Math LLM Evaluation Suite

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

Dataset Evaluation — Visual Summary

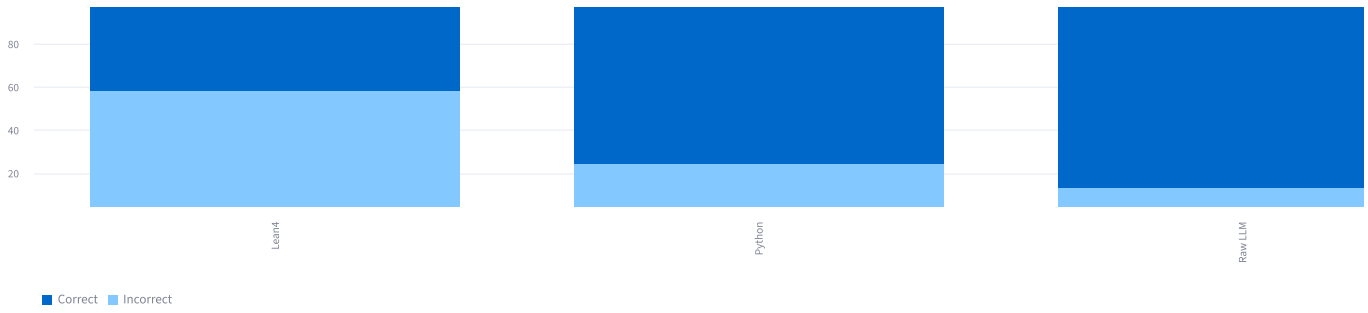
Raw Results Table

	ID	Problem	Ground Truth Answer	Ground Truth Solution	Pure LLM Ar
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	(3, \pi/2)
1	test/intermediate_algebra/	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	\boxed{p - q
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}+	14/3
3	test/number_theory/572.js	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	9
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	Evelyn
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	42
6	test/number_theory/515.js	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	27
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}	90
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^	3√13
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	Error

Summary Metrics

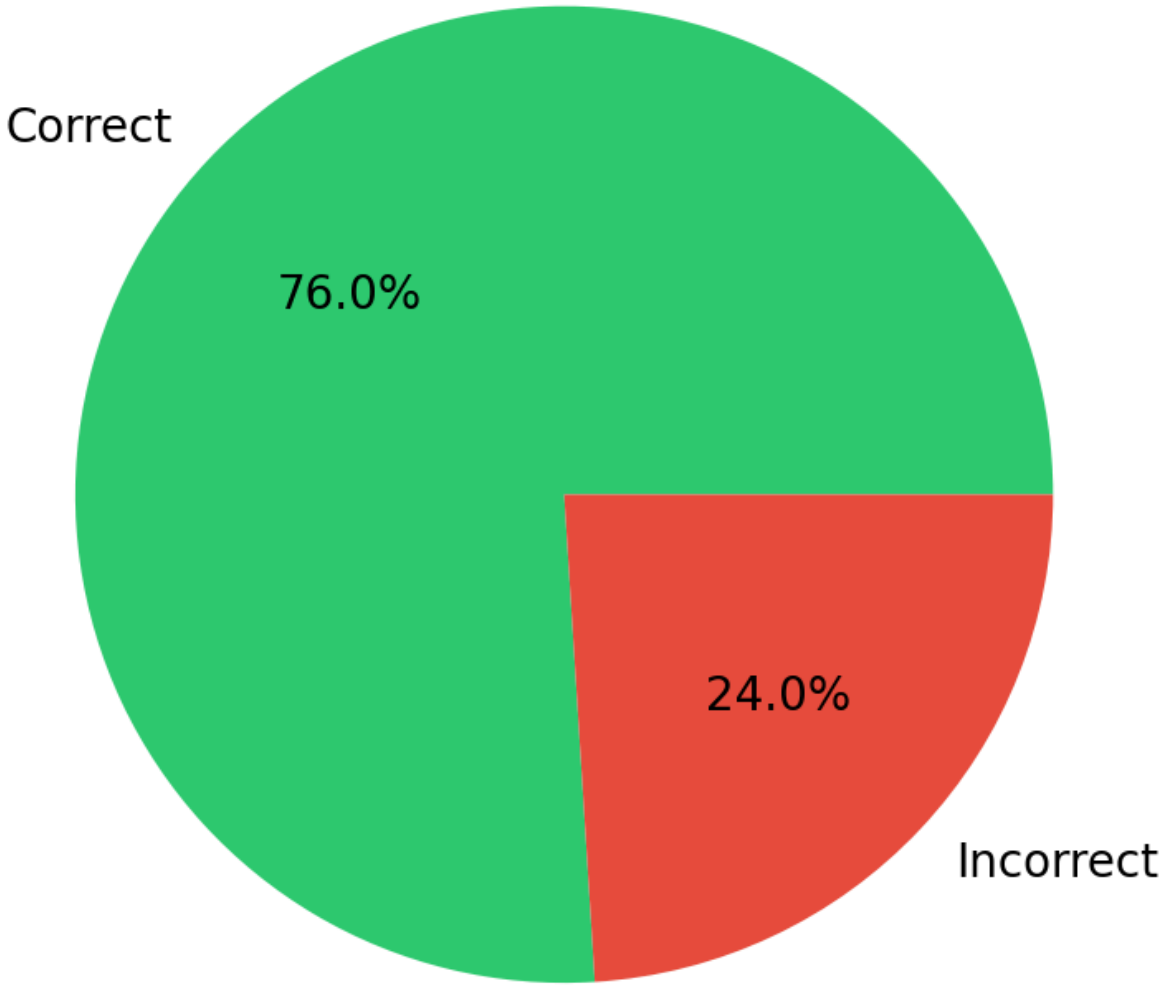
Total Questions	Raw LLM Correct	Python Correct	Lean4 Correct
100	87 (87.0%)	76 (76.0%)	42 (42.0%)

Correctness by Method



Python Correctness Distribution

Python Answer Correctness



Filter Incorrect Predictions

Filter by method:

All Incorrect

63 incorrect predictions found:

	ID	Problem	Ground Truth Answer	Ground Truth Solution	P
0	test/precalculus/807.json	Convert the point $(0,3)$ in rectangular coordinates to polar coordinates. Enter your	\left(3, \frac{\pi}{2} \rig	We have that $r = \sqrt{0^2 + 3^2} = 3.$ Also, if we draw the line connecting the origin	(3
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	\[
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	9
4	test/algebra/1349.json	The results of a cross-country team's training run are graphed below. Which student	\text{Evelyn}	Evelyn covered more distance in less time than Briana, Debra and Angela, so her aver	E
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	2
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}	9
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^	3
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	E
10	test/number_theory/1032.json	What is the least positive integer multiple of 30 that can be written with only the digit	2220	Let M be the least positive multiple of 30 that can be written with only the digits 0 a	2
11	test/intermediate_algebra/1197.json	Let $p(x)$ be a polynomial of degree 5 such that \[p(n) = \frac{n}{n^2 - 1}\]for $n = 2,$	\frac{3}{56}	Let $q(x) = (x^2 - 1) p(x) - x.$ Then $q(x)$ has degree 7, and $q(n) = 0$ for $n = 2$, 3, 4,	T