Math LLM Evaluation Suite

Single Question Mode | Dataset Evaluation Mode | Visualize Auto-Loop Results

Dataset Evaluation — Visual Summary

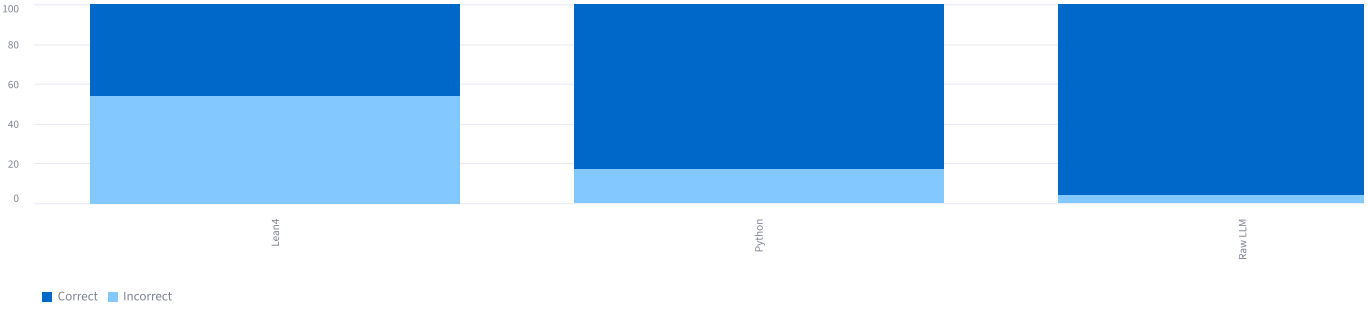
Raw Results Table

	ID	Problem	Ground Truth Answer	Ground Truth Solution
0	test/precalculus/807.json	Convert the point $(0,3)$ in rectangular coordinates to polar coordinates. Enter your	\left(3, \frac{\pi}{2} \right)	We have that $r = \sqrt{0^2 + 3^2} = 3.$ Also, if we draw the line connecting the origin
1	test/intermediate_algebra/1994.json	Define \[p = \sum_{k = 1}^\infty \frac{1}{k^2} \quad \text{and} \quad q = \sum_{k = 1}^	p - q	We count the number of times $\frac{1}{n^3}$ appears in the sum \[\sum_{j = 1}^\infty
2	test/algebra/2584.json	If $f(x) = \frac{3x-2}{x-2}$, what is the value of $f(-2) +f(-1)+f(0)$? Express your answer	\frac{14}{3}	$f(-2)+f(-1)+f(0)=\frac{3(-2)-2}{-2-2}+\frac{3(-1)-2}{-1-2}+\frac{3(0)-2}{0-2}=\frac{-8}{-4}+
3	test/number_theory/572.json	How many positive whole-number divisors does 196 have?	9	First prime factorize $196=2^2\cdot7^2$. The prime factorization of any divisor of 19
4	test/algebra/1349.json	The results of a cross-country team's training run are graphed below. Which student l	\text{Evelyn}	Evelyn covered more distance in less time than Briana, Debra and Angela, so her aver
5	test/prealgebra/1622.json	A regular hexagon can be divided into six equilateral triangles. If the perimeter of one	42	The side length of the hexagon is equal to the side length of one of the equilateral tria
6	test/number_theory/515.json	What is the smallest positive perfect cube that can be written as the sum of three con	27	The sum of three consecutive integers takes the form $(k-1)+(k)+(k+1)=3k$ and hence
7	test/precalculus/927.json	The set of points (x,y,z) that satisfy \[2x = 3y = -z\]is a line. The set of points (x,y,z)	90^\circ	For the first line, let $t = 2x = 3y = -z.$ Then \[\begin{pmatrix} x \\ y \\ z \end{pmatrix}
8	test/algebra/2036.json	What is the distance, in units, between the points $(2, -6)$ and $(-4, 3)$? Express your	3\sqrt{13}	We use the distance formula: \begin{align*} \sqrt{(2 - (-4))^2 + ((-6) - 3)^2} &= \sqrt{6^
9	test/prealgebra/1139.json	The expression $2\cdot 3 \cdot 4\cdot 5+1$ is equal to 121, since multiplication is car	4	By the associative property of multiplication, it doesn't help to insert parentheses tha

Summary Metrics

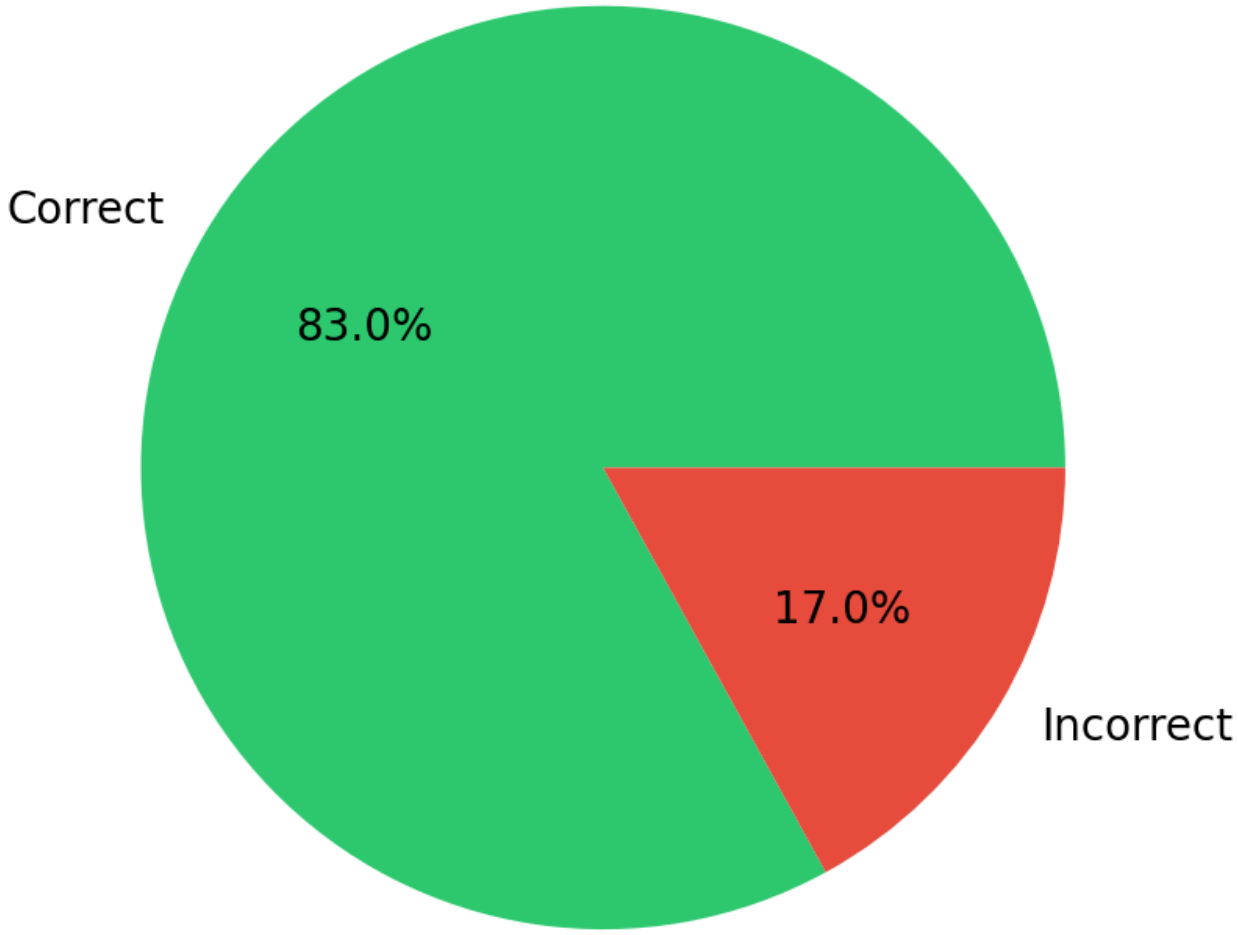
Total Questions	Raw LLM Correct	Python Correct	Lean4 Correct
100	96 (96.0%)	83 (83.0%)	46 (46.0%)

Correctness by Method



Python Correctness Distribution

Python Answer Correctness



Filter Incorrect Predictions

Filter by method:

Raw LLM Incorrect

4 incorrect predictions found:

	ID	Problem	Ground Truth Answer	Ground Truth Solution	Pure
18	test/geometry/434.json	$\overline{BC}$ is parallel to the segment through A, and $AB = BC$. What is the nu	28	Angle $\angle BCA$ and the angle we're trying to measure are alternate interior angle	56
67	test/prealgebra/1233.json	The lengths of two opposite sides of a square are decreased by 40% while the leng	10	Let A be the area of the square. The lengths of one pair of opposite sides was decre	10%
80	test/intermediate_algebra/1849.json	Find the number of integer values of k in the closed interval $[-500,500]$ for which	501	First, note that if $k < 0,$ then $\log(kx)$ is defined for $x \in (-\infty, 0),$ and is strictl	1
96	test/intermediate_algebra/662.json	Find the real roots of \[\frac{(x+ 1)(x - 3)}{5(x + 2)(x - 4)} + \frac{(x + 3)(x - 5)}{9(x + 4)(x -	1 \pm \sqrt{19}	Multiplying out each numerator and denominator, we get \[\frac{x^2 - 2x - 3}{5(x^2 - 2	-6.53