

1. Introduction

This paper records the Bengali input to the MRL 2025 Shared Task on ‘Multilingual Physical Commonsense Reasoning’. The shared task's purpose is to produce datasets in multiple languages that are high-quality and culturally relevant. These datasets can be used to see how models can interpret ordinary physical situations. I propose a Bengali dataset that meets the requirements of the shared task and includes both general and culturally specific components for commonsense reasoning.

2. Language Background

More than 230 million people speak Bengali (Bangla), which is an Indo-Aryan language. Most of them dwell in Bangladesh and the Indian states of Assam, Tripura, and West Bengal. It is the seventh most spoken language in the world and has a lengthy history of writing and art. The collection mainly uses Standard Colloquial Bengali, which is the most common way to speak in Kolkata (India), but it also includes references to regional customs, food, and household objects.

3. Putting the Dataset Together

The dataset appears like a PIQA-style multiple-choice test, with a prompt, two possible solutions, and a label (0 or 1) that tells you which one is correct. The steps to building were:

1. Bengali examples prepared by hand by persons who speak the language. It was made sure that each example incorporates "physical commonsense reasoning," which means thinking about how things work and what happens naturally.

2. Making solutions with minimal lexical differences such that the bad choice still makes sense but is wrong in the end.

Finding a balance between things that are generic and things that are peculiar to your culture so that you can talk about a variety of different subjects.

4. Importance to Culture

The dataset contains components pertaining to Bengali cultural practices and quotidian existence, alongside broad common sense thinking. These are:

Cooking food (such creating pitha, dal, and mustard oil). Traditional household items= (such clay pots, bamboo baskets, and clay lamps).

Holidays and traditions (like the way people celebrate Durga Puja and the way people cultivate in the country).

This cultural embedding makes sure that the dataset isn't just a translation of English PIQA items, but instead shows real Bengali common sense knowledge.

5. The Format of the Dataset

The dataset is in TSV format and has these columns:

prompt: The Bengali circumstance where you have to use common sense to find things out.

solution0: A viable response.

solution1: A different choice for the candidate.

label: A number (0 or 1) that tells you if the answer is right.

People who speak Bengali as their primary language looked over all of the samples by hand. I only provided translated PIQA instances for comparison. They don't add up to the 100 original entries that are needed.

6. Examples of Entries

1. Prompt: What would you do once you paint the wood? Let it dry for a few hours so the colour can dry. Solution 1: Let it dry for a few months so the colour can dry. Tag: 0

2. Prompt: If a glass falls off a table, Solution0: It will break. Answer 1: It will jump. Tag: 0

3. (cultural) Prompt: What do you need to do before you prepare pitha using rice flour? Solution0: To produce a mound, you need to combine it with hot water. Solution1: You need to light the powder on a dry fire.

7. Making sure it's good quality

The dataset was made of good quality by ensuring:

At least one native speaker manually reviewed each entry. Most of the prompts were longer than 25 words, and some were written as more than one sentence to make sure they were all varied lengths.

It was vital that the wrong responses were plausible but plainly wrong so that the differences weren't too glaring. To make sure that people in Bengali-speaking areas could utilise the goods more easily, a group of people was evaluated for "cross-dialect intelligibility."

8. Conclusion

The Bengali dataset for the MRL 2025 Shared Task meets all the standards for cultural integration, physical commonsense reasoning, and manual validation. I think that this change will improve the multilingual benchmark, make language comparisons stronger, and teach us something helpful about how Bengali people utilise common sense reasoning.