rStar: Mutual Reasoning Makes Smaller LLMs Stronger Problem-Solver

Zhenting Qi*, Mingyuan Ma*, Jiahang Xu*,
Li Lyna Zhang, Fan Yang, Mao Yang

Microsoft

HARVARD UNIVERSITY

Intro: Reasoning in SLMs

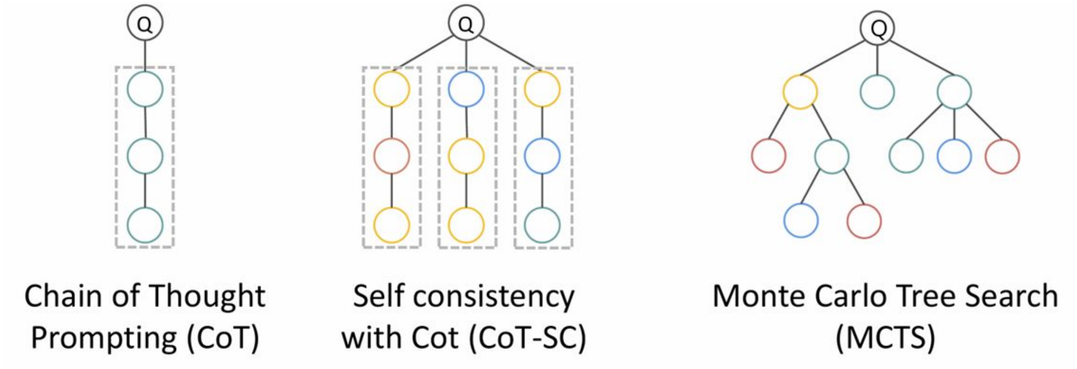
- Solving reasoning tasks is essential. However, currently only powerful models (GPT-4,openAI-o1…) demonstrate strong reasoning abilities.

- While smaller language models (SLMs) are efficient in computation and resource use, enabling faster inference, they often struggle with reasoning tasks.

- Hence, we introduce rStar, Enhancing Reasoning Capabilities of SLMs

Intro: Test-time Computation

- A unified perspective: **Generator** and **Verifier**
- Generator: modify the LLMs’ proposal distribution to sample diverse generations



Chain of Thought Prompting (CoT)

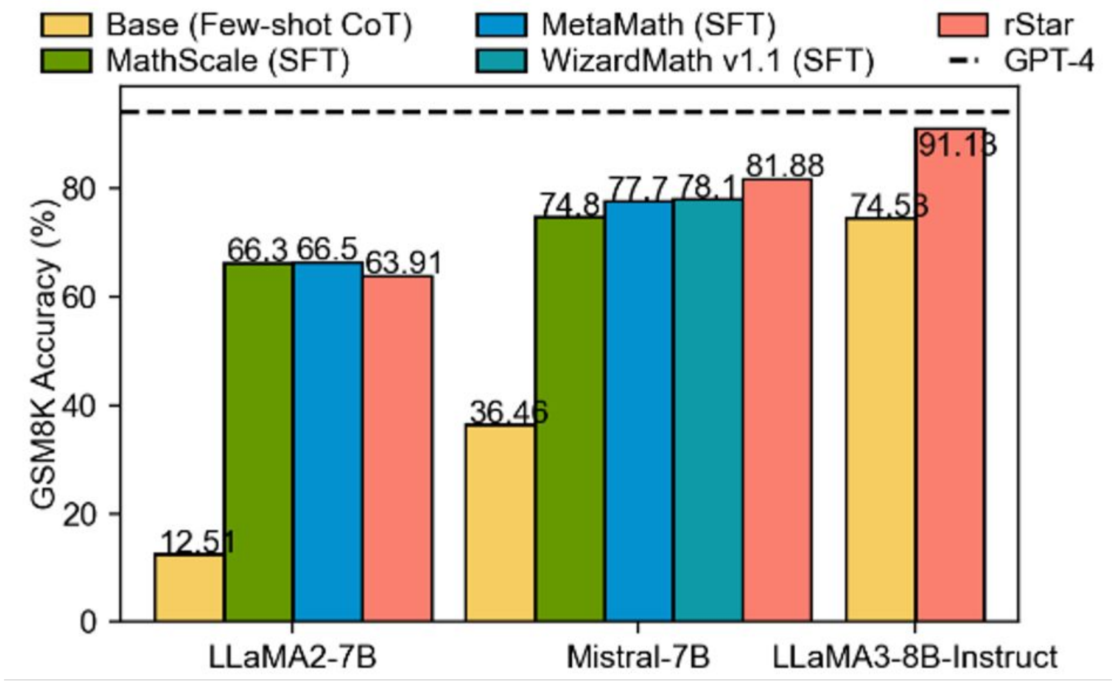
Self consistency with Cot (CoT-SC)

Monte Carlo Tree Search (MCTS)

- Verifier: aggregate or select the best answer from the generated samples
    - Reward model (costly to train😓, lacks task generalization😓)
    - Self-rewarding (sometimes leads to rather random outcomes😓)

Intro: Preliminary Results

rStar makes SLMs highly capable problem-solvers, matching or even surpassing the reasoning performance achieved after domain-specialized SFT.

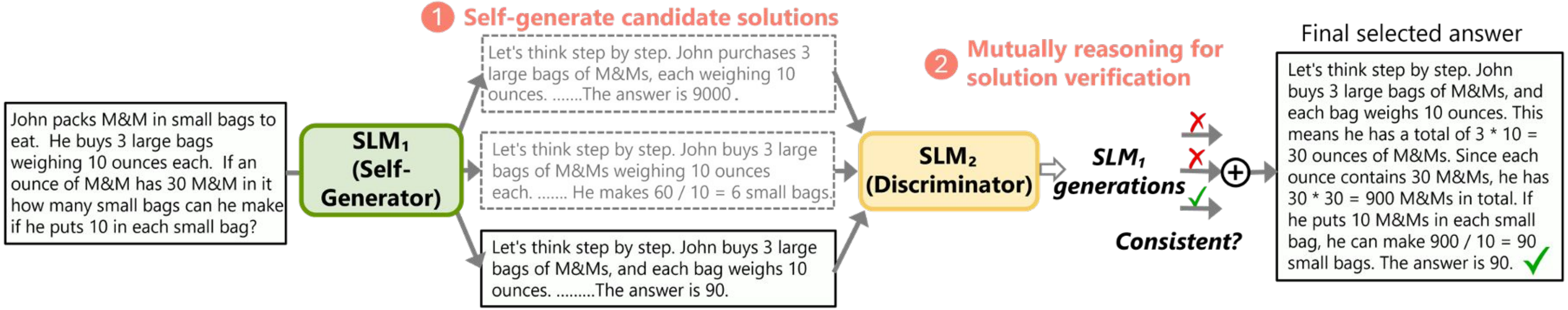


Intro: Insights

- **Insight 1:** *Decomposing complex reasoning tasks into simpler subtasks* helps SLMs handle them more effectively.

- **Insight 2:** *Agreement among peers* (i.e., two SLMs) on derived answers suggests a higher likelihood of correctness.

Method: Overall Framework


1 Self-generate candidate solutions
John packs M&M in small bags to eat. He buys 3 large bags weighing 10 ounces each. If an ounce of M&M has 30 M&M in it how many small bags can he make if he puts 10 in each small bag?
SLM₁ (Self-Generator)
Let's think step by step. John purchases 3 large bags of M&Ms, each weighing 10 ounces. ....The answer is 9000 .
Let's think step by step. John buys 3 large bags of M&Ms weighing 10 ounces each. .... He makes 60 / 10 = 6 small bags.
Let's think step by step. John buys 3 large bags of M&Ms, and each bag weighs 10 ounces. ....The answer is 90.
2 Mutually reasoning for solution verification
SLM₂ (Discriminator)
SLM₁ generations
✗
✗
✓
Consistent?
Final selected answer
Let's think step by step. John buys 3 large bags of M&Ms, and each bag weighs 10 ounces. This means he has a total of 3 * 10 = 30 ounces of M&Ms. Since each ounce contains 30 M&Ms, he has 30 * 30 = 900 M&Ms in total. If he puts 10 M&Ms in each small bag, he can make 900 / 10 = 90 small bags. The answer is 90. ✓

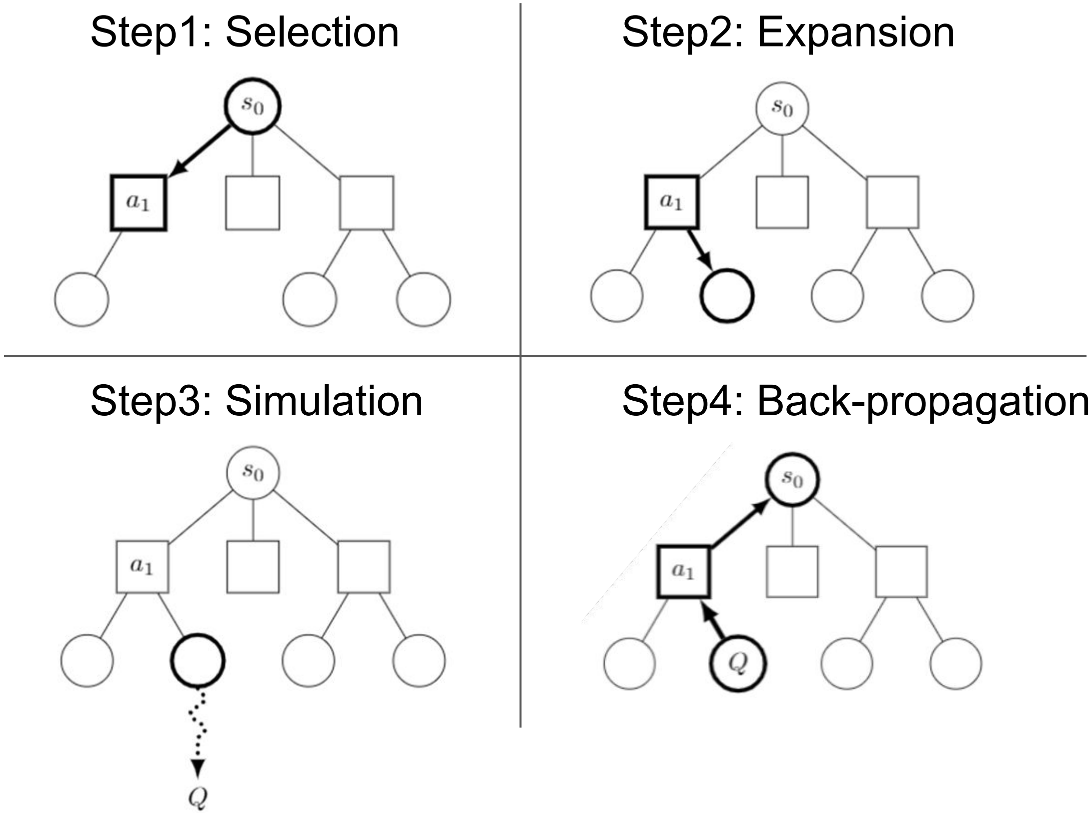
Solution generation:

- augments the target SLM by effectively breaking down a given problem into manageable subtasks
- utilizes MCTS to generate candidate solution trajectories and allow the SLM to perform a human-like reasoning actions

Solution verification:

- uses another SLM to provide unsupervised feedback on each trajectory
- mutually validates the solution consistency

Method: Monte-Carlo Tree Search (MCTS)



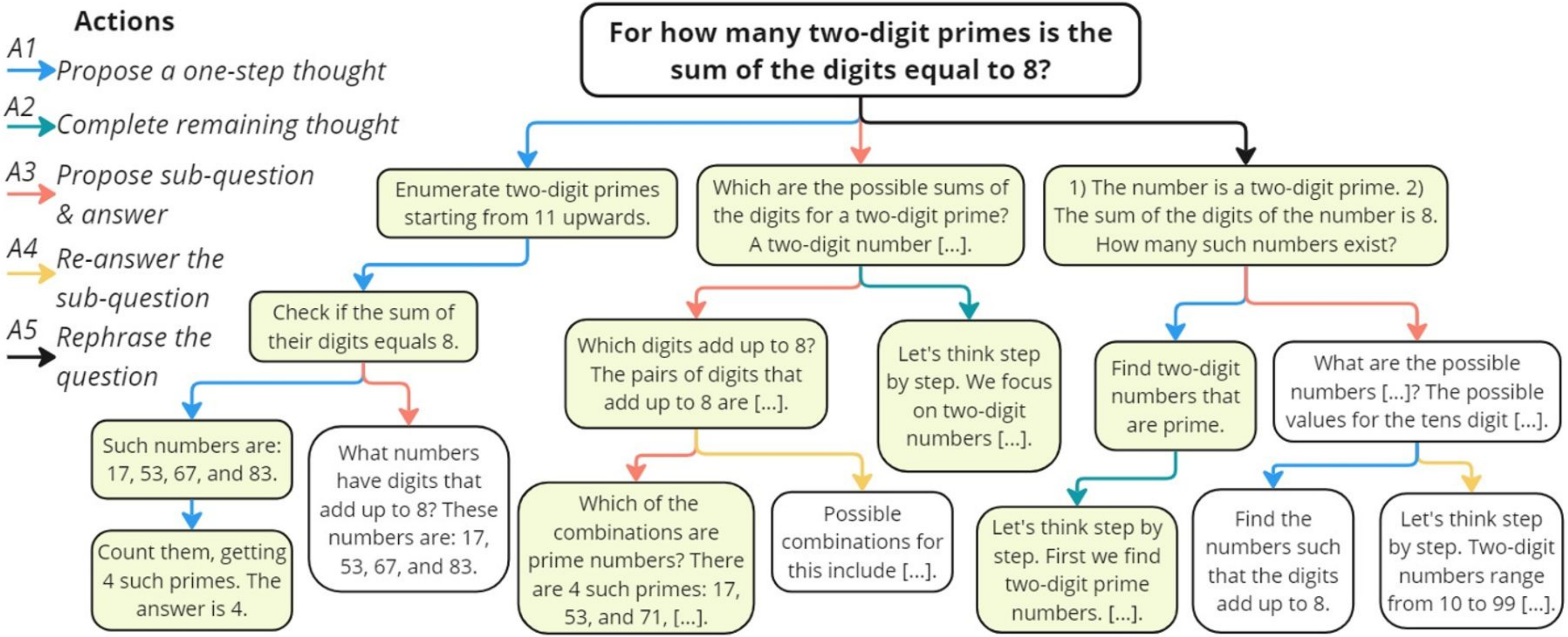
Step1: Selection

Step2: Expansion

Step3: Simulation

Step4: Back-propagation

Method: Enrich Generator’s Action Space



Actions

A1 Propose a one-step thought

A2 Complete remaining thought

A3 Propose sub-question & answer

A4 Re-answer the sub-question

A5 Rephrase the question

For how many two-digit primes is the sum of the digits equal to 8?

Enumerate two-digit primes starting from 11 upwards.

Which are the possible sums of the digits for a two-digit prime? A two-digit number [...].

1) The number is a two-digit prime. 2) The sum of the digits of the number is 8. How many such numbers exist?

Check if the sum of their digits equals 8.

Which digits add up to 8? The pairs of digits that add up to 8 are [...].

Let's think step by step. We focus on two-digit numbers [...].

Find two-digit numbers that are prime.

What are the possible numbers [...]? The possible values for the tens digit [...].

Such numbers are: 17, 53, 67, and 83.

What numbers have digits that add up to 8? These numbers are: 17, 53, 67, and 83.

Count them, getting 4 such primes. The answer is 4.

Which of the combinations are prime numbers? There are 4 such primes: 17, 53, and 71, [...].

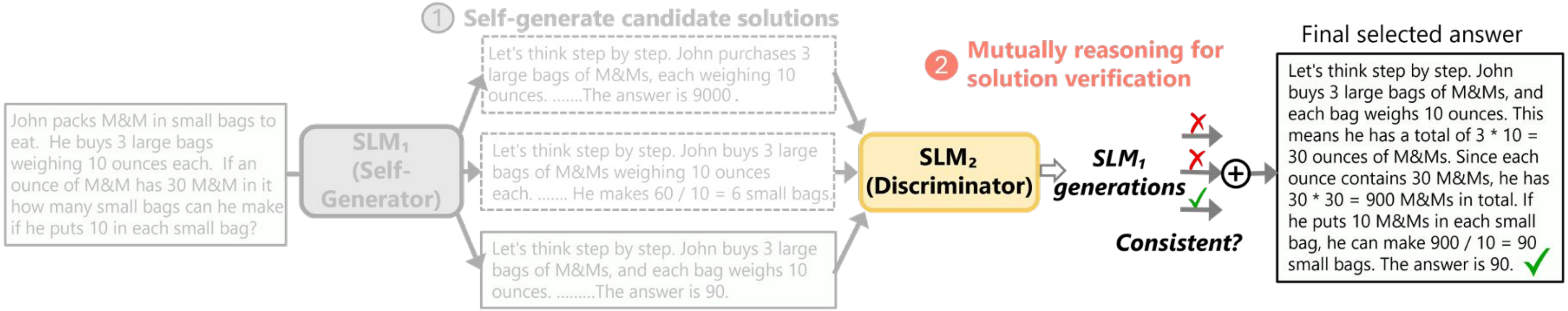
Possible combinations for this include [...].

Let's think step by step. First we find two-digit prime numbers. [...].

Find the numbers such that the digits add up to 8.

Let's think step by step. Two-digit numbers range from 10 to 99 [...].

Method: Discriminator with Mutual Consistency

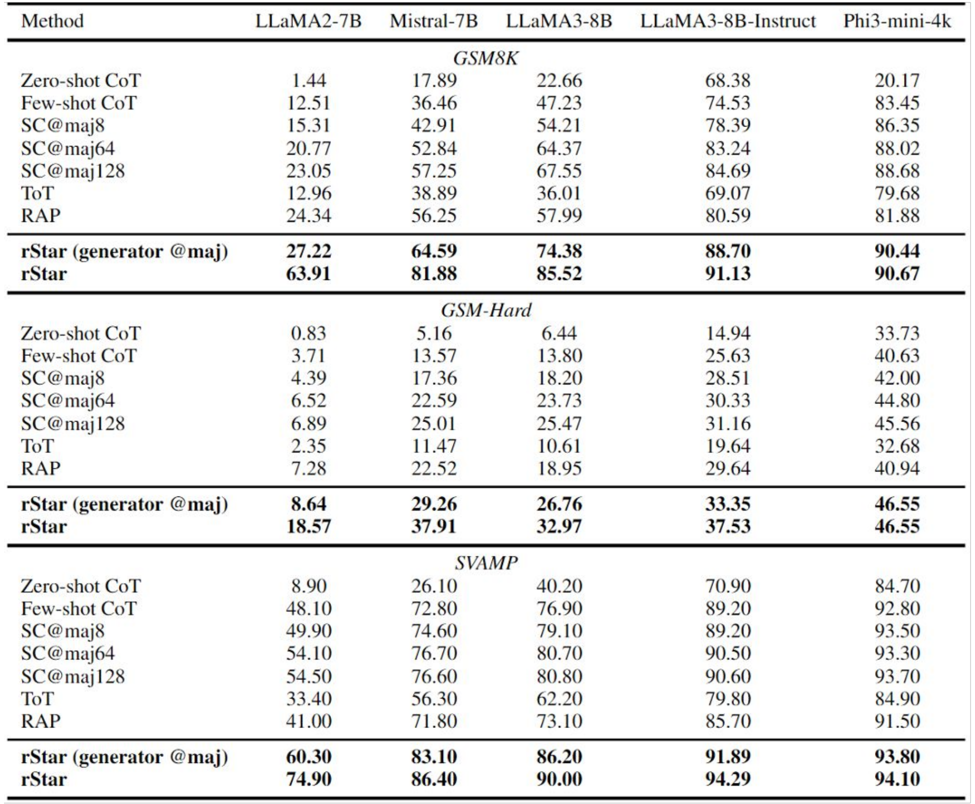


① Self-generate candidate solutions
John packs M&M in small bags to eat. He buys 3 large bags weighing 10 ounces each. If an ounce of M&M has 30 M&M in it how many small bags can he make if he puts 10 in each small bag?
SLM₁ (Self-Generator)
Let's think step by step. John purchases 3 large bags of M&Ms, each weighing 10 ounces. ....The answer is 9000 .
Let's think step by step. John buys 3 large bags of M&Ms weighing 10 ounces each. .... He makes 60 / 10 = 6 small bags.
Let's think step by step. John buys 3 large bags of M&Ms, and each bag weighs 10 ounces. ....The answer is 90.
② Mutually reasoning for solution verification
SLM₂ (Discriminator)
SLM₁ generations
✗
✗
✓
Consistent?
Final selected answer
Let's think step by step. John buys 3 large bags of M&Ms, and each bag weighs 10 ounces. This means he has a total of 3 * 10 = 30 ounces of M&Ms. Since each ounce contains 30 M&Ms, he has 30 * 30 = 900 M&Ms in total. If he puts 10 M&Ms in each small bag, he can make 900 / 10 = 90 small bags. The answer is 90. ✓

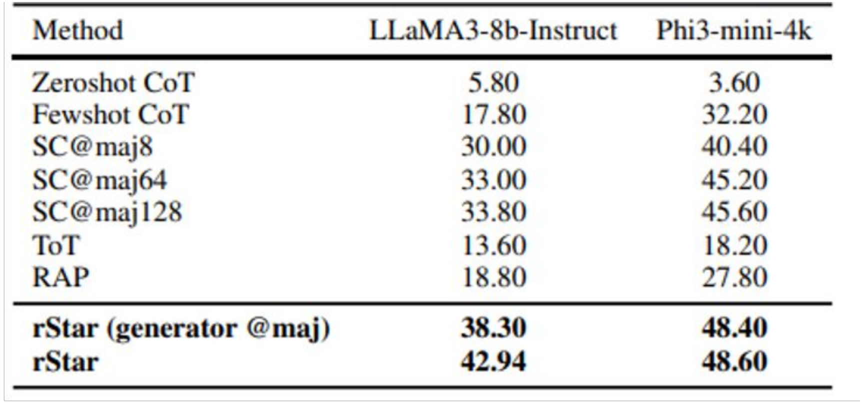
Mutual consistency: use another SLM to act as a discriminator, providing unsupervised feedback on each candidate trajectory

- Partial reasoning steps as the hints: *reduce the reasoning difficulty for another SLM*
- Mirrors human experience (derived answers from the same initial steps indicates a high likelihood of correctness)

Experiments: Mathematical Tasks

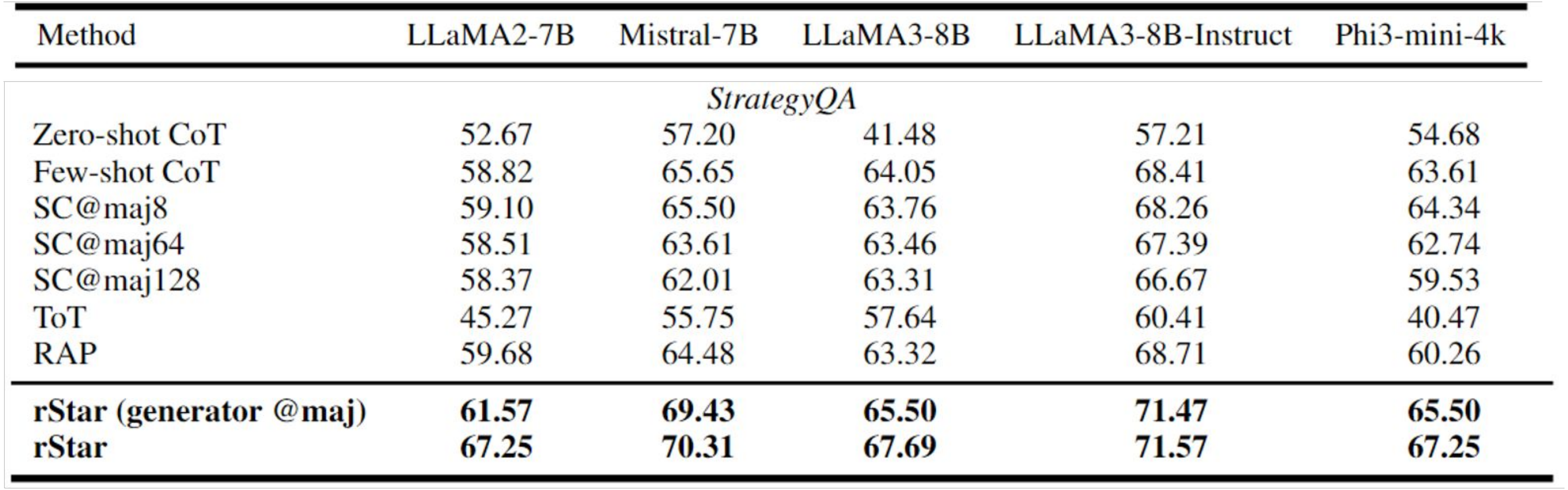


Method	LLaMA2-7B	Mistral-7B	LLaMA3-8B	LLaMA3-8B-Instruct	Phi3-mini-4k
			GSM8K		
Zero-shot CoT	1.44	17.89	22.66	68.38	20.17
Few-shot CoT	12.51	36.46	47.23	74.53	83.45
SC@maj8	15.31	42.91	54.21	78.39	86.35
SC@maj64	20.77	52.84	64.37	83.24	88.02
SC@maj128	23.05	57.25	67.55	84.69	88.68
ToT	12.96	38.89	36.01	69.07	79.68
RAP	24.34	56.25	57.99	80.59	81.88
rStar (generator @maj)	**27.22**	**64.59**	**74.38**	**88.70**	**90.44**
rStar	**63.91**	**81.88**	**85.52**	**91.13**	**90.67**
			GSM-Hard		
Zero-shot CoT	0.83	5.16	6.44	14.94	33.73
Few-shot CoT	3.71	13.57	13.80	25.63	40.63
SC@maj8	4.39	17.36	18.20	28.51	42.00
SC@maj64	6.52	22.59	23.73	30.33	44.80
SC@maj128	6.89	25.01	25.47	31.16	45.56
ToT	2.35	11.47	10.61	19.64	32.68
RAP	7.28	22.52	18.95	29.64	40.94
rStar (generator @maj)	**8.64**	**29.26**	**26.76**	**33.35**	**46.55**
rStar	**18.57**	**37.91**	**32.97**	**37.53**	**46.55**
			SVAMP		
Zero-shot CoT	8.90	26.10	40.20	70.90	84.70
Few-shot CoT	48.10	72.80	76.90	89.20	92.80
SC@maj8	49.90	74.60	79.10	89.20	93.50
SC@maj64	54.10	76.70	80.70	90.50	93.30
SC@maj128	54.50	76.60	80.80	90.60	93.70
ToT	33.40	56.30	62.20	79.80	84.90
RAP	41.00	71.80	73.10	85.70	91.50
rStar (generator @maj)	**60.30**	**83.10**	**86.20**	**91.89**	**93.80**
rStar	**74.90**	**86.40**	**90.00**	**94.29**	**94.10**



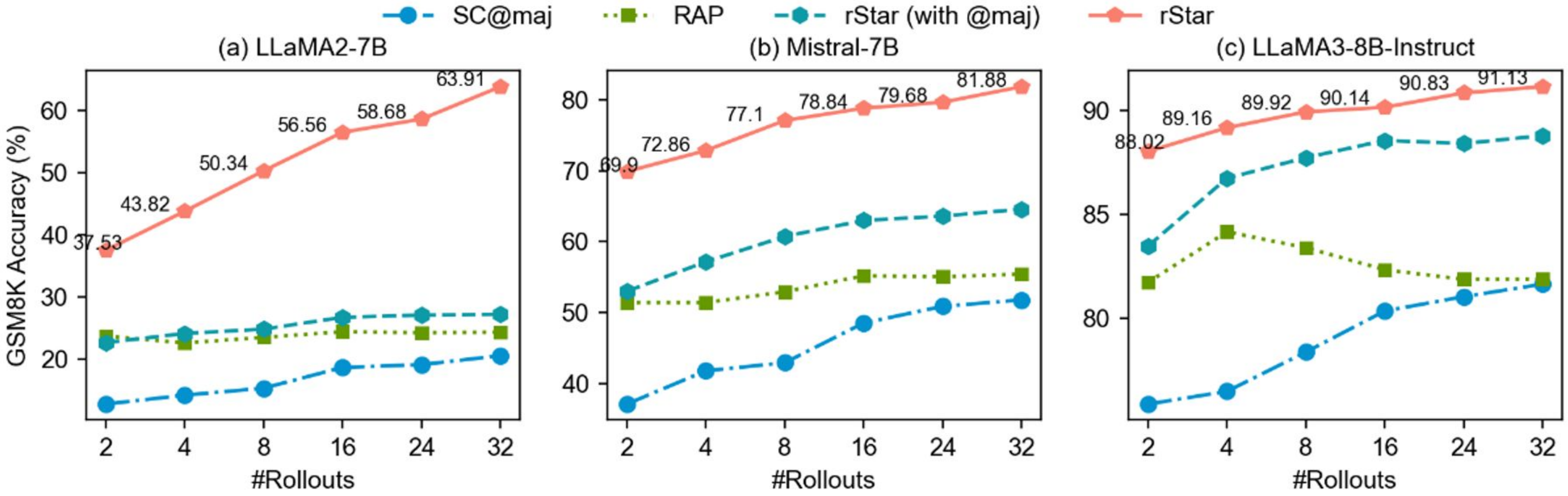
Method	LLaMA3-8b-Instruct	Phi3-mini-4k
Zeroshot CoT	5.80	3.60
Fewshot CoT	17.80	32.20
SC@maj8	30.00	40.40
SC@maj64	33.00	45.20
SC@maj128	33.80	45.60
ToT	13.60	18.20
RAP	18.80	27.80
rStar (generator @maj)	**38.30**	**48.40**
rStar	**42.94**	**48.60**

Experiments: Commonsense Reasoning Tasks



Method	LLaMA2-7B	Mistral-7B	LLaMA3-8B	LLaMA3-8B-Instruct	Phi3-mini-4k
			StrategyQA		
Zero-shot CoT	52.67	57.20	41.48	57.21	54.68
Few-shot CoT	58.82	65.65	64.05	68.41	63.61
SC@maj8	59.10	65.50	63.76	68.26	64.34
SC@maj64	58.51	63.61	63.46	67.39	62.74
SC@maj128	58.37	62.01	63.31	66.67	59.53
ToT	45.27	55.75	57.64	60.41	40.47
RAP	59.68	64.48	63.32	68.71	60.26
rStar (generator @maj)	**61.57**	**69.43**	**65.50**	**71.47**	**65.50**
rStar	**67.25**	**70.31**	**67.69**	**71.57**	**67.25**

Experiments: Test-Time Scaling



Thank you!