

Entry view

Provides an easy entry point to our tool, with explanation of its main components.

a Entry view to the CLEAR interface

CLEAR: Evaluation Results Analyzer

Easily explore and filter your dataset based on discovered issues and score.

What you can do here:

- View recurring problems discovered in your dataset
- Filter rows by issue type and score range
- See stats and score distribution for filtered data
- Drill down into individual examples by clicking on rows

Filtering Mechanism

Provides a dynamic ability to filter the data by issues types, and score range.

c The filtering mechanism control panel

Select Filters

Use the filters below to narrow down instances based on issues and score:

Click **Apply Filter** to update the results or **Clear Filters** to reset everything.

Include ANY of (OR) Must ALSO have (AND) Exclude ANY of (NOT)

Choose an option Mathematical er... x Logical errors de... x

Select instances with failures in ONLY the above issues

Select score range

Filter Summary

- ALL of ['Mathematical errors in calculations, including rounding and final steps', 'Logical errors despite clear reasoning']
- Score between 0.0 and 0.5

Found 33 entries matching entries

Model Behavior

Provides instance-level info on data slices to enable issue-based error analysis

Insight: By filtering for instances marked with specific issues, we can choose examples that help us understand the model's behavior and its error patterns. We can also examine the available instance metadata and the judge's explanations.

e Detailed view of selected instance

- Please select an entry from the table or dropdown:
- From the table:** Click the grey strip to the left of the first column.
- From the dropdown:** Select an option or start typing the desired index or question.

Clear selection

	question
20	I have 10 liters of orange drink that are two-thirds water and I wish
160	A tank of water has a depth of 17 feet on Monday. On Tuesday, the
203	Carl, Æs favorite food is cheese. He ate a sandwich every day this w
205	Brinley is in Mr. Bert's math class. Mr. Bert gives six tests each seme
218	Adrian's age is three times the age of Harriet, and Harriet is half the
<input checked="" type="checkbox"/>	287 A 76-star flag has three rows of 8 stars, two rows of 6 stars and the
301	A simple folding newspaper or tabloid can be made by folding a pie
322	Jam has three boxes full of pencils and 2 loose pencils which give a
324	Ralph has a record store where people can trade their own records
373	A company bought \$400000 worth of equipment from a retailer bus

Details for Math Entry Index: 287

Question: A 76-star flag has three rows of 8 stars, two rows of 6 stars and the rest are 5-star rows. How many rows of 5 stars are there altogether on the flag?

Response: To find the number of 5-star rows, we first need to determine how many rows of stars are on the flag in total.

The flag has:

- * Three rows of 8 stars
- * Two rows of 6 stars

So, there are already $3 \cdot 8 + 2 \cdot 6 = 24 + 12 = 36$ rows accounted for.

The question states that the 76-star flag has these rows and more 5-star rows. Therefore, the number of 5-star rows must be the total number of rows minus the number of rows already accounted for.

Number of 5-star rows = Total number of rows - Accounted rows
 $= 76$ (the total number of stars divided by the number of stars in a row) - 36
 $= 76 - 36$
 $= 40$

So, there are 40 rows of 5 stars on the flag.

Model Input (Prompt):

Full Evaluation Text:

Evaluation Summary:

The model answer contains errors in reasoning and calculations. It incorrectly calculates the total number of stars and misapplies subtraction to find the number of 5-star rows.

Ground Truth:

The rows of 8 stars have a total of: 3 rows * 8 stars/row = $3 \cdot 8 = 24$ stars.
 The rows of 6 stars have a total of: 2 rows * 6 stars/row = $2 \cdot 6 = 12$ stars.
 Including the rows of 8 and 6 stars, we have 24 stars + 12 stars = $24 + 12 = 36$ stars.
 Now we need to know how many stars are missing, 76 stars - 36 stars = $76 - 36 = 40$ stars.
 Now we need to distribute the missing stars in rows of 5 stars, 40 stars / 5 rows = $40 / 5 = 8$ rows of 5 stars.
 ####

Score: 0.0

Recurring Issues:

- Incorrect understanding of problem statements leading to flawed reasoning
- Mathematical errors in calculations, including rounding and final steps
- Misunderstanding or incorrect application of mathematical concepts or methods
- Logical errors despite clear reasoning

f Instance-level view

Issues View

Provides overview of model issues to help identify the system error patterns.

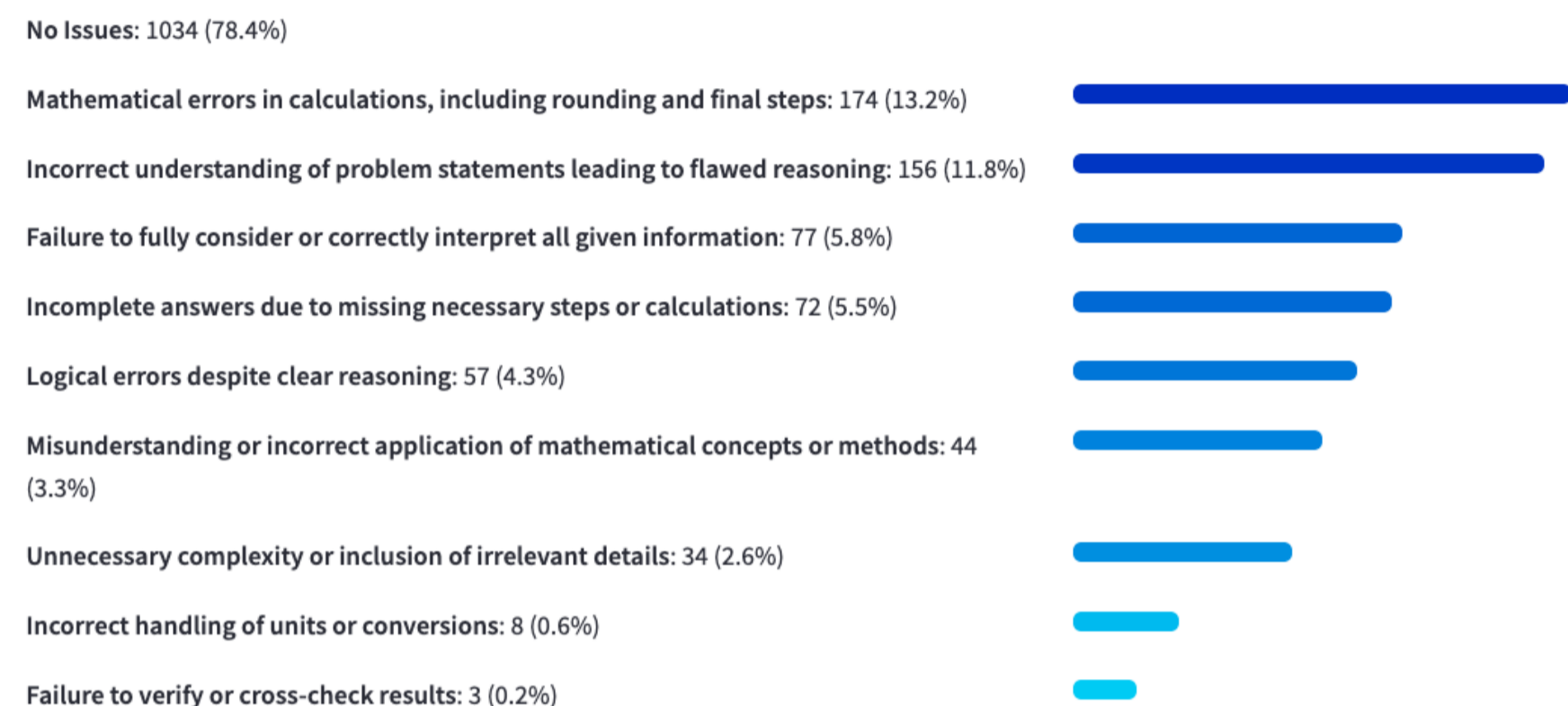
Insight: The model suffers from calculation errors as its major problem

b The issues distribution as recognized by our approach

Issues Distribution

Total evaluations processed: 1319

Frequency of each issue (sorted high to low):



Comparison View

Provides a comparison view of the issues distribution in the full dataset compared to the filtered one.

Insight: Misunderstanding of the problem appear in most cases that also exhibit mathematical and logical errors.

d Issues frequencies comparison of the full and filter data

Comparison of Issue Frequencies:

