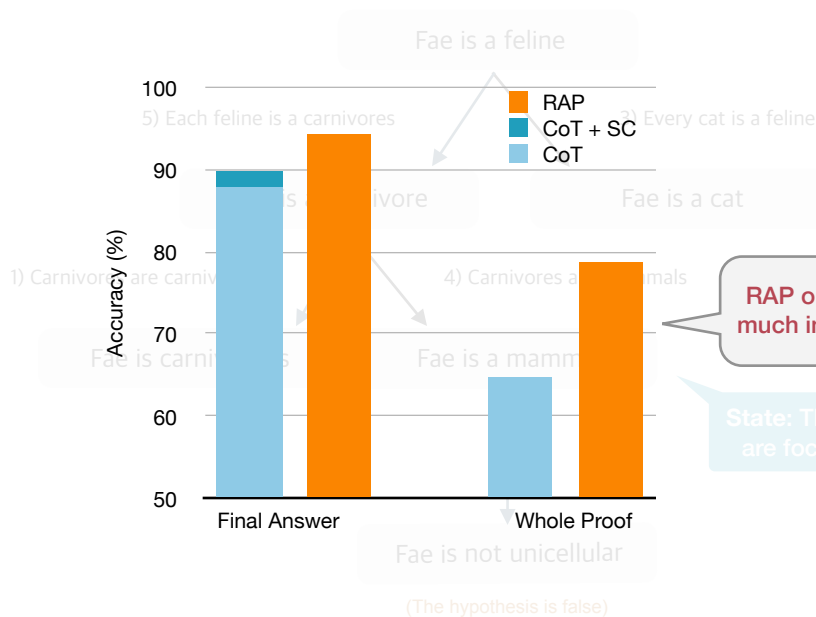
Fact: Fae is a feline

Hypothesis: Fae is unicellular?

State: The fact we
are focusing on

RAP on Logical Reasoning (PrOntoQA)

Action: selecting a rule from the rule set



Rules:

- (1) Carnivores are carnivorous
- (2) Animals are not unicellular
- (3) Every cat is a feline
- (4) ...

Fact: Fae is a feline

Hypothesis: Fae is unicellular?

RAP outperforms CoT much in proof accuracy

State: The fact we are focusing on

Large Language Model Step-by-step Reasoning

Q1: Different formulations and implementations?

Solving Math Word Problems via Cooperative Reasoning induced Language Models

Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

Jason Wei Xuezhi Wang Dale Schuurmans Maarten Bosker
Brian Ichter Fei Xia Ed H. Chi Quoc V. Le Denny Zhou
Google Research, Brain Team
{jasonwei, dennyzhou}@google.com

Xinyu Zhu^{◇*} Junjie Wang^{♣*} Lin Zhang[♡] Yuxiang Zhai
Ruyi Gan[♡] Jiaying Zhang[♡] Yujiu Yang^{◇†}
[◇]Tsinghua University [♣]Waseda University
[♡]International Digital Economy Academy
zhuxy21@emails.tsinghua.edu.cn yang.yujiu@sz.tsinghua.edu.cn
wj1020181822@toki.waseda.jp joe10495@asagi.waseda.jp
{zhanglin, ganrui, zhangjiaying}@idea.edu.cn

Tree of Thoughts: Deliberate Problem Solving with Large Language Models

Shunyu Yao Dian Yu Jeffrey Zhao Izhak Shafran
Princeton University Google DeepMind Google DeepMind Google DeepMind
Thomas L. Griffiths Yuan Cao Karthik Narasimhan
Princeton University Google DeepMind Princeton University

Reasoning with Language Model is Planning with World Model

Shibo Hao^{*♣} Yi Gu^{*♣} Haodi Ma[◇] Joshua Jiahua Hong[♣]
Zhen Wang^{*♣} Daisy Zhe Wang[◇] Zhiting Hu[♣]
[♣]UC San Diego, [◇]University of Florida
[♣]Mohamed bin Zayed University of Artificial Intelligence
{s5hao, yig025, jjhong, zhw085, zhh019}@ucsd.edu
{ma.haodi, daisiyw}@ufl.edu

GRACE: Discriminator-Guided Chain-of-T

Muhammad Khalifa^{*}, Lajanugen Logeswaran[†], Moontae Lee^{††},
Honglak Lee^{*†}, Lu Wang^{*}

University of Michigan^{*}, LG AI Research[†], University

TOOLCHAIN*: EFFICIENT ACTION SPACE NAVIGATION IN LARGE LANGUAGE MODELS WITH A* SEARCH

Yuchen Zhuang^{1*}, Xiang Chen², Tong Yu², Saayan Mitra²
Victor Bursztyn², Ryan A. Rossi², Somdeb Sarkhel², Chao Zhang¹
Georgia Institute of Technology¹ Adobe Research²
yczhuang@gatech.edu, {xiangche, tyu, smitra}@adobe.com
{soaresbu, ryrossi, sarkhel}@adobe.com, chaozhang@gatech.edu

AlphaZero-Like Tree-Search can Guide Large Language Model Decoding and Training

Xidong Feng^{*1} Ziyu Wan^{*2} Muning Wen² Stephen Marcus McAleer³
Ying Wen² Weinan Zhang² Jun Wang¹

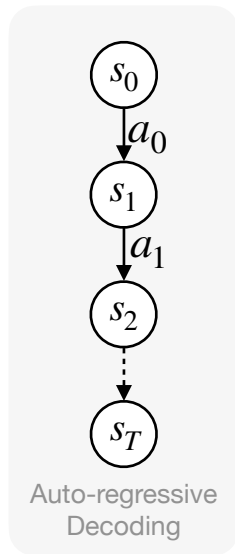
Outline

Reasoning with LLMs:

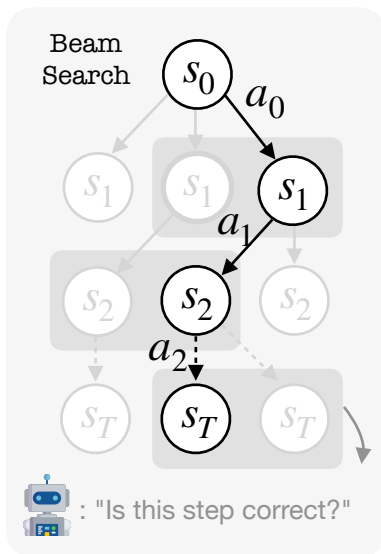
- Algorithms
- **Library**
- Evaluation
- Analysis

Step-by-step reasoning algorithms

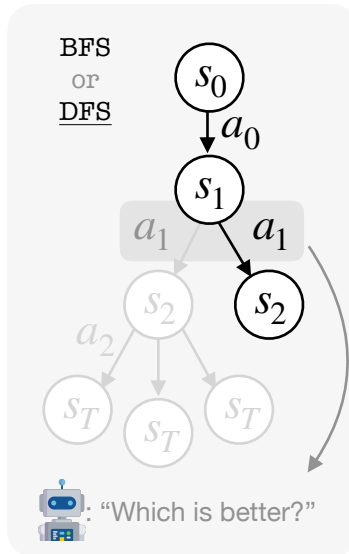
Chain-of-Thoughts
[Wei et al., 2022]



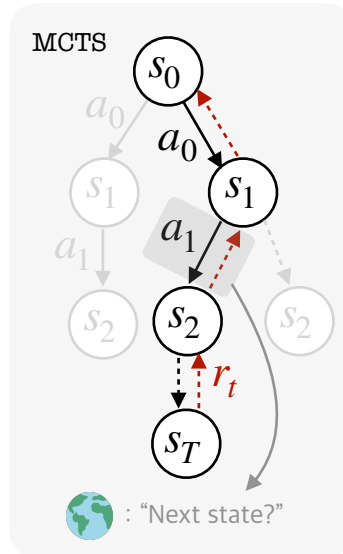
Self-eval Beam Search
[Xie et al., 2023]



Tree-of-thoughts
[Yao et al., 2023]



Reasoning-via-planning
[Hao et al., 2023]



Unified formulation of reasoning algorithms

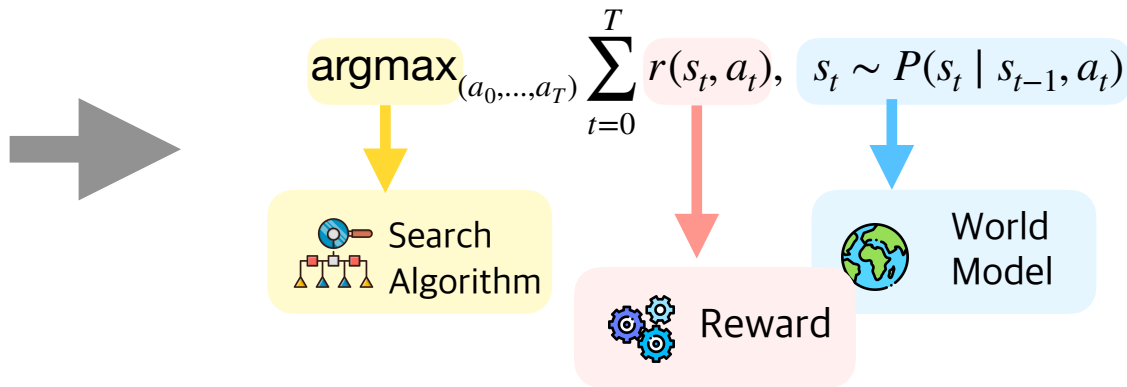
Chain-of-Thoughts
[Wei et al., 2022]

Self-eval Beam Search
[Xie et al., 2023]

Tree-of-thoughts
[Yao et al., 2023]

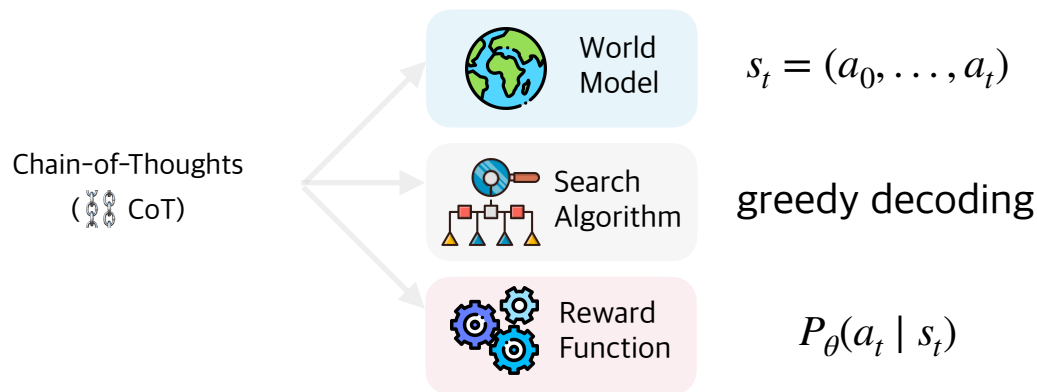
Reasoning-via-planning
[Hao et al., 2023]

...



A Formulation of Step-by-step Reasoning

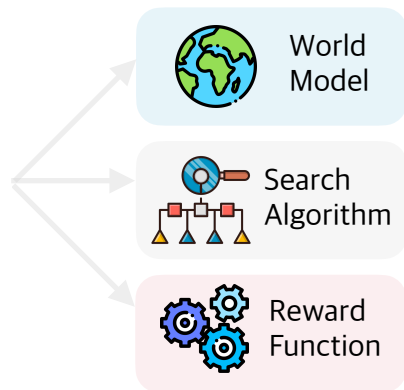
$$\operatorname{argmax}_{(a_0, \dots, a_T)} \sum_{t=0}^T r(s_t, a_t), \quad s_t \sim P(s_t \mid s_{t-1}, a_t)$$



A Formulation of Step-by-step Reasoning

$$\operatorname{argmax}_{(a_0, \dots, a_T)} \sum_{t=0}^T r(s_t, a_t), \quad s_t \sim P(s_t \mid s_{t-1}, a_t)$$

Chain-of-Thoughts
(🧠 CoT)



$$s_t = (a_0, \dots, a_t)$$

greedy decoding

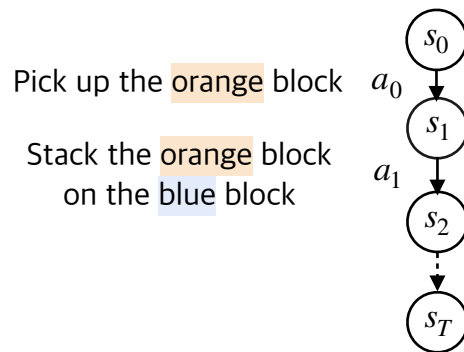
$$P_{\theta}(a_t \mid s_t)$$

Task:



Manipulates the blocks such that:

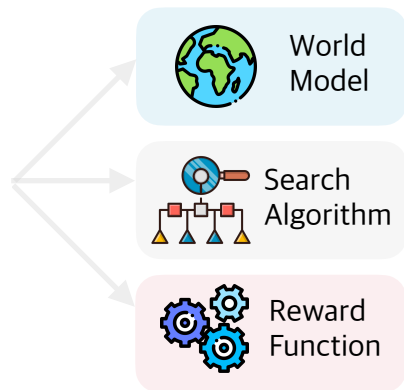
- Orange block on the blue block;
- Yellow block is on the orange block.



A Formulation of Step-by-step Reasoning

$$\operatorname{argmax}_{(a_0, \dots, a_T)} \sum_{t=0}^T r(s_t, a_t), \quad s_t \sim P(s_t \mid s_{t-1}, a_t)$$

Chain-of-Thoughts
(🧠 CoT)

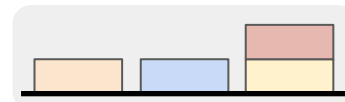


$$s_t = (a_0, \dots, a_t)$$

greedy decoding

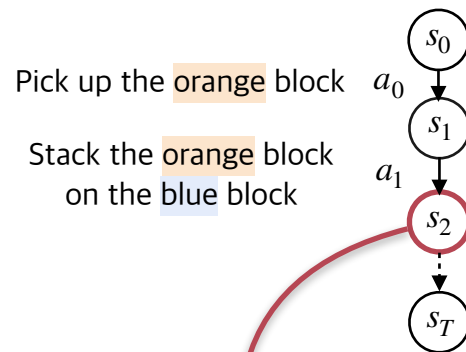
$$P_{\theta}(a_t \mid s_t)$$

Task:



Manipulates the blocks such that:

- Orange block on the blue block;
- Yellow block is on the orange block.

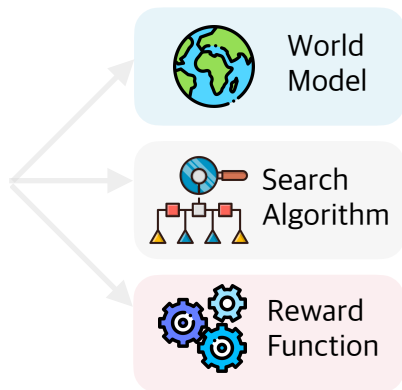


(Pick up the orange block, Stack the orange block on the blue block)

A Formulation of Step-by-step Reasoning

$$\operatorname{argmax}_{(a_0, \dots, a_T)} \sum_{t=0}^T r(s_t, a_t), \quad s_t \sim P(s_t \mid s_{t-1}, a_t)$$

Tree-of-Thoughts
( ToT)



$$s_t = (a_0, \dots, a_t)$$

BFS / DFS

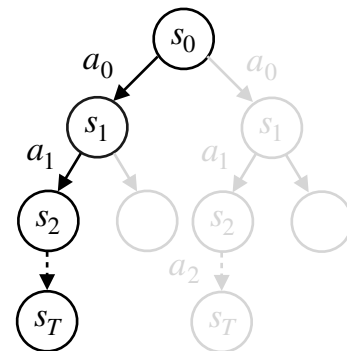
$$P_{\theta}(\text{"good"} \mid s_t, a_t)$$

Task:



Manipulates the blocks such that:

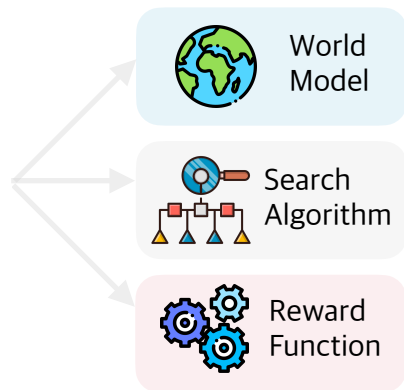
- Orange block on the blue block;
- Yellow block is on the orange block.



A Formulation of Step-by-step Reasoning

$$\operatorname{argmax}_{(a_0, \dots, a_T)} \sum_{t=0}^T r(s_t, a_t), \quad s_t \sim P(s_t \mid s_{t-1}, a_t)$$

Tree-of-Thoughts
(🌲 ToT)



$$s_t = (a_0, \dots, a_t)$$

BFS / DFS

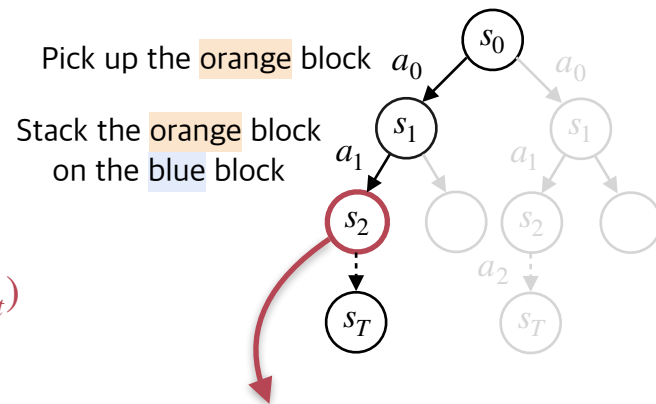
$$P_{\theta}(\text{"good"} \mid s_t, a_t)$$

Task:



Manipulates the blocks such that:

- Orange block on the blue block;
- Yellow block is on the orange block.

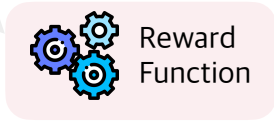
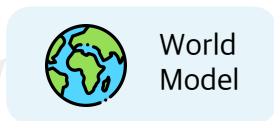


(Pick up the orange block, Stack the orange block on the blue block)

A Formulation of Step-by-step Reasoning

$$\operatorname{argmax}_{(a_0, \dots, a_T)} \sum_{t=0}^T r(s_t, a_t), \quad s_t \sim P(s_t \mid s_{t-1}, a_t)$$

Reasoning-via-
Planning
(RAP)



$$s_t \sim P_\theta(s_t \mid s_{t-1}, a_{t-1})$$

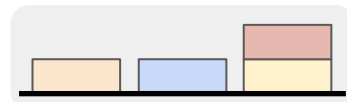
MCTS

$$P_\theta(\text{"good"} \mid s_t, a_t)$$

$$P_\theta(a_t \mid s_t)$$

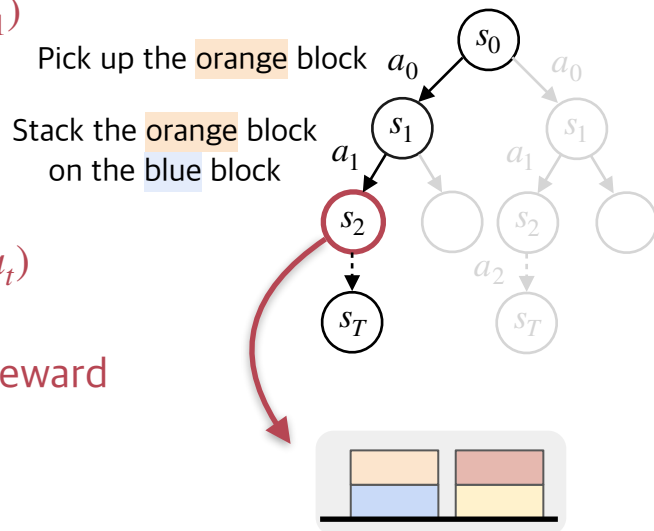
Other task-specific reward

Task:

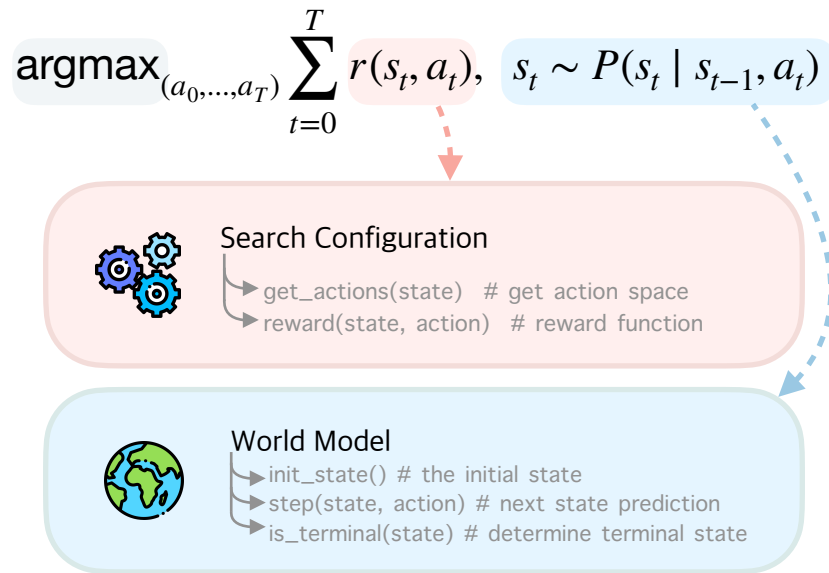


Manipulates the blocks such that:

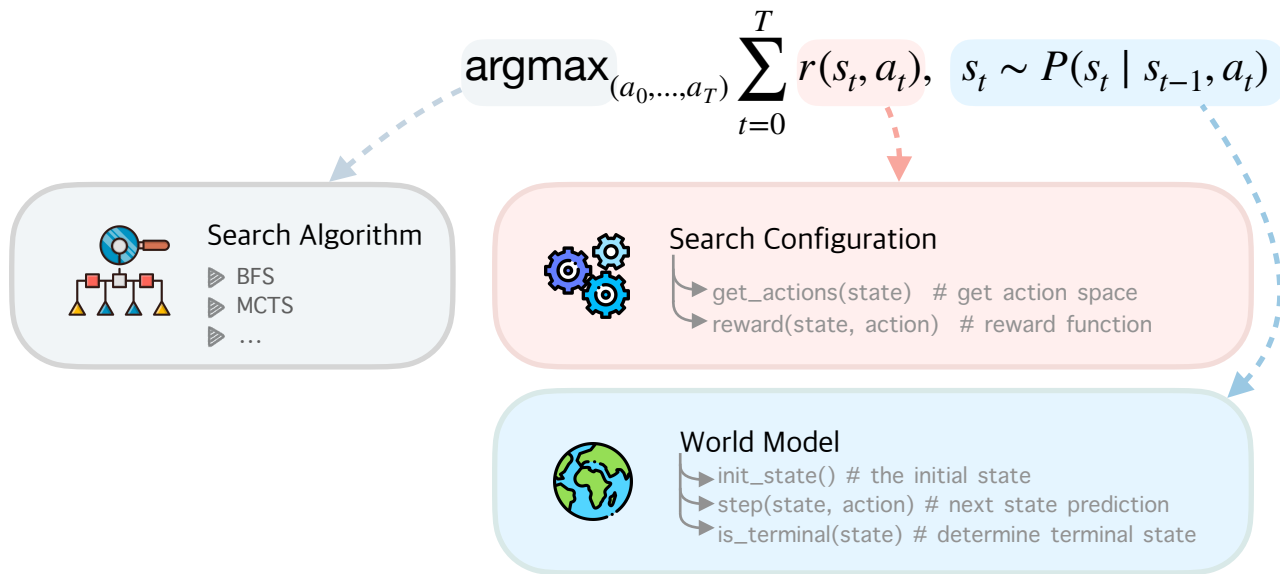
- Orange block on the blue block;
- Yellow block is on the orange block.



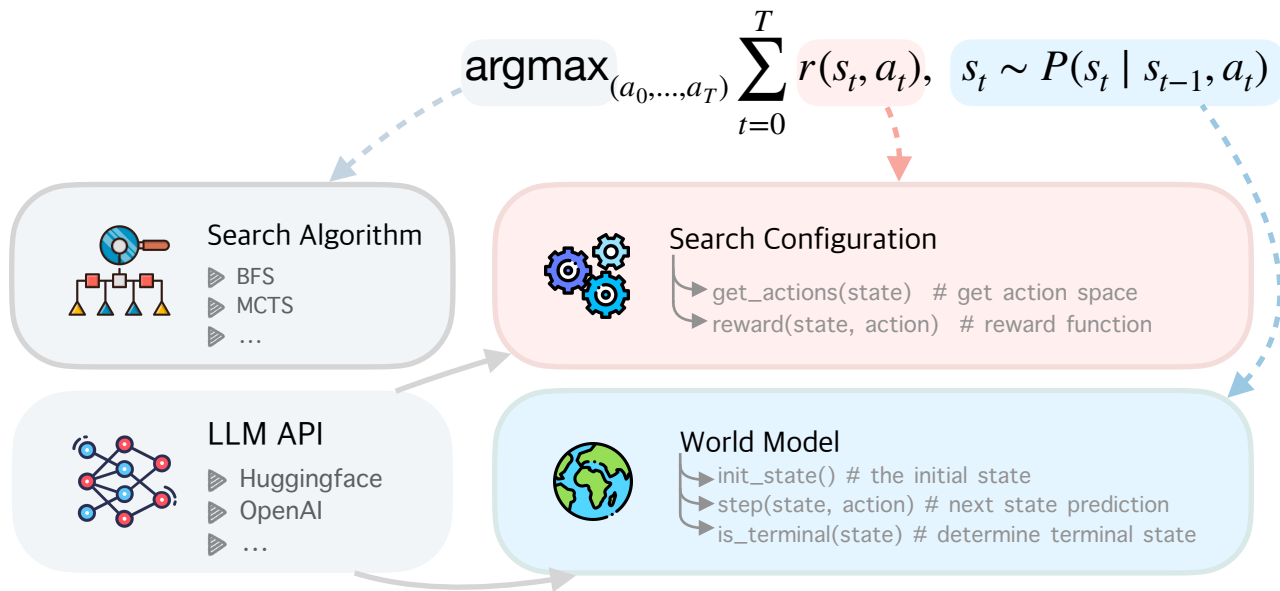
LLM Reasoners: A library for complex reasoning with LLMs



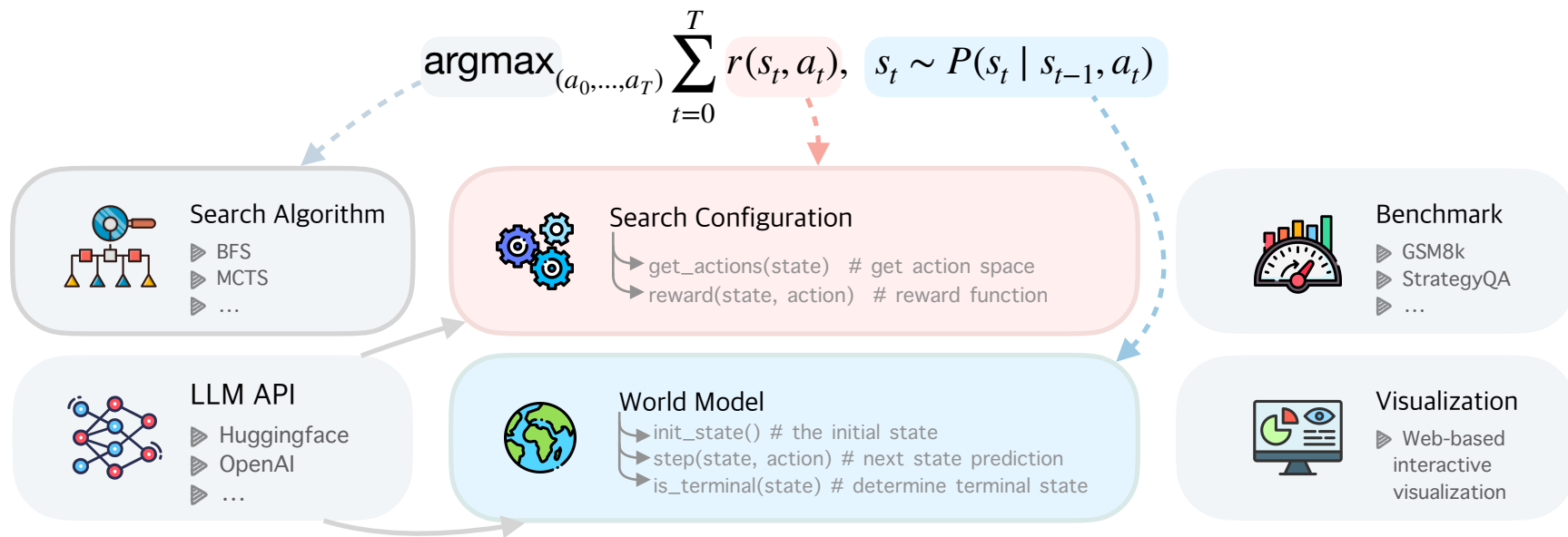
LLM Reasoners: A library for complex reasoning with LLMs

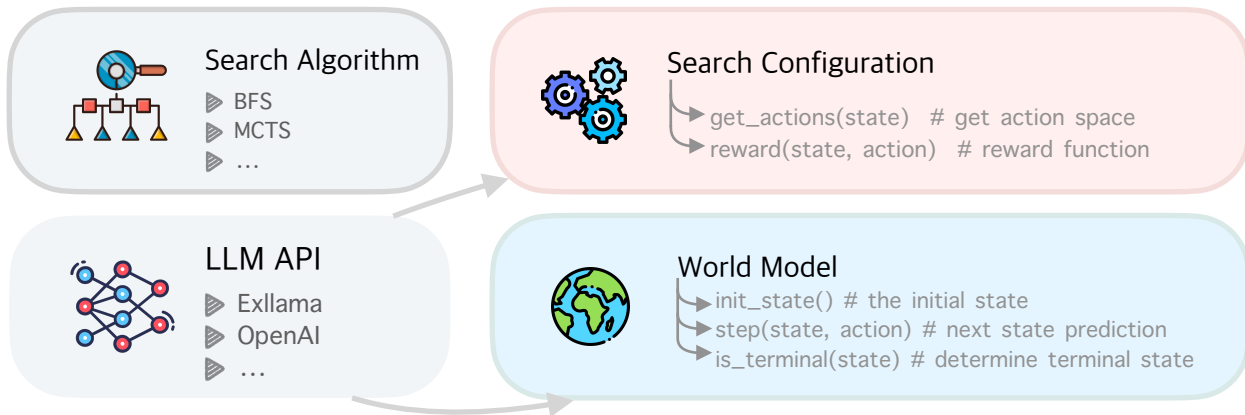


LLM Reasoners: A library for complex reasoning with LLMs



LLM Reasoners: A library for complex reasoning with LLMs





```

from reasoners import SearchConfig, WorldModel
from reasoners.algorithm import MCTS
from reasoners.lm import Llama2Model
from reasoners import Reasoner

class MyWorldModel(WorldModel):
    def step(self, state, action):
        return self.llm.generate(self.next_state_prompt.format(state, action))
    ...

class MyConfig(SearchConfig):
    def reward(self, state, action):
        self_eval = self.llm.generate(self.eval_prompt.format(state, action))
        return self_eval
    ...

reasoner = Reasoner(
    world_model=MyWorldModel(), search_config=MyConfig(), search_algo= MCTS()
)

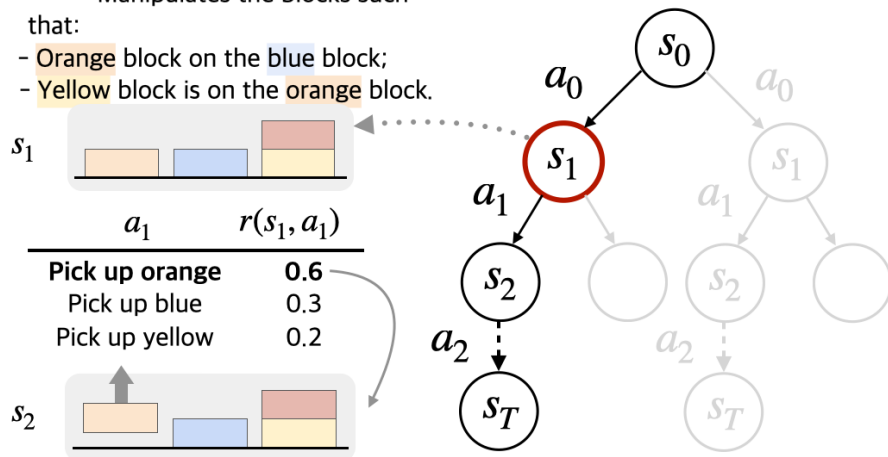
```

Task:

Manipulates the blocks such

that:

- Orange block on the blue block;
- Yellow block is on the orange block.

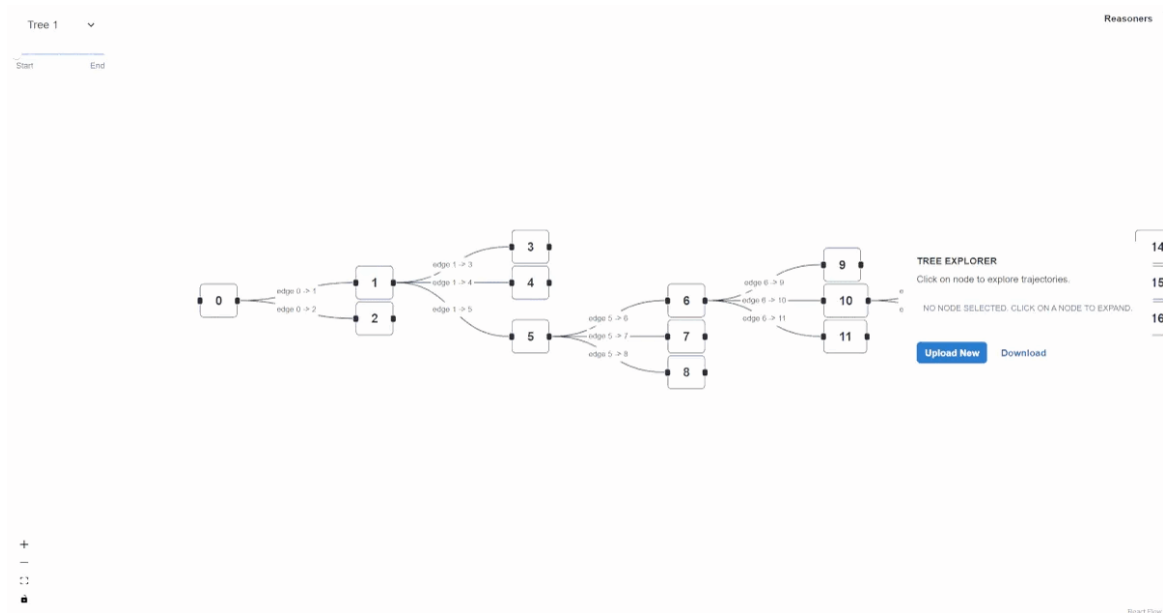


LLM Reasoners: A library for complex reasoning with LLMs



Visualization

► Web-based
interactive
visualization



Outline

Reasoning with LLMs:

- Algorithms
- Library
- **Evaluation**
- Analysis

Large Language Model Step-by-step Reasoning

Q2: How to evaluate step-by-step reasoning?

Q: Does Amtrak operate four wheel vehicles?

1. Amtrak transports people with trains and buses.
2. A bus is a four wheel vehicle.
3. The answer is **yes**.

Ground Truth

1. Amtrak operates **trains, which are four wheel vehicles.**
2. Thus, Amtrak operates four wheel vehicles.
3. So the answer is **yes**.

Llama-2 70B

Correct answer **but incorrect reasoning**

(39% of the cases in StrategyQA)

Reasoning Chain Evaluation

Previous methods:

- Compare to human-written reference (Celikyilmaz et al., 2020)
- Train a model to evaluate (Golovneva et al., 2022)
- Prompt GPT-4 to evaluate (He et al., 2023)

Reasoning Chain Evaluation

Previous methods:

- Compare to **human-written reference** (Celikyilmaz et al., 2020)
- Train a model to evaluate (Golovneva et al., 2022) **Training data**
- Prompt GPT-4 to evaluate (He et al., 2023, Tyen et al., 2023)

Human-written demonstration

- **Need additional human efforts**

Reasoning Chain Evaluation

Previous methods:

- Compare to human-written reference (Celikyilmaz et al., 2020)
- Train a model to evaluate (Golovneva et al., 2022)
- **Prompt GPT-4 to evaluate (He et al., 2023, Tyen et al., 2023)**

Instruction to GPT-4
not adaptive to
different tasks

LLMs cannot *find* reasoning errors, but can *correct* them!

Gladys Tyen^{*1}, Hassan Mansoor², Victor Cărbune², Peter Chen^{†2}, Tony Mak^{†2}

¹University of Cambridge, Dept. of Computer Science & Technology, ALTA Institute

²Google Research

gladys.tyen@cl.cam.ac.uk

{hassan, chenfeif, tonymak, vcarbune}@google.com

- Need additional human efforts
- **Unsatisfactory evaluation accuracy**



Automatic Reasoning Chain Evaluation (AutoRace)

Q: Can one ignite helium?

1. Helium is an odorless and tasteless gas.
2. Helium has no color.
3. So the answer is no.



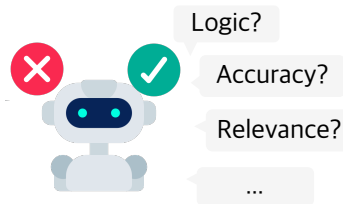
Is this answer correct?

The given answer is partially correct...

Automatic Reasoning Chain Evaluation (AutoRace)

Q: Can one ignite helium?

1. Helium is an odorless and tasteless gas.
2. Helium has no color.
3. So the answer is no.

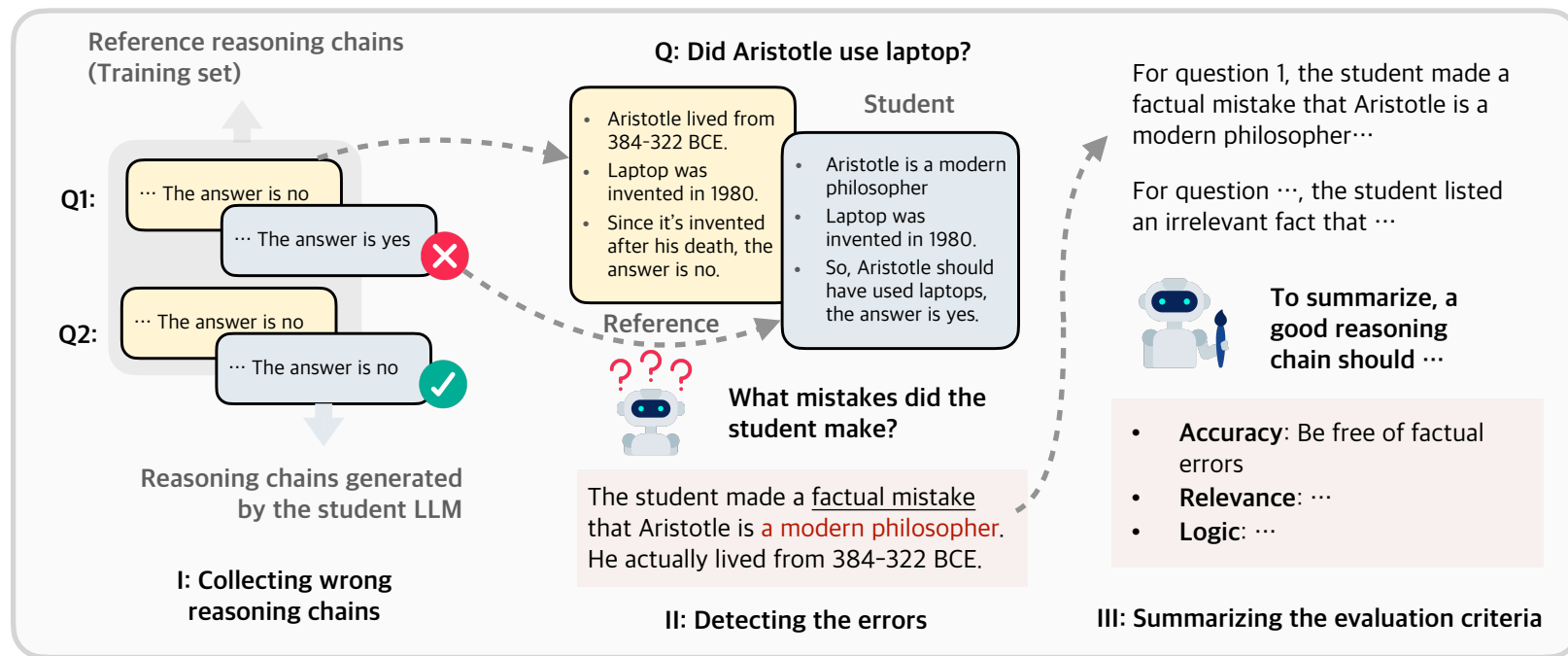


Following the criteria,
evaluate the reasoning chain
step by step.

- Accuracy: ..., correct.
- Relevance: The information in the first two steps are irrelevant to the question.
- Logic: The final step cannot be inferred from the previous steps.

So, the reasoning is **INCORRECT**.

Automatic Reasoning Chain Evaluation (AutoRace)



Criterion List Construction

Automatic Reasoning Chain Evaluation

1. Helium is an odorless and tasteless gas.
2. Helium has no color.
3. So the answer is no.

For question 1, the student made a factual mistake that Aristotle is a modern philosopher...

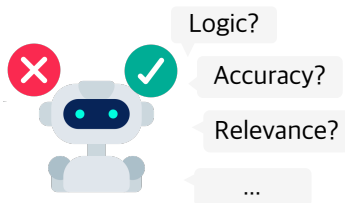
For question ..., the student listed an irrelevant fact that ...



To summarize, a good reasoning chain should ...

- **Accuracy:** Be free of factual errors
- **Relevance:** ...
- **Logic:** ...

Q: Can one ignite helium?



Following the criteria,
evaluate the reasoning chain
step by step.

- Accuracy: ..., correct.
- Relevance: The information in the first two steps are irrelevant to the question.
- Logic: The final step cannot be inferred from the previous steps.

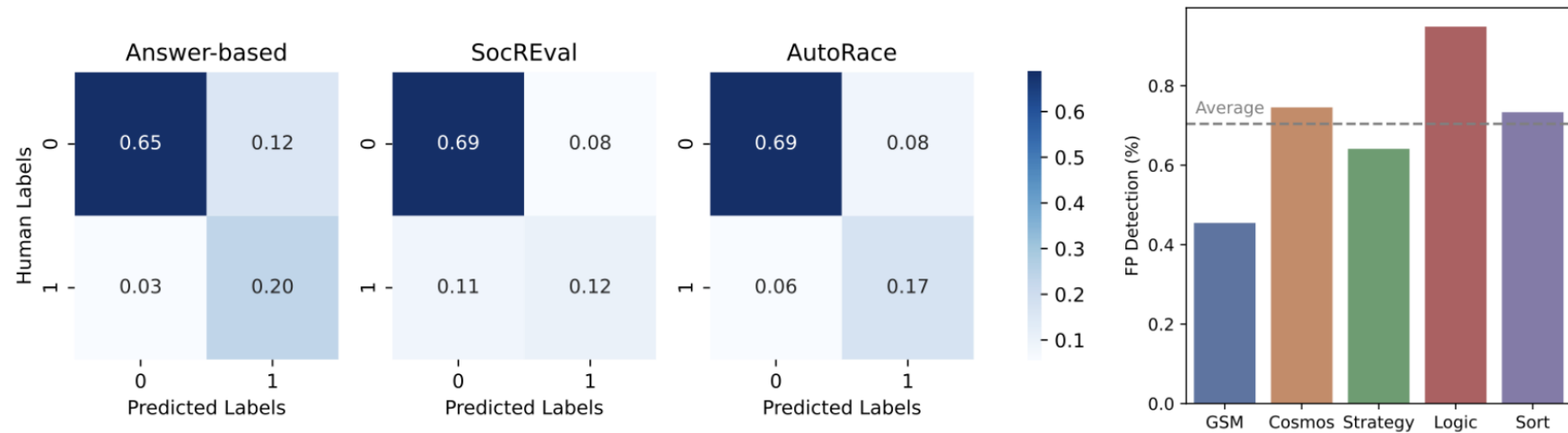
So, the reasoning is **INCORRECT**.

Automatic Reasoning Chain Evaluation (AutoRace)

Method	Math		Common		Logical		Average	Fully Auto.
	GSM8k	Arith	Strategy	Cosmos	Logic	Sort		
<i>Answer-based</i>	0.94	0.94	0.76	0.67	0.87	0.94	0.85	
SocREval	0.89	0.85	0.71	0.80	0.89	0.77	0.82	✗
Direct (trace)	0.90	0.38	0.80	0.72	0.21	0.36	0.56	✗
Direct (step)	0.85	0.43	0.83	0.73	0.75	0.33	0.65	✗
CoT (step)	0.78	0.74	0.78	0.72	1.00	0.86	0.81	✗
AutoRace (Ours)	0.91	0.85	0.79	0.78	0.97	0.86	0.86	✓

- No additional human efforts needed
- Better average accuracy
- Remain robust across different tasks

Automatic Reasoning Chain Evaluation (AutoRace)



- Balancing precision and recall
- Detect 70% of the false positive chains

Automatic Reasoning Chain Evaluation (AutoRace)

False positive chains recognized by AutoRace:

Type	Question	Answer	Explanation
A	Henry made two stops during his 60-mile bike trip. He first stopped after 20 miles. His second stop was 15 miles before the end of the trip. How many miles did he travel between his first and second stops?	Step 1 - Henry traveled 20 miles + 15 miles = 35 miles between his first and second stops. Step 2 - Henry traveled 60 miles - 35 miles = 25 miles without stopping. Step 3 - The answer is 25	There is a hallucination in an early reasoning step, but the following reasoning steps ignore the mistake and reach the correct answer.
B	Is "A Tale of Two Cities" a parody of the Bible?	"A Tale of Two Cities" was published in 1859, and the Bible was published much earlier. Thus , "A Tale of Two Cities" is not a parody of the Bible. So the answer is no.	The reasoning chain has obvious or multiple mistakes, but hits the correct answer by chance.
C	Is clementine pith highly sought after?	Clementine pith is not highly sought after. So the answer is no.	The reasoning chain is not informative at all, though the answer is correct.

Outline

Reasoning with LLMs:

- Algorithms
- Library
- Evaluation
- Analysis

Experimental Results

Method	Math			Logical	Common	Embodied
	GSM8k*	AQuA*	Game24	PrOnto	StrategyQA*	Blocks
CoT	0.37 (0.54)	0.09 (0.34)	0.04	0.58	0.34 (0.76)	0.05
ToT (BFS)	0.53 (0.58)	0.15 (0.42)	0.04	0.52	0.41 (0.76)	0.09
ToT (DFS)	0.45 (0.52)	0.10 (0.36)	0.07	0.44	0.42 (0.76)	0.08
RAP	0.58 (0.64)	0.20 (0.47)	0.07	0.59	0.28 (0.77)	0.51

For three datasets marked with*, we evaluate the reasoning chain with both AutoRace and the answer correctness (in brackets)
Other tasks use oracle evaluator (e.g., program/simulator)

Experimental Analysis

From auto-regressive decoding to reward-guided search

Method	Math			Logical	Common	Embodied
	GSM8k*	AQuA*	Game24	PrOnto	StrategyQA*	Blocks
CoT	0.37 (0.54)	0.09 (0.34)	0.04	0.58	0.34 (0.76)	0.05
ToT (BFS)	0.53 (0.58)	0.15 (0.42)	0.04	0.52	0.41 (0.76)	0.09
ToT (DFS)	0.45 (0.52)	0.10 (0.36)	0.07	0.44	0.42 (0.76)	0.08
RAP	0.58 (0.64)	0.20 (0.47)	0.07	0.59	0.28 (0.77)	0.51

Overall improved performance with search

Experimental Analysis

From auto-regressive decoding to reward-guided search

Method	Math			Logical	Common	Embodied
	GSM8k*	AQuA*	Game24	PrOnto	StrategyQA*	Blocks
CoT	0.37 (0.54)	0.09 (0.34)	0.04	0.58	0.34 (0.76)	0.05
ToT (BFS)	0.53 (0.58)	0.15 (0.42)	0.04	0.52	0.41 (0.76)	0.09
ToT (DFS)	0.45 (0.52)	0.10 (0.36)	0.07	0.44	0.42 (0.76)	0.08
RAP	0.58 (0.64)	0.20 (0.47)	0.07	0.59	0.28 (0.77)	0.51

Less false positive

Experimental Analysis

Less false positive chains

A mechanism to “regret”



From auto-regressive decoding to reward-guided search

Type	Question	Answer	Explanation
A	Henry made two stops during his 60-mile bike trip. He first stopped after 20 miles. His second stop was 15 miles before the end of the trip. How many miles did he travel between his first and second stops?	Step 1 - Henry traveled 20 miles + 15 miles = 35 miles between his first and second stops . Step 2 - Henry traveled 60 miles - 35 miles = 25 miles without stopping. Step 3 - The answer is 25	There is a hallucination in an early reasoning step, but the following reasoning steps ignore the mistake and reach the correct answer.
B	Is "A Tale of Two Cities" a parody of the Bible?	"A Tale of Two Cities" was published in 1859, and the Bible was published much earlier. Thus , "A Tale of Two Cities" is not a parody of the Bible. So the answer is no.	The reasoning chain has obvious or multiple mistakes, but hits the correct answer by chance.
C	Is clementine pith highly sought after?	Clementine pith is not highly sought after . So the answer is no.	The reasoning chain is not informative at all, though the answer is correct.

Experimental Analysis

From auto-regressive decoding to reward-guided search

Method	Math			Logical	Common	Embodied
	GSM8k*	AQuA*	Game24	PrOnto	StrategyQA*	Blocks
CoT	0.37 (0.54)	0.09 (0.34)	0.04	0.58	0.34 (0.76)	0.05
→ ToT (BFS)	0.53 (0.58)	0.15 (0.42)	0.04	0.52	0.41 (0.76)	0.09
→ ToT (DFS)	0.45 (0.52)	0.10 (0.36)	0.07	0.44	0.42 (0.76)	0.08
RAP	0.58 (0.64)	0.20 (0.47)	0.07	0.59	0.28 (0.77)	0.51

The breadth of search matters more than the depth

Experimental Analysis

The impact of world model

Method	Math			Logical	Common	Embodied
	GSM8k*	AQuA*	Game24	PrOnto	StrategyQA*	Blocks
CoT	0.37 (0.54)	0.09 (0.34)	0.04	0.58	0.34 (0.76)	0.05
ToT (BFS)	0.53 (0.58)	0.15 (0.42)	0.04	0.52	0.41 (0.76)	0.09
ToT (DFS)	0.45 (0.52)	0.10 (0.36)	0.07	0.44	0.42 (0.76)	0.08
RAP	0.58 (0.64)	0.20 (0.47)	0.07	0.59	0.28 (0.77)	0.51

A world model improves performance in most domains

Most effective in **Embodied Planning** task

Experimental Analysis

The impact of prompting format

Method	Math			Logical	Common	Embodied
	GSM8k*	AQuA*	Game24	PrOnto	StrategyQA*	Blocks
CoT	0.37 (0.54)	0.09 (0.34)	0.04	0.58	0.34 (0.76)	0.05
ToT (BFS)	0.53 (0.58)	0.15 (0.42)	0.04	0.52	0.41 (0.76)	0.09
ToT (DFS)	0.45 (0.52)	0.10 (0.36)	0.07	0.44	0.42 (0.76)	0.08
RAP	0.58 (0.64)	0.20 (0.47)	0.07	0.59	0.28 (0.77)	0.51

Experimental Analysis

Q: Did either Kublai Khan or his grandfather practice monogamy?

Chain-of-thoughts:

Step 1: Both Kublai Khan and his grandfather, Genghis Khan, were polygamous.

Step 2: Thus, neither Kublai Khan nor his grandfather practiced monogamy.

Step 3: So the answer is no.



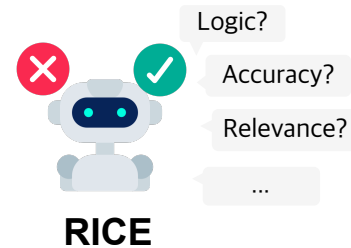
RAP:

Step 1: Did Kublai Khan have a harem? - Kublai Khan had a harem of 7,000 women. So the answer is yes.

Step 2: Did Genghis Khan have a harem? - Genghis Khan had a harem of 500 women. So the answer is yes.

Step 3: Does having a harem of women mean practicing polygamy? - Having a harem of women means practicing polygamy. So the answer is yes.

Step 4: ... So the answer is no.



Experimental Analysis

The impact of prompting format

Type	Question	Answer	Explanation
A	Henry made two stops during his 60-mile bike trip. He first stopped after 20 miles. His second stop was 15 miles before the end of the trip. How many miles did he travel between his first and second stops?	Step 1 - Henry traveled 20 miles + 15 miles = 35 miles between his first and second stops . Step 2 - Henry traveled 60 miles - 35 miles = 25 miles without stopping. Step 3 - The answer is 25	There is a hallucination in an early reasoning step, but the following reasoning steps ignore the mistake and reach the correct answer.
B	Is "A Tale of Two Cities" a parody of the Bible?	"A Tale of Two Cities" was published in 1859, and the Bible was published much earlier. Thus , "A Tale of Two Cities" is not a parody of the Bible. So the answer is no.	The reasoning chain has obvious or multiple mistakes, but hits the correct answer by chance.
C	Is clementine pith highly sought after?	Clementine pith is not highly sought after . So the answer is no.	The reasoning chain is not informative at all, though the answer is correct.

Experimental Analysis

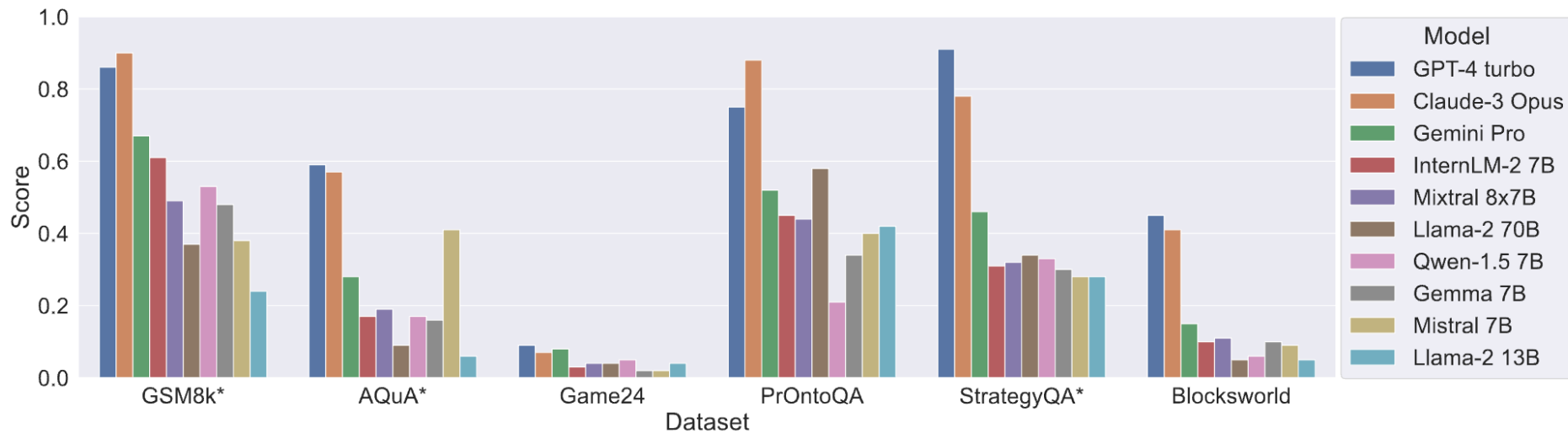
The impact of prompting format

Method	Math			Logical	Common	Embodied
	GSM8k*	AQuA*	Game24	PrOnto	StrategyQA*	Blocks
CoT	0.37 (0.54)	0.09 (0.34)	0.04	0.58	0.34 (0.76)	0.05
ToT (BFS)	0.53 (0.58)	0.15 (0.42)	0.04	0.52	0.41 (0.76)	0.09
ToT (DFS)	0.45 (0.52)	0.10 (0.36)	0.07	0.44	0.42 (0.76)	0.08
RAP	0.58 (0.64)	0.20 (0.47)	0.07	0.59	0.28 (0.77)	0.51

Easier to trigger false positives

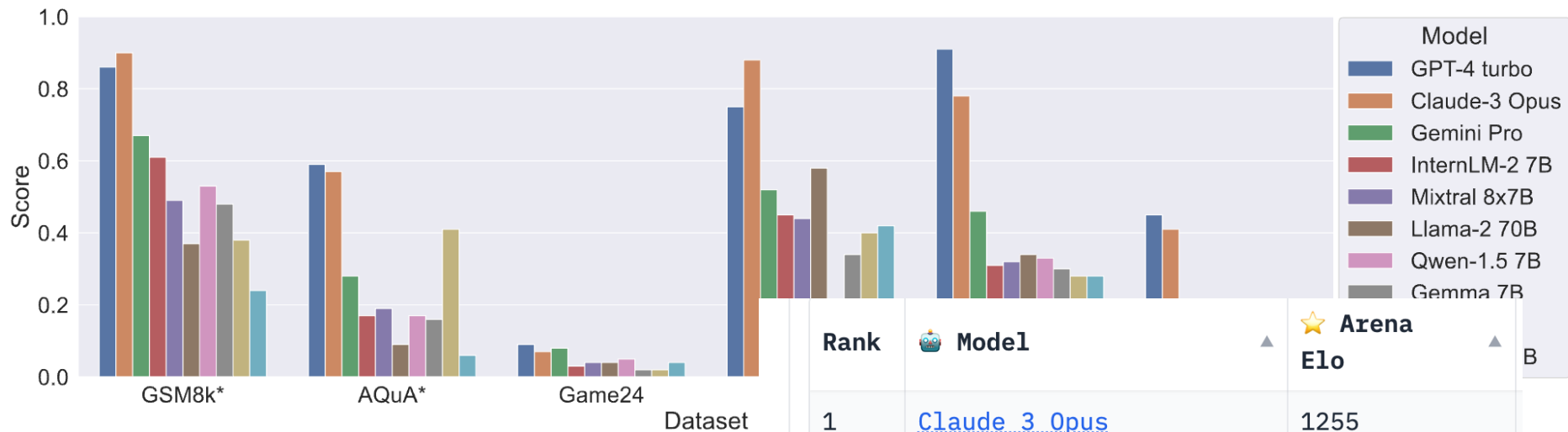
But only for certain datasets, where
the details are not necessary

AutoRace Leaderboard



Top models can almost solve GSM8k, StrategyQA
But fails on long-range reasoning tasks requiring planning

AutoRace Leaderboard



Top 3 models align with ChatArena

- Strong reasoning ability is necessary for SOTA chatbot

Rank	Model	Arena Elo
1	Claude 3 Opus	1255
1	GPT-4-1106-preview	1252
	GPT-4-0125-preview	1249
4	Bard (Gemini Pro)	1204

Summary

Reasoning with LLMs:

- Algorithms
- Library
- Evaluation
- Analysis



llm-reasoners.net