

SPARK: Step-by-step Proof Assistant for Reasoning and Knowledge

Language Warriors

Aditya Raghuvanshi
Srivatsa

(2021114009)
(2024701003)

Motivation

Mathematical reasoning is a cornerstone of scientific and technological advancement, yet existing AI systems often struggle with rigorous, step-by-step problem-solving and verification. Even after giving questions we majorly receive wrong answers.

This research bridges the gap by developing an interactive, math assistant that combines LLM reasoning with formal verification.

- By enabling step-by-step explanations, auto-verification, and cross-domain adaptability, this system empowers us to tackle complex problems with confidence.
- It also advances AI fairness and interpretability, ensuring reliable and ethical mathematical reasoning tools for diverse applications.
- We also aim to democratise the reasoning capabilities of LLM by experimenting within the limits of resource by working on small LLMs and consumer grade GPU.

Experiments

1. Baseline :
 - a. Qwen 2-1.5B-Instruct
 - b. Qwen 2.5-MATH-1.5B-Instruct - 4-bit quantized
 - c. Gemini-2.0-Flash (for Comparisions)
2. Multi-step reason generation using theory augmentation assisted by judge LLMs.
 - a. Linear step Generation
3. Explore RL based and RL-free methods
 - a. GRPO - RL Method Deepseek

Note :

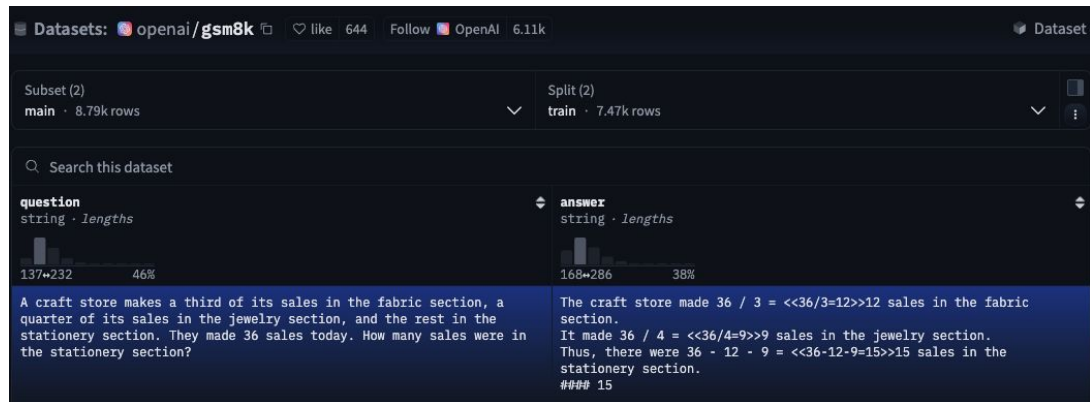
- We will make use of GPU efficient fine tuning of quantized SLMs

Process

Dataset Preparation

1. gsm8k

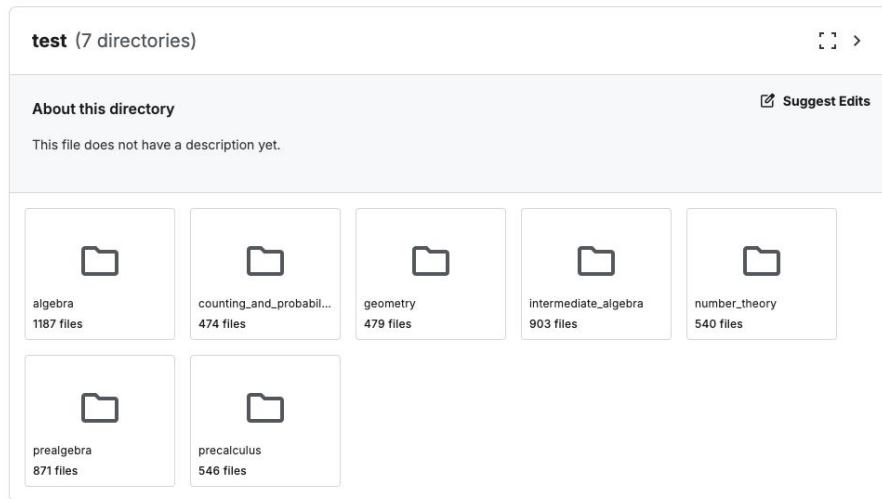
- Run a regex to retrieve the answers from the solutions for future comparisons.



Dataset Preparation

1. Extracted all the questions from concept-wise directories to csv format
2. Filtered it with only level 4 & level 5 questions
3. Zero shot prompting to gemini-2.0 to retrieve answers from solutions.

Final dataset size - 2538 Questions



```
▼ "root" : { 4 items
  "problem" : string "How many vertical asymptotes does the graph of  $y=\frac{2}{x^2+x-6}$  have?"
  "level" : string "Level 3"
  "type" : string "Algebra"
  "solution" :
    string "The denominator of the rational function factors into  $x^2+x-6=(x-2)(x+3)$ . Since the numerator is always nonzero, there is a vertical asymptote whenever the denominator is  $\$0$ , which occurs for  $x = 2$  and  $x = -3$ . Therefore, the graph has  $\boxed{2}$  vertical asymptotes."
}
```

Baseline Models and Comparisons

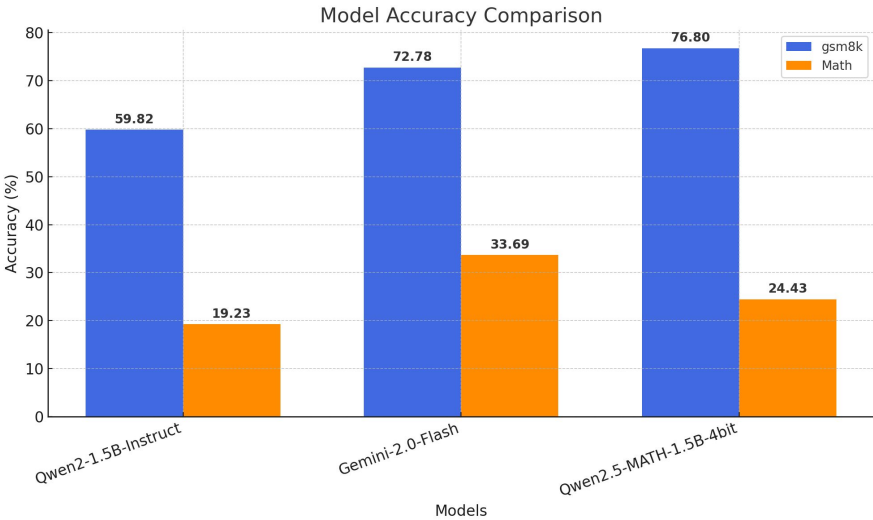
We start by establishing strong baseline performances using the following models:

- **Qwen2-1.5B-Instruct** (Standard Instruction-Tuned Model)
- **Qwen2.5-MATH-1.5B-Instruct (4-bit Quantized)** (Mathematics-specialized fine-tuned model)
- **Gemini-2.0-Flash** (Used for comparison as a high-performing reference model)

Each of these models was evaluated on the **gsm8k** and **Math** datasets to measure their capabilities in handling arithmetic and algebraic reasoning ranging from very basic level questions to high level competitive examinations.

Results

Experiment	Model	Dataset	Accuracy
Baseline & Comparisons	Qwen2-1.5B-Instruct	gsm8k	59.82%
		Math	19.23%
	Qwen2.5-MATH-1.5B-4bit	gsm8k	76.80%
		Math	24.43%
	Gemini-2.0-Flash (Comparison Model)	gsm8k	72.78%
		Math	33.69%



Post Result Analyses

1. Formatting Issues in Answers

One of the key observations in Qwen's outputs is inconsistent formatting of numerical answers. Since Qwen2-1.5B is a relatively small model, it does not always adhere to a structured output format, leading to issues such as:

Expected Output: 6600

Possible Model Outputs:

- 6,600 (Comma-separated format)
- 6600.0 (Floating-point representation)
- 6600/1 (Fraction form)

Post Result Analyses

Incomplete Answers Due to Output Length Limits

```
6,"In triangle $PQRS$, point $T$ is on $PR$ and point $S$ is on $PQ$ such that $TS \parallel RQ$. The measure of $\angle RPQ$ is $65^\circ$, and the measure of $\angle TSQ$ is $145^\circ$. What is the measure of $\angle PRQ$?","system
Please reason step by step, and put your final answer within \boxed{}.
user
In triangle $PQRS$, point $T$ is on $PR$ and point $S$ is on $PQ$ such that $TS \parallel RQ$. The measure of $\angle RPQ$ is $65^\circ$, and the measure of $\angle TSQ$ is $145^\circ$. What is the measure of $\angle PRQ$?
assistant
```

To solve for the measure of $\angle PRQ$ in triangle PQR , we start by analyzing the given information and using properties of parallel lines and similar triangles.

Given:

- $TS \parallel RQ$
- $\angle RPQ = 65^\circ$
- $\angle TSQ = 145^\circ$

Since $TS \parallel RQ$, we can use the properties of parallel lines and corresponding angles. Specifically, $\angle TSQ$ is an exterior angle to triangle PQR at vertex Q . The measure of an exterior angle of a triangle is equal to the sum of the measures of the two non-adjacent interior angles.

$$\angle TSQ = \angle PQR + \angle PRQ$$

Given that $\angle TSQ = 145^\circ$, we can substitute this value into the equation:

$$145^\circ = \angle PQR + \angle PRQ$$

Next, we need to find the measure of $\angle PQR$. Since $TS \parallel RQ$ and PT is on PR and PS is on PQ , triangle PQR is similar to triangle TSQ by the AA (Angle-Angle) similarity criterion. This means that corresponding angles are equal.

However, we can also use the fact that the sum of the angles in any triangle is 180° . In triangle PQR , we have:

$$\angle PQR + \angle PRQ + \angle RPQ = 180^\circ$$

Substituting the known value of $\angle RPQ = 65^\circ$, we get:

$$\angle PQR + \angle PRQ + 65^\circ = 180^\circ$$

Rearranging the equation to solve for $\angle PQR + \angle PRQ$, we get:

$$\angle PQR + \angle PRQ = 115^\circ$$

We already know from the exterior angle property that:

$$\angle TSQ = \angle PQR + \angle PRQ$$

Post Result Analyses

Was the Model's Solution Directionally Correct?

- We also evaluate whether the model was moving toward the correct solution, even when it failed to complete the problem.

Key cases:

- Correct approach but stopped early → Fixable via longer generation
- Incorrect approach from the start → Indicates fundamental gaps in reasoning

Error Categorization: Why Do Models Get It Wrong?

1. Incorrect & Incomplete Solutions

- Some model outputs were partially correct but failed to reach a final answer.
- Often caused by early stopping or loss of logical coherence in multi-step solutions.

2. Conceptual Knowledge Gaps

- Some errors were due to missing fundamental mathematical concepts, particularly in higher-level algebra and proofs in the Math dataset.

Possible Solutions

Step by Step Thinking - Property Augmentation

Step 1:

1. Use GPT4 models to generate step by step thought process towards solution
2. Each step will consist of a mathematical property, followed by the evaluation of property towards solving the problem

Step by Step Thinking - Fine Tuning of SLMs with step data

Steps 2:

1. With the dataset prepared previously will finetune small language models at 4-bit quantisation.

GRPO

Group Relative Policy Optimization (GRPO) is a reinforcement learning (RL) algorithm designed to improve large language models' mathematical reasoning capabilities efficiently.

How its different from PPO ? :

1. Eliminating the Value Model (Critic)
2. Instead of estimating rewards for individual outputs independently, GRPO assigns rewards based on their relative quality within a group of generated responses.

How will GRPO benefit us ?

1. Formatting Issues (e.g., "6,600" vs. "6600.0" vs. "6600/1")

GRPO can be applied to improve formatting consistency by incorporating reward functions that favor standardized answer formats. **However**, GRPO won't automatically enforce a fixed format unless the reward model explicitly prioritizes a particular one.

2. Incomplete Answers

GRPO can help mitigate output truncation by rewarding more complete solutions

Note : It won't fix fundamental knowledge gaps → that's why we have the Linear Idea

Thank You