

## 1011 A Prompts used in experiments

### 1012 A.1 Prompt for MCTS

1013 The prompt used for MCTS is shown in Table 6.

Table 6: Prompt used for VLM during MCTS procedure. We provide two examples of in-context learning to force VLM to follow the reasoning format.

#### MCTS Prompt Template:

Answer the question **\*\*step by step\*\*** and provide the final answer at the end, each step should end with **\*\*<end>\*\*** and put your final answer within  $\square$ . Below are two examples:  
Question: BoatsRUs built 7 canoes in January of this year and then each subsequent calendar month they built twice the number of canoes they had built the previous month. How many total canoes were built by BoatsRUs by the end of May of this year?

### Step1: To find the result of the total number of canoes built by BoatsRUs by the end of May, I need to find the number of canoes built in each month from January to May and then add them up. <end>

### Step2: To find the number of canoes built in each month, I need to use the formula for the number of canoes built in a given month, which is the number of canoes built in the previous month times 2. <end>

### Step3: So, the number of canoes built in January is 7, the number of canoes built in February is 7 times 2, which is 14, the number of canoes built in March is 14 times 2, which is 28, the number of canoes built in April is 28 times 2, which is 56, and the number of canoes built in May is 56 times 2, which is 112. <end>

### Step4: Now, I can add up these numbers to get the total number of canoes built by BoatsRUs by the end of May: 7 plus 14 plus 28 plus 56 plus 112, which is 217. <end>

### Final Answer: The answer is:  $\square$  217.

Question: Find the number of blue circles in the figure.

### Step 1: To find the result of the number of blue circles, I need to interpret the figure. The figure is a Venn diagram with two labeled sets: - One set labeled "blue" contains all the shapes that are blue in color. - The other set labeled "circle" contains all the shapes that are circular in shape. The overlapping region of the Venn diagram contains shapes that are both blue and circular. <end>

### Step 2: The overlapping region contains shapes that meet both criteria: Blue color and Circle shape. From the diagram: - There is **\*\*one blue circle\*\*** in the overlapping region. <end>

### Final Answer: The answer is:  $\square$  1.

Remember to answer the question **\*\*step by step\*\***! Here is your question:

Question: {QUESTION}

### 1014 A.2 Prompt for Critic Model

1015 The prompt used for critic model during MCTS is shown in Table 7.

### 1016 A.3 Prompt for RFT

1017 The prompt used for RFT is shown in Table 8.

## 1018 B More experiments

### 1019 B.1 Reward curves of VLM with different training data

1020 We compare the reward curves during RFT of ThinkLite-VL-Random11k, ThinkLite-VL-Fullset,  
1021 ThinkLite-VL-Iter5Only, and ThinkLite-VL, as shown in Figure 5. Although ThinkLite-VL-

Table 7: Critic prompt for MCTS simulation results evaluation.

<p><b>Critic Prompt Template:</b></p> <p>Please help me judge the correctness of the generated answer and the corresponding rationale.</p> <p>Question: {}</p> <p>Ground truth answer: {}</p> <p>Generated rationale and answer: {}</p> <p>Your output should only be one sentence: the generated answer is true or false.</p>
--

Table 8: Prompt template used for reinforcement learning fine-tuning.

<p><b>Prompt Template:</b></p> <p>You FIRST think about the reasoning process as an internal monologue and then provide the final answer. The reasoning process MUST BE enclosed within &lt;think&gt; &lt;/think&gt; tags. The final answer MUST BE put in <input type="checkbox"/>.</p>
--

1022 Random11k and ThinkLite-VL-Fullset achieve higher rewards during training, their actual benchmark  
1023 performances are inferior to ThinkLite-VL. This observation suggests that incorporating a large  
1024 number of easy samples into training rapidly improves rewards but fails to enhance the model’s  
1025 reasoning ability. Moreover, ThinkLite-VL exhibits notably lower rewards compared to ThinkLite-  
1026 VL-Iter5Only, indicating that the unsolved data identified by our MCTS-based sample selection  
1027 strategy indeed pose significant challenges to the VLM. By progressively learning to solve these chal-  
1028 lenging problems during training—even if not all are solved completely—the reasoning capabilities  
1029 of VLMs can be substantially improved.

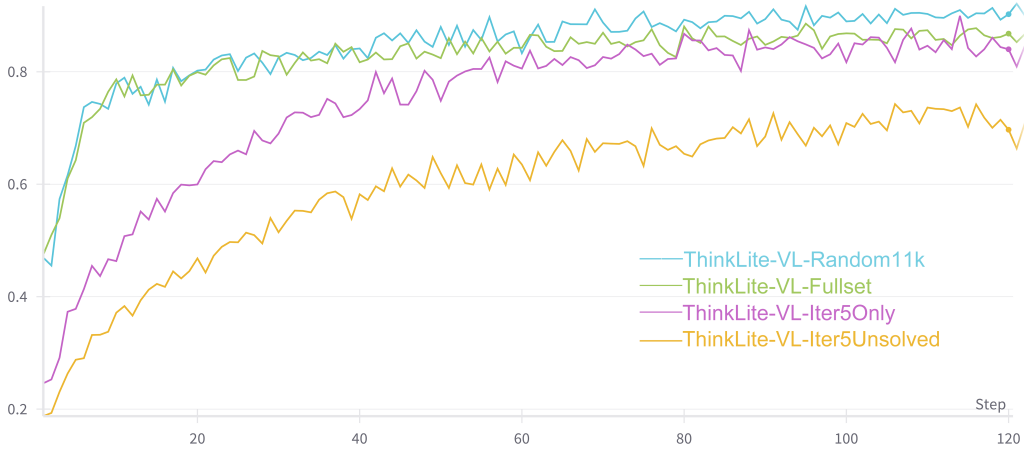


Figure 5: Comparison of reward curves of 7B models trained with different data during RFT. Iter5+Unsolved 11k dataset presents the most challenging learning setting for VLM, highlighting the difficulty of the samples selected by MCTS-based sample selection.

## 1030 B.2 Ablation Study of Data Difficulty

1031 In this section, we investigate how training data difficulty affects model performance. We present the  
1032 average performance of models trained using different difficulty data in Table 9. Notably, the model

1033 trained with the Iter5+Unsolved subset achieves the highest average score of 63.89, outperforming all  
 1034 other settings. When expanding the difficulty threshold (e.g., Iter10, Iter20, Iter30, and Iter40), the  
 1035 model performance consistently declines, suggesting that medium-difficulty samples are important  
 1036 for improving model reasoning ability. As the difficulty of the training data decreases, the model’s  
 1037 performance also declines. This trend suggests that the inclusion of an excessive number of easy  
 1038 samples may weaken the training signal during RFT and ultimately hurt the model’s reasoning ability.

Table 9: ThinkLite-VL-7B performance under different training data difficulty settings. Iter5+Unsolved achieves the best performance.

Difficulty level	Data size	Avg. score
Fullset	70k	63.13
Iter1+Unsolved	18k	63.29
Iter5+Unsolved	11k	63.89
Iter10+Unsolved	8k	62.65
Iter20+Unsolved	6.8k	62.61
Iter30+Unsolved	6.1k	62.39
Iter40+Unsolved	5.8k	62.26
Unsolved	5.6k	62.04

## 1039 C Case Studies

1040 In this section, we present samples of varying difficulty levels selected by the MCTS-based sample  
 1041 selection method across different datasets, as shown in Tables 15 through 14. The difficulty levels are  
 1042 determined based on the number of reasoning iterations required by the VLM to arrive at the correct  
 1043 answer during the MCTS process, providing reference examples for understanding how the method  
 1044 distinguishes between easy and challenging samples.

## 1045 D Limitations

1046 While sample selection has effectively enhanced the reasoning capabilities of vision-language models  
 1047 (VLMs), the overall training efficiency remains a key limitation, both in data filtering and rein-  
 1048 forcement learning stages. Common approaches such as self-consistency-based selection and our  
 1049 proposed MCTS-based strategy require substantial time for sample filtering. Additionally, the GRPO  
 1050 training process incurs significant computational overhead due to the large number of rollout samples  
 1051 needed for target value estimation. These efficiency challenges can potentially be mitigated through  
 1052 parallelized sample generation and processing, which we leave as an avenue for future work.

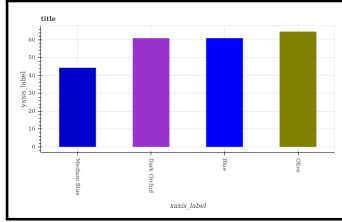
## 1053 E Societal impacts

1054 ThinkLite-VL can positively impact society by enabling more data-efficient and accessible develop-  
 1055 ment of advanced visual reasoning in vision-language models (VLMs). However, negative societal  
 1056 risks include the potential misuse of its MCTS-based sample selection insights for crafting sophis-  
 1057 ticated disinformation. Responsible development must also guard against over-reliance on these  
 1058 enhanced, yet still fallible, reasoning systems.

---

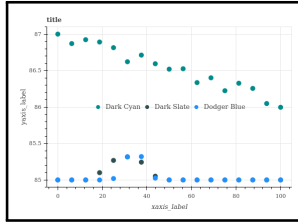
**Example 3: Different difficulty samples from FigureQA**

---



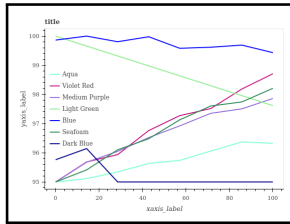
Iter0 **Question:** Is Medium Blue less than Dark Orchid?  
**Ground Truth Answer:** Yes.

---



Iter29 **Question:** Does Dodger Blue intersect Dark Slate?  
**Ground Truth Answer:** Yes.

---



Unsolved **Question:** Does Violet Red have the maximum area under the curve?  
**Ground Truth Answer:** No.

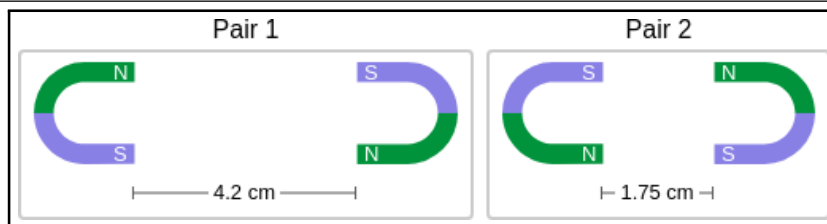
---

Table 10: Example of samples with different difficulties decided by MCTS-based sample selection from FigureQA.

---

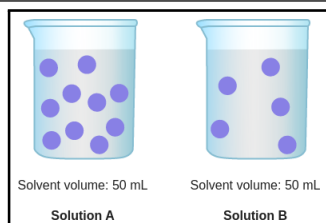
**Example 4: Different difficulty samples from ScienceQA**

---



Iter0 **Question:** Think about the magnetic force between the magnets in each pair. Which of the following statements is true? Choices: (A) The magnitude of the magnetic force is greater in Pair 2. (B) The magnitude of the magnetic force is greater in Pair 1. (C) The magnitude of the magnetic force is the same in both pairs.  
**Ground Truth Answer:** A.

---



Iter13 **Question:** Which solution has a higher concentration of purple particles? Choices: (A) neither; their concentrations are the same (B) Solution A (C) Solution B  
**Ground Truth Answer:** B.

---



Unsolved **Question:** What is the direction of this push? Choices: (A) away from the hockey stick (B) toward the hockey stick  
**Ground Truth Answer:** A.

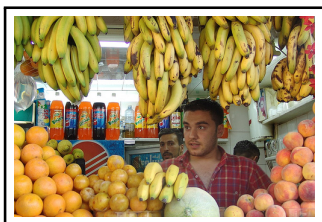
---

Table 11: Example of samples with different difficulties decided by MCTS-based sample selection from ScienceQA.

---

**Example 5: Different difficulty samples from OK-VQA**

---



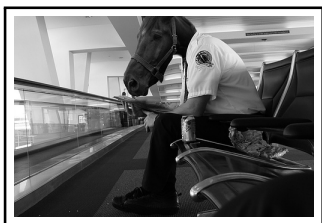
Iter0 **Question:** What food group is pictured here?  
**Ground Truth Answer:** fruit.

---



Iter20 **Question:** What is the length of the surfboard the man in the black shorts at the back of the line of people is holding?  
**Ground Truth Answer:** 7 feet.

---



Unsolved **Question:** What is this guy's profession?  
**Ground Truth Answer:** security.

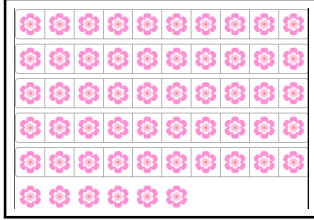
---

Table 12: Example of samples with different difficulties decided by MCTS-based sample selection from OK-VQA.

---

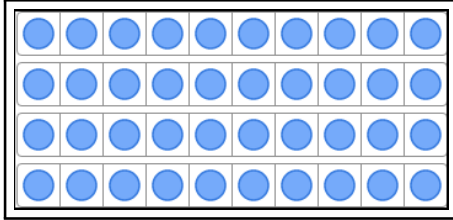
**Example 6: Different difficulty samples from IconQA**

---



Iter0 **Question:** How many flowers are there?  
**Ground Truth Answer:** 56.

---



Iter10 **Question:** How many dots are there?  
**Ground Truth Answer:** 40.

---



Unsolved **Question:** How many stars are there?  
**Ground Truth Answer:** 19.

---

Table 13: Example of samples with different difficulties decided by MCTS-based sample selection from IconQA.

---

**Example 7: Different difficulty samples from TabMWP**

---

red confetti	\$11 per pound
gold confetti	\$12 per pound
rainbow confetti	\$10 per pound
silver confetti	\$12 per pound
green confetti	\$12 per pound

Iter0 **Question:** Adriana wants to buy 3 pounds of silver confetti. How much will she spend?  
**Ground Truth Answer:** 36.

---

Spinning a wheel numbered 1 through 5	
Number spun	Frequency
1	2
2	9
3	4
4	11
5	3

Iter22 **Question:** A game show viewer monitors how often a wheel numbered 1 through 5 stops at each number. How many people are there in all?  
**Ground Truth Answer:** 29.

---

Ties per rack	
Stem	Leaf
3	2 5 6 8 9
4	0 4 6 8 8 8
5	1 4
6	5 8
7	5 6 7 9 9

Unsolved **Question:** The employee at the department store counted the number of ties on each tie rack. How many racks have at least 30 ties but fewer than 70 ties?  
**Ground Truth Answer:** 15.

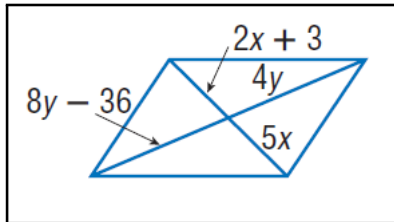
---

Table 14: Example of samples with different difficulties decided by MCTS-based sample selection from TabMWP.

---

**Example 1: Different difficulty samples from Geometry3K**

---

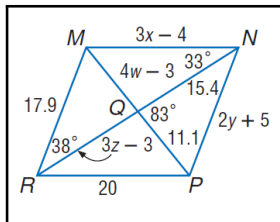


Iter0

**Question:** Find  $y$  so that the quadrilateral is a parallelogram.

**Ground Truth Answer:** 9.

---

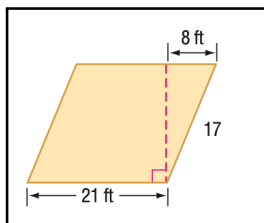


Iter16

**Question:** Use parallelogram  $MNPQ$  to find  $y$ .

**Ground Truth Answer:** 6.45.

---



Unsolved

**Question:** Find the area of the parallelogram. Round to the nearest tenth if necessary.

**Ground Truth Answer:** 315.

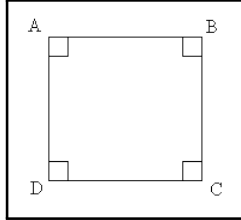
---

Table 15: Example of samples with different difficulties decided by MCTS-based sample selection from GeoQA.

---

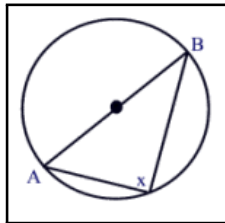
**Example 2: Different difficulty samples from Geos**

---



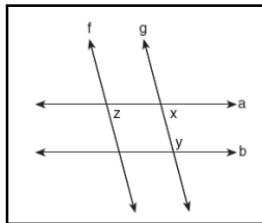
Iter0 **Question:** What is the area of the following square, if the length of BD is  $2 * \sqrt{2}$ ? Choices: (A) 1 (B) 2 (C) 3 (D) 4 (E) 5.  
**Ground Truth Answer:** D.

---



Iter7 **Question:** Given the circle at the right with diameter AB, find x. Choices: (A) 30 degrees (B) 45 degrees (C) 60 degrees (D) 90 degrees (E) None  
**Ground Truth Answer:** D.

---



Unsolved **Question:** In the diagram at the right, lines f and g are parallel, and lines a and b are parallel.  $x = 75$ . What is the value of  $y + z$ ? Choices: (A) 75 (B) 105 (C) 150 (D) 180 (E) None  
**Ground Truth Answer:** D.

---

Table 16: Example of samples with different difficulties decided by MCTS-based sample selection from Geos.