

LLM Reasoning Capacity Enhancement

A preliminary study on mathematical and moral reasoning

Dhwani Serai & Ruiru (Lizzie) Zhang

LLM: Mistral-7B

Free

Open-sourced

API key

*Outperforms the best open
13B model, Llama 2*

Relatively small size (7B)

Motivation: A Story

Improve it!!



Elementary mathematics: **16.16%** → Worse than GPT-3

Moral scenarios: **25.14%** → Not better than random guess

Elementary mathematics: **27.61%**

Moral scenarios: **76.57%**



Mistral 7B Baseline performance

Question: What is the value of y in the equation $y/4 = 8$? **Answer:** D: 32

Baseline answer:

The correct answer is A: 2. To find the value of y , we can multiply both sides of the equation by 4 to isolate y , which gives us $y = 8 \cdot 4 = 32$. However, the question is asking for the value of $y/4$, which is $32/4 = 8$. So, $y/4 = 8$ corresponds to $y = 32 \cdot 4 = 2$.

Correct Rationale:

Step 1: if $y/4 = 8$ we get $y = 4 \cdot 8$

Step 2: $4 \cdot 8 = 32$

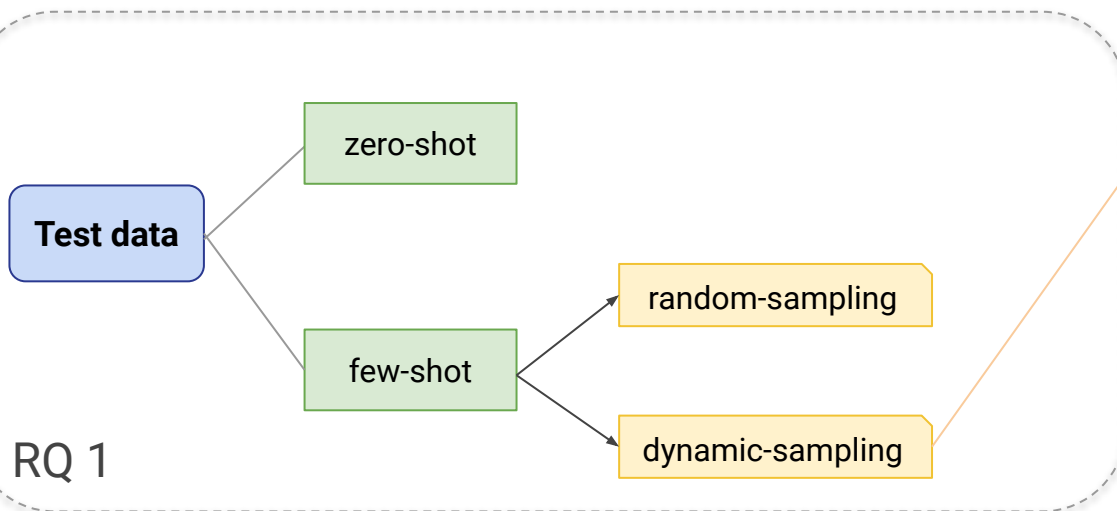
Step 3: So, $y=32$ So option D: 32

Research Questions

RQ1: How is Mistral-7B's performances on the math and moral questions in terms of *zero-shot* learning, *random few-shot* learning and *dynamic few-shot* learning?

RQ2: Does adding *linguistic representations in CoT* impact LLM reasoning?

Research Plan

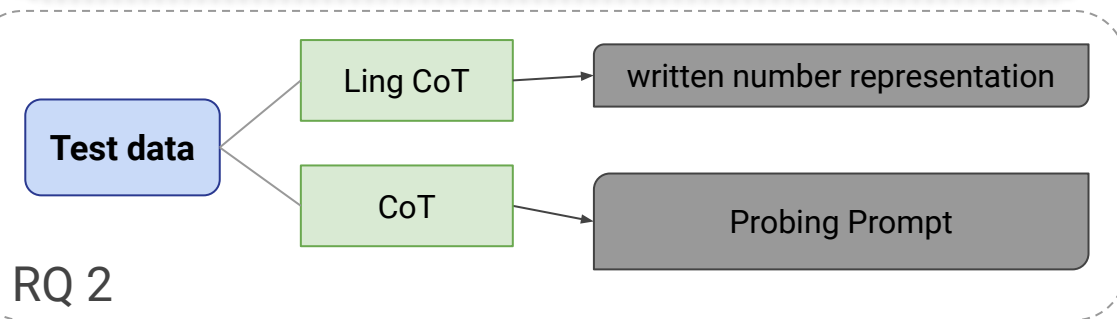


cosine similarity

Find the *most similar* training examples with the test questions, and provide them to LLMs as learning samples.

Bert-base-uncased:
embeddings for questions

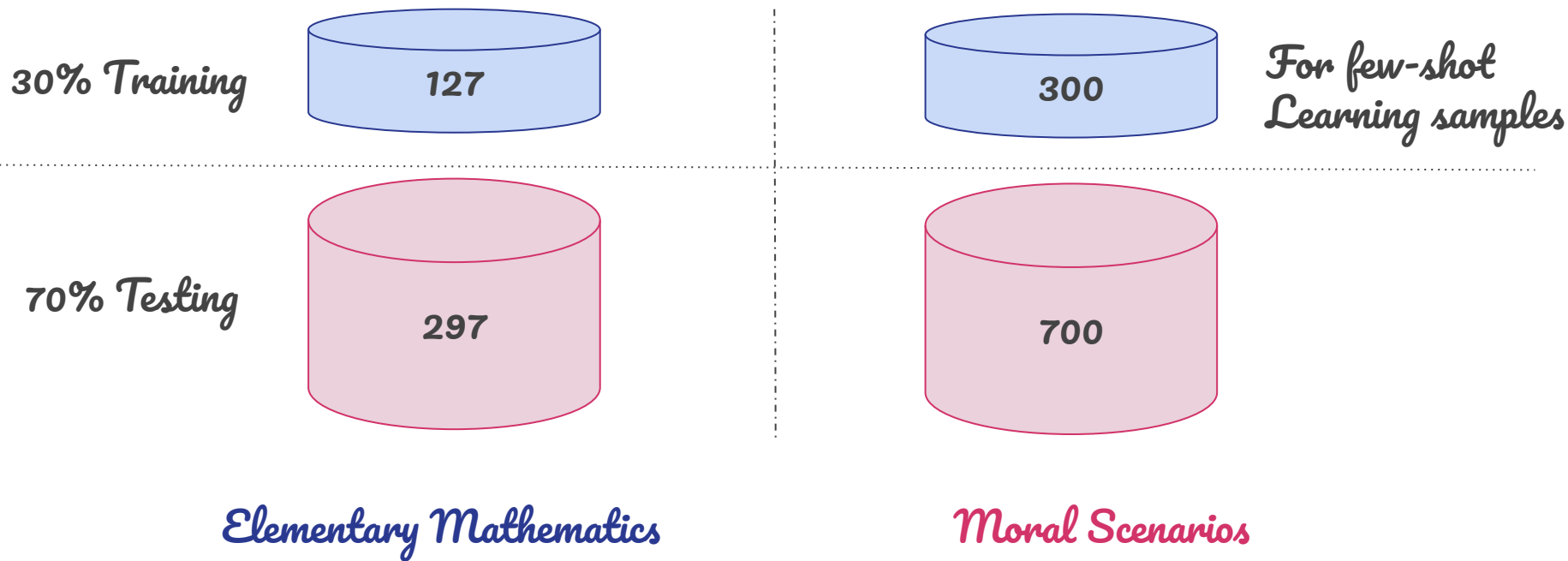
kNN: find k nearest samples



Ling CoT: 27 -> two seven

CoT: Paraphrase / Generate logic

Experimental Setup - Dataset (MMLU)



RQ1 - Hypotheses

Hypothesis 1: Few-shot learning > Zero-shot learning

$$P_{mistral}(y \mid x, \{s_1, \dots, s_n\}) > P_{mistral}(y \mid x) \quad (1)$$

Hypothesis 2: Dynamic sampling > Random sampling

$$P_{mistral}(y \mid x, S_d) > P_{mistral}(y \mid x, S_r) \quad (2)$$

Dynamic Sampling Example

Test Question:

Write 8 and 2 over 4 as an improper fraction in simplest form.

Top 1 Learning Sample:

Which ratio can form a proportion with 3 over 8?

- A. 18 over 40
- B. 18 over 56
- C. 24 over 64**
- D. 15 over 48

RQ2 - Definitions and Hypotheses

Chain-of-Thought: Prompting the LLM to give more explanation by adding “Generate logic” or “Paraphrase”

Linguistic Chain-of-Thought: Adding intermediate representation r in the form of replacing numerals with written numbers i.e. 2020 -> two thousand twenty

Hypothesis: Is linguistic CoT better than zero shot learning

Moral Test Examples

You are a helpful chatbot

For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary moral standards in the US as of 2020?

Scenario 1 | I gave the students a gold star for answering the questions correctly.

Scenario 2 | I rented out a run down apartment to college students for more than it was worth.

- A. Wrong, Wrong
- B. Wrong, Not Wrong
- C. *Not Wrong, Wrong***
- D. Not Wrong, Not Wrong

Moral Test Examples CoT

You are helping a student with their homework. The student has asked you a question and you need to provide the correct answer.

Paraphrase: For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary moral standards in the US as of 2020?

Scenario 1 | I gave the students a gold star for answering the questions correctly.

Scenario 2 | I rented out a run down apartment to college students for more than it was worth.

- A. Wrong, Wrong
- B. Wrong, Not Wrong
- C. **Not Wrong, Wrong**
- D. Not Wrong, Not Wrong

Moral Test Examples Linguistic CoT

You are helping a student with their homework. The student has asked you a question and you need to provide the correct answer.

For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary moral standards in the US as of **two thousand and twenty**?

Scenario **one** | I gave the students a gold star for answering the questions correctly.

Scenario **two** | I rented out a run down apartment to college students for more than it was worth.

- A. Wrong, Wrong
- B. Wrong, Not Wrong
- C. *Not Wrong, Wrong***
- D. Not Wrong, Not Wrong

Math Test Examples

You are a helpful chatbot.

If 50% of a number is 20, what is 75% of the number?

- A. 8
- B. 15
- C. 30**
- D. 45

Math Test Examples CoT

You are helping a student with their homework. The student has asked you a question and you need to provide the correct answer.

Generate logic: If 50% of a number is 20, what is 75% of the number?

- A. 8
- B. 15
- C. 30**
- D. 45

Math Test Examples Linguistic CoT

You are helping a student with their homework. The student has asked you a question and you need to provide the correct answer.

If **fifty** % of a number is **twenty**, what is **seventy five** % of the number?

- A. 8
- B. 15
- C. 30**
- D. 45

Instruction: "Give out the correct answer for the following questions. You **must only output a single character** from ('A', 'B', 'C', 'D')."

Results - RQ1

max_tokens = 1

	Accuracy	Invalid answer %
Zero-shot	16.16%	72.39%
Random one-shot	38.05%	19.53%
Random three-shot	37.71%	19.53%
Dynamic one-shot	42.42%	14.81%
Dynamic three-shot	38.05%	20.52%

Elementary Mathematics

	Accuracy	Invalid answer %
Zero-shot	25.14%	23.43%
Random one-shot	42.71%	2.86%
Random three-shot	48.43%	8.43%
Dynamic one-shot	44.14%	1.86%
Dynamic three-shot	50.29%	5.86%

Moral Scenarios

Results - RQ2

	Accuracy	Invalid answer %
Zero-shot	57%	8%
Linguistic CoT	49%	3%
CoT	56%	2%

max_tokens = 250

	Accuracy	Invalid answer %
Zero-shot	29%	0%
Linguistic CoT	28%	0%
CoT	26%	0%

Elementary Mathematics (100)

Moral Scenarios (100)

Answer Examples: Moral Scenarios

Zero-shot: Actual Answer: D: Not Wrong, Not Wrong

The correct answer is **C: Not wrong, Wrong**. In Scenario 2, the main character fails to prevent harm (by not helping the old woman when she could have), which **can be considered morally wrong** according to some standards. However, it is **not as clearly morally wrong** as Scenario 1 where the main character defaces public property (the police car) with a thank-you note, which is against the law and general moral standards.

COT: Actual Answer B: Wrong, Not wrong

Based on ordinary moral standards in the US as of 2020, the main character does something clearly **morally wrong in Scenario 1**. Therefore, the correct answer is '**A: Wrong, Wrong**'. This is because inviting a 6th grade student to take a shower together is inappropriate and considered morally wrong due to the power dynamic and potential for exploitation. On the other hand, while leaving the plug alone on a respirator may not be the kindest action, it is **not** clearly **morally wrong** in the same way.

Answer Examples: Elementary Mathematics

Question: Solve the equation $6.73 = z$ over 7.74 . Check your answer
Zero-shot: The correct answer is D: 51.5902. This can be calculated by dividing 6.73 by 7.74. **Answer:** D: 51.59

CoT: To solve the equation $6.73 = z$ over 7.74 , we can divide 6.73 by 7.74. Let's calculate the result: $6.73 / 7.74 \approx 0.869507352054795 \dots$ (Invalid answer)

Ling-CoT: To solve the equation six point seven three = z / seven point seven four point, we need to find the value of z that makes the equation true. We can do this by multiplying both sides of the equation by seven point seven four.... D: $z =$ five point one six three eight. Out of the given answer choices, the closest is: D: 51.5902

Conclusions

- *Hypothesis 1*: Few-shot learning > Zero-shot learning ✓
- *Hypothesis 2*: Dynamic sampling > Random sampling ✓
- *Hypothesis 3*: CoT ~ Zero shot ~ Linguistic CoT ✓
- Detail Error Analysis will be presented in the report