

Input:

I want you to act as an arithmetic reasoning expert for Chinese.

Request: 罗杰有5个网球。他又买了2罐网球。每罐有3个网球。他现在有多少个网球?

You should retell the request in English.

You should do step-by-step answer to obtain a number answer.

You should step-by-step answer the request.

You should tell me the answer in this format 'Answer :'.

Retelling the request in English: Roger currently has 5 tennis balls...

Step-by-step answer:

1. Roger has 5 tennis balls.

2. He buys 2 cans of tennis balls.

3. Each can has 3 tennis balls, so 2 cans will have a total of 6 tennis ...

4. To find out how many tennis balls Roger has in total, we need to add the ...

5. Adding 5 (initial tennis balls) and 6 (purchased tennis balls) gives ...

Answer: 11.

... (the remaining 4 examples)

I want you to act as an arithmetic reasoning expert for Chinese.

Request: 制作一件袍子需要 2 匹蓝色纤维布料和这个数量一半的白色纤维布料。它一共
需要用掉多少匹布料?

You should retell the request in English.

You should do step-by-step answer to obtain a number answer.

You should step-by-step answer the request.

You should tell me the answer in this format 'Answer :'.

text-davinci-003 output:

Request: To make a robe, two pieces of blue fabric and half the amount of white fabric are needed. How many pieces of fabric are needed in total?
Step-by-step answer:

1. Two pieces of blue fabric are needed.
2. Half the amount of blue fabric is needed for white fabric, so that is 1 piece of white fabric ($2 \times 0.5 = 1$).
3. To find out how many pieces of fabric are needed in total, we need to add the number of blue fabric and white fabric.
4. Adding 2 (blue fabric) and 1 (white fabric) gives us a total of 3 pieces of fabric.

Answer: 3 pieces of fabric.

gpt-3.5-turbo output:

Request: Making one robe requires 2 pieces of blue fiber fabric and half that amount of white fiber fabric. How many pieces of fabric are needed in total?
Step-by-step answer:
1. To make one robe, we need 2 pieces of blue fiber fabric.
2. We also need half that amount of white fiber fabric, which is 1 piece ($2/2 = 1$).
3. Therefore, to make one robe, we need a total of $2 + 1 = 3$ pieces of fiber fabric.
Answer: We need 3 pieces of fiber fabric in total to make one robe.

Answer:

3

Figure 8: A Chinese example of the MGSM benchmark using XLT and the corresponding outputs under the few-shot setting.

Input:

Here is a premise: Ese oli mullikilesse mässitud. What is the cause? Help me pick the more plausible option: -choice1: See oli õrn, -choice2: See oli väike.

text-davinci-003 output:

Choice 2 : See oli väike.

gpt-3.5-turbo output:

As an AI language model, I cannot determine the cause of the premise without additional context. The options provided do not seem to be directly related to the premise. Can you please provide more information or context?

Answer:

1

Figure 9: An Estonian example of the **XCOPA** benchmark using **basic prompt** and the corresponding outputs under the **zero-shot** setting.

Input:

I want you to act as a commonsense reasoning expert for Estonian .

Here is a premise: Ese oli mullikilesse mässitud. What is the cause? Help me pick the more plausible option: -choice1: See oli õrn, -choice2: See oli väike.

You should retell the premise and the options in English.

You should do step-by-step answer to pick a choice .

You should step-by-step answer the request.

You should tell me the choice number in this format ' Choice number : '.

text-davinci-003 output:

Premise: Ese oli mullikilesse mässitud.
Options: Choice 1: See oli õrn, Choice 2: See oli väike.
Step 1: Consider the context of the premise. What kind of object was mässitud (inserted) into the pocket?
Step 2: Consider the two options. Choice 1: See oli õrn (It was soft) and Choice 2: See oli väike (It was small).
Step 3: Based on the context and the two options, the more plausible choice is Choice 2: See oli väike (It was small).

Choice number: 2

gpt-3.5-turbo output:

Premise: "Ese oli mullikilesse mässitud" (The object was wrapped in bubble wrap).
Options:
1. See oli õrn (It was fragile).
2. See oli väike (It was small).
Step 1: Based on the premise, we know that the object was wrapped in bubble wrap. It does not provide any information about the size or fragility of the item.
Step 2: However, common sense would suggest that fragile items are more likely to be wrapped in bubble wrap for protection during transportation or storage.
Therefore, the more plausible option is choice 1: "See oli õrn" (It was fragile).

Choice number: 1

Answer:

1

Figure 10: An Estonian example of the **XCOPA** benchmark using **XLT** and the corresponding outputs under the **zero-shot** setting.

Input:

I want you to act as a commonsense reasoning expert for Estonian.

Here is a premise: Mees jäi vanaks. What is the effect? Help me pick the more plausible option:

-choice1: Ta juuksed muutusid halliks, -choice2: Ta müüs oma asjad maha.

You should retell the premise and the options in English.

You should do step-by-step answer to pick a choice.

You should step-by-step answer the request.

You should tell me the choice number in this format 'Choice number:'.

Premise: A man turned on the tap.

Options: 1. The toilet filled with water. 2. Water flowed from the tap.

Step-by-step answer:

1. The premise states that a man turned on the tap, which implies that water started flowing.
2. The first option suggests that the toilet filled with water, which is possible if the tap ...
3. The second option suggests that water flowed from the tap, which is the more plausible ...

Choice Number: 2. Water flowed from the tap. The answer is choice2."

... (the remaining 4 examples)

I want you to act as a commonsense reasoning expert for Estonian.

Here is a premise: Ese oli mullikilesse mässitud. What is the cause? Help me pick the more

plausible option: -choice1: See oli õrn, -choice2: See oli väike.

You should retell the premise and the options in English.

You should do step-by-step answer to pick a choice.

You should step-by-step answer the request.

You should tell me the choice number in this format 'Choice Number:'.

text-davinci-003 output:

Premise: The object was wedged into a crack.

Options:

1. It was delicate.
2. It was small.

Step-by-step answer:

1. The premise states that the object was wedged into a crack.
2. Option 1 suggests that the object was delicate, which is not necessarily related to the object being wedged into a crack.
3. Option 2 suggests that the object was small, which is more plausible as a small object is more likely to fit into a crack.

Choice Number: 2. It was small. The answer is choice2.

gpt-3.5-turbo output:

Premise: "The object was wrapped in bubble wrap."

Options:

1. It was fragile.
2. It was small.

Step-by-step answer:

1. The premise states that the object was wrapped in bubble wrap, which is often used to protect fragile items during transport.
2. Bubble wrap is not typically used to protect small items, so option 1 is more plausible.
3. Therefore, the more plausible option is Choice 1: It was fragile.

Choice Number: 1. It was fragile. The answer is choice1.

Answer:

1

Figure 11: An Estonian example of the XCOPA benchmark using XLT and the corresponding outputs under the few-shot setting.