

Technical Appendix

A Data

A.1 Data Sources

Problem Set	Benchmark	Domain	Questions	Source
ARC Challenge Test	ARC	Science	1,173	[Clark <i>et al.</i> , 2018]
AQUA-RAT	AGI Eval	Math	254	[Zhong <i>et al.</i> , 2023]
Hellaswag Val	Hellaswag	Common Sense Reasoning	10,042	[Zellers <i>et al.</i> , 2019]
LogiQA (English)	AGI Eval	Logic	651	[Liu <i>et al.</i> , 2020]
LSAT-AR	AGI Eval	Law (Analytic Reasoning)	230	[Wang <i>et al.</i> , 2021]
LSAT-LR	AGI Eval	Law (Logical Reasoning)	510	[Wang <i>et al.</i> , 2021]
LSAT-RC	AGI Eval	Law (Reading Comprehension)	260	[Wang <i>et al.</i> , 2021]
MedMCQA Valid	MedMCQA	Medicine	6,150	[Pal <i>et al.</i> , 2022]
SAT-English	AGI Eval	English	206	[Zhong <i>et al.</i> , 2023]
SAT-Math	AGI Eval	Math	220	[Zhong <i>et al.</i> , 2023]

Table 1: Source of problem sets used to create the MCQA test set

A.2 Sample MCQA Problem

```
{
  "id": 3,
  "source": "agi-eval/aqua-rat",
  "source_id": 35,
  "topic": "Math",
  "context": "",
  "question": "A rectangular solid, 3 x 4 x 15, is inscribed in a sphere,
              so that all eight of its vertices are on the sphere.
              What is the diameter of the sphere?",
  "choices": {
    "A": " 13.3542",
    "B": " 15.8113",
    "C": " 18.3451",
    "D": " 19.5667",
    "E": " 20.8888"},
  "answer": "B",
  "solution": "In an inscribed rectangle in a sphere, we will have a line joining
              the opposite vertices as the diameter. According to the Pythagoras theorem,
              sides 3, 4 give diagonal as 5 ==> with 5 and 15, we get 5sqrt(10).
              5sqrt(10) or 15.8113 is the diameter of the sphere.\nanswer = B"
}
```

Figure 1: Sample of a multiple-choice question in standardized data format – with whitespace added for readability.

B Prompts

B.1 Answer-Only Prompt

```
[System Prompt]
You are an intelligent assistant.
Your task is to answer the following multiple-choice questions.
You MUST answer the question using the following format 'Action: Answer("[choice]")'
The parameter [choice] is the letter or number of the answer you want to select (e.g. "A", "B", "C",
or "D").
For example, 'Answer("C")' will select choice "C" as the best answer.
You MUST select one of the available choices; the answer CANNOT be "None of the Above".

[Example Problem]
Question: What is the capital of the state where Johns Hopkins University is located?
Choices:
  A: Baltimore
  B: Annapolis
  C: Des Moines
  D: Las Vegas

[Example Solution]
Action: Answer("B")
```

Figure 2 - Sample of the answer-only system prompt and one-shot example

B.2 Verbose CoT Prompt

```
[System Prompt]
You are an intelligent assistant.
Your task is to answer the following multiple-choice questions.
Think step-by-step through the problem to ensure you have the correct answer.
Then, answer the question using the following format 'Action: Answer("[choice]")'
The parameter [choice] is the letter or number of the answer you want to select (e.g. "A", "B",
"C", or "D").
For example, 'Answer("C")' will select choice "C" as the best answer.
You MUST select one of the available choices; the answer CANNOT be "None of the Above".

[Example Problem]
Question: What is the capital of the state where Johns Hopkins University is located?
Choices:
  A: Baltimore
  B: Annapolis
  C: Des Moines
  D: Las Vegas

[Example Solution]
Thought:
  Johns Hopkins University is located in Baltimore.
  Baltimore is a city located in the State of Maryland.
  The capital of Maryland is Annapolis.
  Therefore, the capital of the state where Johns Hopkins University is located is Annapolis.
  The answer is B: Annapolis.
Action: Answer("B")
```

Figure 3 - Sample of standard (i.e., verbose) CoT system prompt and one-shot example

B.3 Concise CoT Prompt

```
[System Prompt]
You are an intelligent assistant.
Your task is to answer the following multiple-choice questions.
Think step-by-step through the problem to ensure you have the correct answer.
Then, answer the question using the following format 'Action: Answer("[choice]")'
The parameter [choice] is the letter or number of the answer you want to select (e.g. "A", "B", "C",
or "D").
For example, 'Answer("C")' will select choice "C" as the best answer.
You MUST select one of the available choices; the answer CANNOT be "None of the Above".
Be concise.

[Example Problem]
Question: What is the capital of the state where Johns Hopkins University is located?
Choices:
  A: Baltimore
  B: Annapolis
  C: Des Moines
  D: Las Vegas

[Example Solution]
Thought:
  Johns Hopkins University is located in Baltimore, Maryland.
  The capital of Maryland is Annapolis.
Action: Answer("B")
```

Figure 4 - Sample of CCoT system prompt and one-shot example