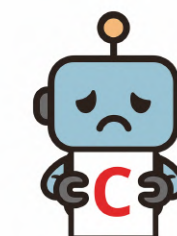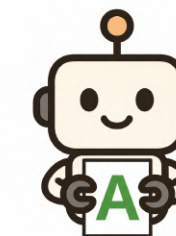## Evaluation 1: VQA

Select the most suitable option to fill in the blank so that a certain pattern is shown.

Evaluation 2: Chain-of-thought Error Detection

Groundtruth Chain-of-thought: Step 1..., Step 2..., Step 3..., Step 4..., Hence, the correct answer is A.

Randomly pick a step then randomly inject a type of error

Magic magic, make the chosen step and following steps erroneous!

Corrupted Chain-of-thought: Step 1..., Step 2..., Step 3..., Step 4..., Hence, the correct answer is C.

Model model, since which step does the reasoning start to go wrong?