A*-Thought: Efficient Reasoning via Bidirectional Compression for Low-Resource Settings

Xiaoang Xu[1], Shuo Wang[2*], Xu Han[2,4,5], Zhenghao Liu[3], Huijia Wu[1], Peipei Li[1], Zhiyuan Liu[2,4,5], Maosong Sun[2,4,5], Zhaofeng He[1*]

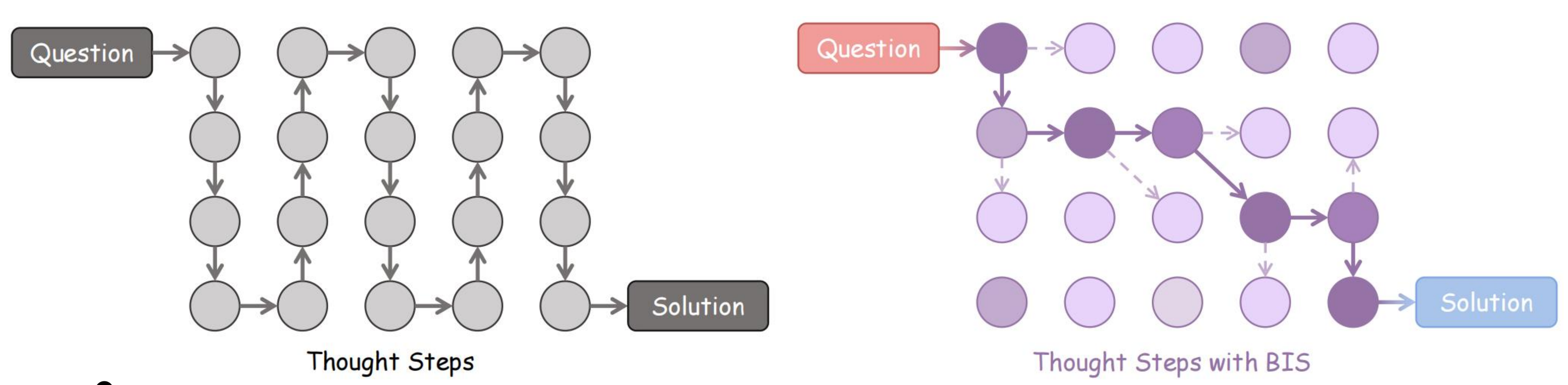
[1] BUPT [2] Dept. of Comp. Sci. & Tech., Tsinghua University [3] Northeastern University [4] Institute for AI, Tsinghua University [5] Beijing National Research Center for Information Science and Technology

Scan QR Code

Chain-of-Thought v.s. A*-Thought

CoT: Redundant & Mazy

A*-Thought: High-Quality & Informative



We retained the more important steps!

Step-Level Bidirectional Importance Score (BIS)

Attention-Level

$$\mathrm{ATTN}(\mathbf{y}|\mathbf{x}) = \frac{1}{H|\mathbf{y}||\mathbf{x}|}\sum_{h=1}^{H}\sum_{j=1}^{|\mathbf{y}|}\sum_{i=1}^{|\mathbf{x}|} a_h(y_j, x_i)$$

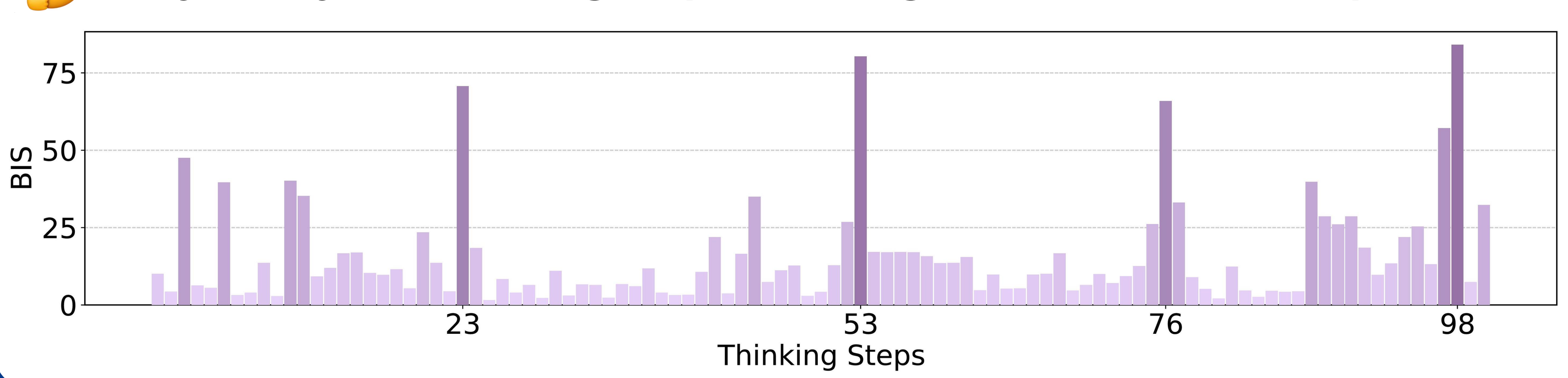
Modle-Level

$$\mathrm{NLL}(\mathbf{y}|\mathbf{x}) = -\frac{1}{|\mathbf{y}|}\sum_{j=1}^{|\mathbf{y}|}\log P(y_j|\mathbf{x}, \mathbf{y}_{<j})$$

Question-Solution Global Information

$$\mathrm{BIS}\left(\mathbf{t}^{(n)}\right) = \frac{(1-\alpha)\,\mathrm{ATTN}\left(\mathbf{q}|\mathbf{t}^{(n)}\right) + \alpha\mathrm{ATTN}\left(\mathbf{s}|\mathbf{t}^{(n)}\right)}{(1-\alpha)\,\mathrm{NLL}\left(\mathbf{q}|\mathbf{t}^{(n)}\right) + \alpha\mathrm{NLL}\left(\mathbf{s}|\mathbf{t}^{(n)}\right)}$$

Only a very few thinking steps have high BIS … How to sample them?



Path-Level A* Search for Compressing CoT

(1) Initialization

Given the thought sequeue $\mathcal{Q}$ sorted by BIS, the first step is dequeued from $\mathcal{Q}$ to form thinking span $\mathbf{r}^{(n)} = \langle \mathbf{t}^{(n-1)}, \mathbf{t}^{(n)}, \mathbf{t}^{(n+1)} \rangle$ as the root node of the search tree $\mathcal{T}$.

(2) Verification

A validation model $\mathcal{V}$ is introduced to determine whether the current thought path $\mathbf{t}'_k$ successfully leads to the solution $\mathbf{S}$:

$$\mathcal{V}(\mathbf{q}+\mathbf{t}'_k)\begin{cases}\neq \mathbf{s}, & \text{expand } \mathcal{T}\\ = \mathbf{s}, & \text{return } \mathbf{t}' = \mathbf{t}'_k\end{cases}$$

Notations

- question: $\mathbf{q}$
- solution: $\mathbf{S}$
- current thinking step: $\mathbf{t}'_k$
- thinking span: $\mathbf{r}^{(n)} = \langle \mathbf{t}^{(n-1)}, \mathbf{t}^{(n)}, \mathbf{t}^{(n+1)} \rangle$
- self-information: $\mathcal{I}(\mathbf{s}|\mathbf{q}, \mathbf{t}'_k) = -\frac{1}{|\mathbf{s}|}\log P_{\mathcal{V}}(\mathbf{s}|\mathbf{q}, \mathbf{t}'_k)$

(3) Exploration

If $\mathbf{t}'_k$ doesn't pass verification, each node in the first W spans $\{\mathbf{r}_1, \ldots, \mathbf{r}_W\}$ that dequeued from $\mathcal{Q}$ is appended to $\mathbf{t}'_k$, then select the node that with the minimal cost function $f(\cdot)$ in candidate thinking paths:

$$\hat{\mathbf{r}}_w = \underset{w\in\{1,\cdots,W\}}{\operatorname{argmin}} \; f(\mathbf{t}'_k + \mathbf{r}_w)$$

Current Cost Function G

$$g(\mathbf{t}'_k) = -\frac{\beta}{|\mathbf{t}'_k|}\log P_{\mathcal{V}}(\mathbf{t}'_k|\mathbf{q})$$

Future Cost Function H

$$h(\mathbf{t}'_k) = \mathcal{I}(\mathbf{s}|\mathbf{q}, \mathbf{t}'_k)$$

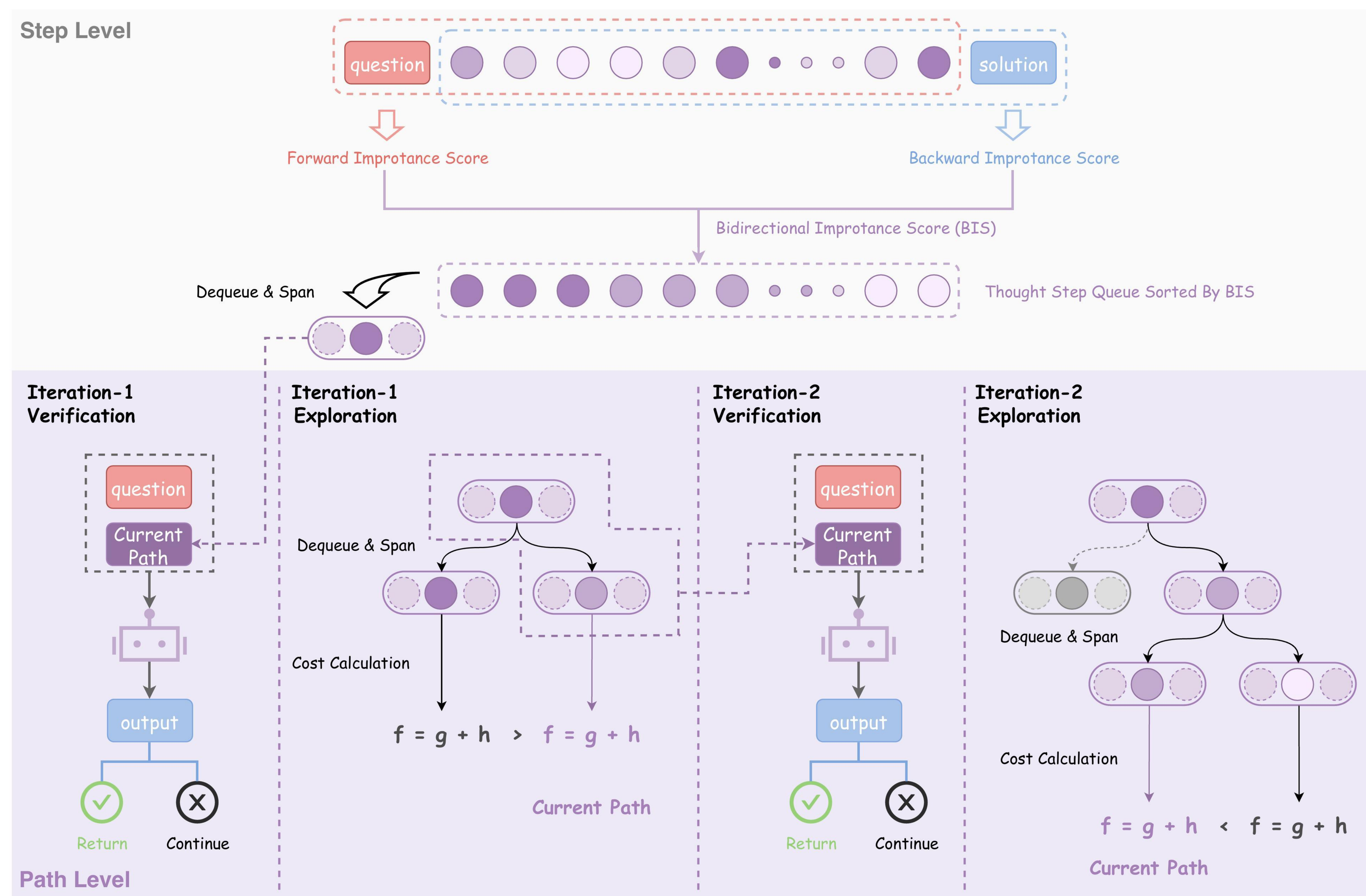
Cost Function F = G + H

$$f(\mathbf{t}'_k + \mathbf{r}_w) = g(\mathbf{t}'_k + \mathbf{r}_w) + h(\mathbf{t}'_k + \mathbf{r}_w)$$

A*-Thought Framework

1. How to identify high-quality and informative steps at the *step-level*?

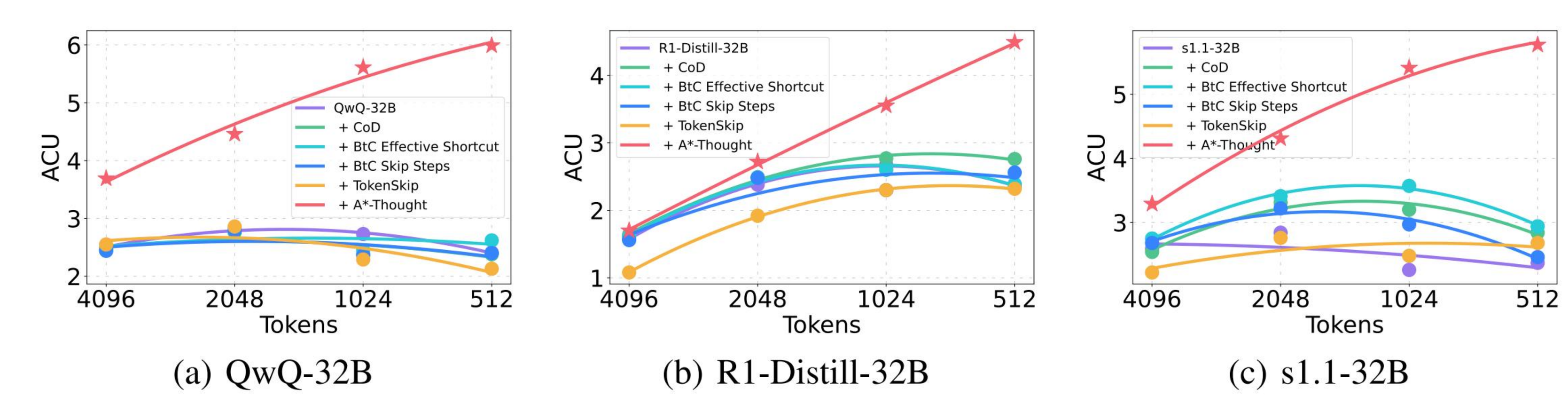
We design a ***step-level bidirectional importance score (BIS)*** to evaluate the criticality of individual sentences. This scoring mechanism serves to significantly enhance effectiveness of the A* search procedure compared to standard sampling.



2. How to effectively assemble individual thinking steps into a concise and effective reasoning path at the *path-level*?

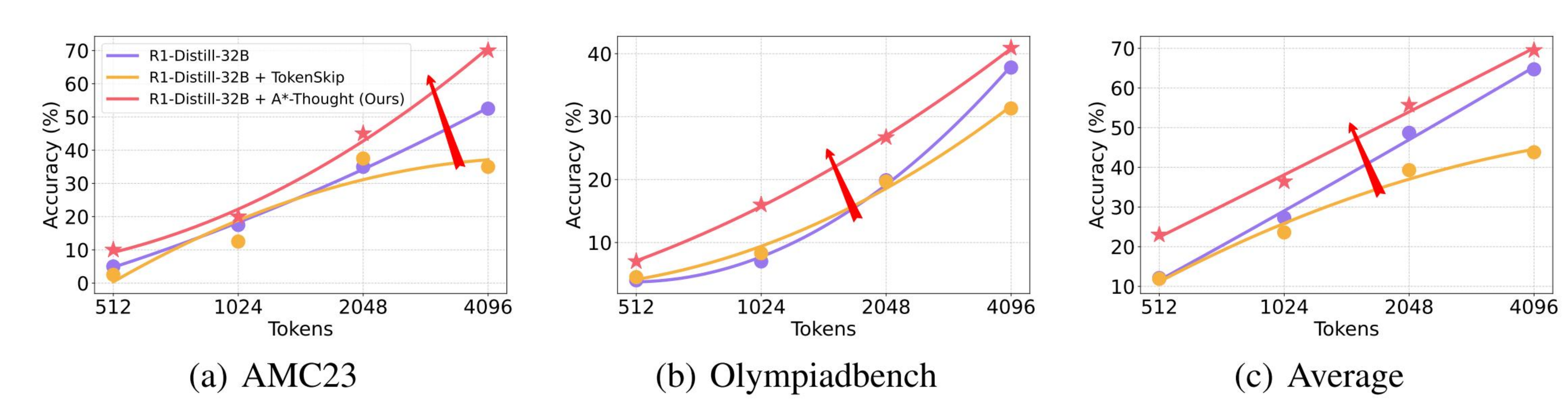
We introduce a ***path-level A* search algorithm*** tailored for compressing lengthy CoTs from LRMs. It strategically considers both current path quality and estimated future costs to optimize LRM performance under stringent output length constraints.

Efficiency & Performance



(a) QwQ-32B (b) R1-Distill-32B (c) s1.1-32B

Figure 4: ACU on different methods, which reflects performance-to-efficiency ratio of LRMs.



(a) AMC23 (b) Olympiadbench (c) Average

Figure 5: Performance of R1-Distill-32B augmented using TokenSkip and A*-Thought. "Average" denotes the average accuracy of the model in MATH500, AMC23, OlympiadBench, and GSM8K.

Results

Table 1: Experimental results of different long-to-short methods across several benchmarks. The best results are shown in **bold**, and the second-best results are underlined.

Methods	MATH500 Acc.(↑)	MATH500 Len.(↓)	AMC23 Acc.(↑)	AMC23 Len.(↓)	OlympiadBench Acc.(↑)	OlympiadBench Len.(↓)	GSM8K Acc.(↑)	GSM8K Len.(↓)	Average Acc.(↑)	Average Len.(↓)	ACU
Budget: 512 Tokens											
QwQ-32B	10.8	512.00	2.5	512.00	3.3	512.00	27.6	511.97	11.1	511.99	2.16
QwQ-32B w/ s1K-1.1	9.6	512.00	7.5	512.00	3.4	512.00	28.8	512.00	12.3	512.00	2.41
+ CoD	10.6	512.00	5.0	512.00	4.2	512.00	<u>29.0</u>	511.96	12.2	511.99	2.38
+ BtC Effective Shortcut	10.2	512.00	<u>12.5</u>	512.00	4.2	512.00	26.7	511.95	<u>13.4</u>	511.99	<u>2.62</u>
+ BtC Skip Steps	9.6	512.00	5.0	512.00	<u>5.6</u>	512.00	28.9	511.95	12.3	511.99	2.40
+ TokenSkip	<u>10.8</u>	511.05	2.5	512.00	3.9	512.00	26.4	508.11	10.9	510.79	2.13
+ A*-Thought	**33.2**	**491.92**	**15.0**	**508.60**	**12.0**	**509.74**	**57.4**	**451.76**	**29.4**	**490.51**	**5.99**
Budget: 1024 Tokens											
QwQ-32B	16.6	1016.85	15.0	1024.00	6.4	1023.93	49.1	951.96	21.8	1004.19	2.17
QwQ-32B w/ s1K-1.1	<u>24.8</u>	1023.52	<u>17.5</u>	1024.00	<u>8.9</u>	1023.94	60.1	999.80	<u>27.8</u>	1017.82	<u>2.73</u>
+ CoD	<u>24.8</u>	1023.37	5.0	1024.00	7.3	1023.64	60.1	996.84	24.3	1016.96	2.39
+ BtC Effective Shortcut	23.4	1022.88	7.5	1024.00	7.7	1023.92	<u>61.3</u>	1000.44	25.0	1017.81	2.45
+ BtC Skip Steps	23.4	1023.25	5.0	1024.00	7.6	1024.00	59.9	1000.93	24.0	1018.05	2.36
+ TokenSkip	22.4	995.96	12.5	1024.00	6.4	1019.61	49.7	934.74	22.8	993.58	2.29
+ A*-Thought	**50.8**	**858.28**	**37.5**	**928.25**	**22.3**	**954.74**	**81.9**	**688.69**	**48.1**	**857.49**	**5.61**
Budget: 2048 Tokens											
QwQ-32B	51.2	1844.96	25.0	1978.60	18.4	2021.95	80.4	1245.68	43.8	1772.80	2.47
QwQ-32B w/ s1K-1.1	60.0	1887.15	<u>35.0</u>	2000.95	23.3	2012.14	88.7	1474.00	51.8	1843.56	2.81
+ CoD	60.2	1894.54	30.0	2022.35	<u>25.5</u>	2018.02	89.5	1490.23	51.3	1856.29	2.76
+ BtC Effective Shortcut	<u>60.8</u>	1884.67	<u>35.0</u>	2004.65	23.7	2012.43	<u>89.8</u>	1473.25	<u>52.3</u>	1843.75	2.84
+ BtC Skip Steps	58.8	1884.96	<u>35.0</u>	2005.67	23.2	2013.05	89.2	1490.39	51.6	1848.52	2.79
+ TokenSkip	53.6	1685.34	<u>35.0</u>	1923.25	19.7	1943.68	86.7	1272.03	48.8	1706.08	<u>2.86</u>
+ A*-Thought	**69.2**	**1271.76**	**45.0**	**1540.30**	**30.3**	**1625.89**	**91.2**	**843.69**	**58.9**	**1320.41**	**4.46**
Budget: 4096 Tokens											
QwQ-32B	75.4	2798.67	55.0	3456.05	36.5	3645.22	85.8	1348.24	63.2	2812.05	2.25
QwQ-32B w/ s1K-1.1	79.6	2693.27	<u>65.0</u>	3485.95	**42.4**	3500.66	**95.2**	1624.11	**70.6**	2826.00	2.50
+ CoD	**80.2**	2719.00	60.0	3354.28	42.0	3488.67	<u>95.0</u>	1655.80	<u>69.3</u>	2804.44	2.47
+ BtC Effective Shortcut	79.6	2696.72	57.5	3355.43	<u>42.4</u>	3493.28	94.8	1636.45	68.6	2795.47	2.45
+ BtC Skip Steps	<u>80.2</u>	2710.83	57.5	3399.93	41.8	3494.41	94.9	1651.37	68.6	2814.14	2.44
+ TokenSkip	74.4	2336.29	52.5	3156.68	37.8	3289.44	94.8	1412.87	64.9	2548.82	<u>2.55</u>
+ A*-Thought	78.8	**1699.78**	**65.0**	**2385.85**	40.1	**2546.45**	93.1	**874.54**	<u>69.3</u>	**1876.66**	**3.69**

Analysis

Table 4: Amount of the training data and the corresponding time. ρ denotes the compression rate.

Training Set	ρ	Training Time (s) QwQ-32B	R1-Distill-32B	s1.1-32B
s1K-1.1	100.00%	13819.60	13784.32	13638.07
+ TokenSkip	85.41% (↓14.59%)	12846.14 (↓7.04%)	12700.96 (↓7.86%)	12679.52 (↓7.03%)
+ A*-Thought	31.31% (↓68.69%)	10468.20 (↓24.25%)	10809.50 (↓21.58%)	10550.03 (↓22.64%)

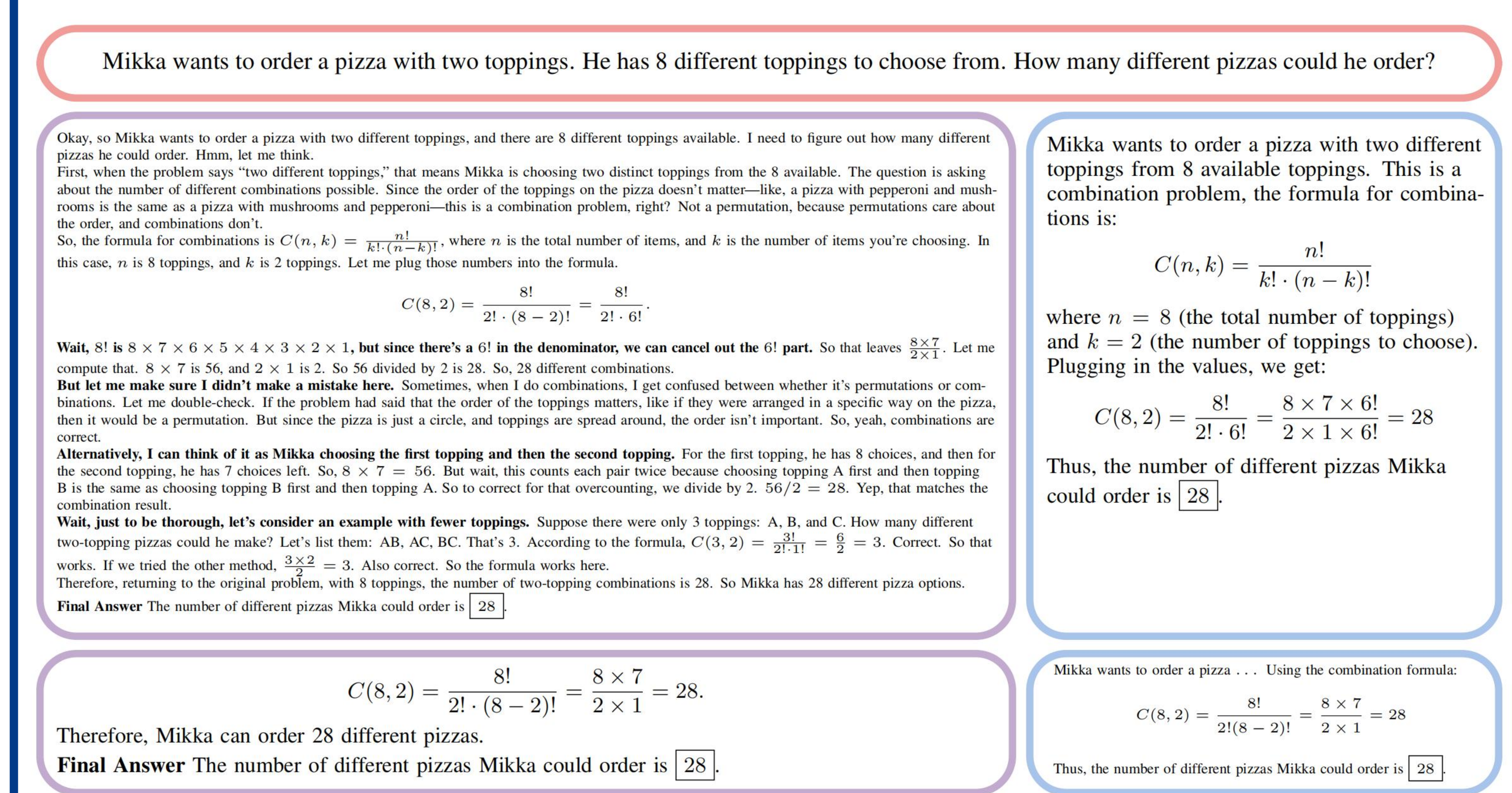


Figure 7: Responses generated by QwQ-32B models trained with and without A*-Thought. Red box represents the question, purple box represents the thinking process, blue box represents the solution.

Email for more questions: ***xuxiaoang@bupt.edu.cn***