

Math LLM Evaluation Suite

 Single Question Mode  Dataset Evaluation Mode  Visualize Auto-Loop Results

Dataset Evaluation — Visual Summary

Raw Results Table

ID	ID	Problem	Ground Truth Answer	Ground Truth Solution	Pure LLM Answer
0	test/precalculus/807.json	Convert the point $\langle 0, 3 \rangle$ in rectangular coordinates to polar coordinates. Enter your answer in the form $\langle r, \theta \rangle$.	$\left(3, \frac{\pi}{2} \right)$	We have that $r = \sqrt{0^2 + 3^2} = 3$. Also, if we draw the line connecting the origin $(0, 0)$ to the point $(0, 3)$, we see that the angle between the positive x-axis and this line is $\frac{\pi}{2}$.	$\left(3, \frac{\pi}{2} \right)$
1	test/intermediate_algebra/1349.json	Define $p = \sum_{k=1}^{\infty} \frac{1}{k^2}$. What is the value of $p - q$?	$p - q$	We count the number of times $\frac{1}{n^3}$ appears in the sum $\sum_{j=1}^{\infty} \frac{1}{j^2}$.	$p - q$
2	test/algebra/2584.json	If $f(x) = \frac{1}{x-2}$, what is the value of $f(-2) + f(-1) + f(0)$? Express your answer as a common fraction.	$\frac{1}{3}$	$f(-2) + f(-1) + f(0) = \frac{1}{-2-2} + \frac{1}{-1-2} + \frac{1}{0-2} = \frac{-1}{-4} + \frac{-1}{-3} + \frac{-1}{-2} = \frac{1}{4} + \frac{1}{3} + \frac{1}{2} = \frac{13}{12}$	$\frac{13}{12}$
3	test/number_theory/572.json	How many positive whole-number divisors does 196 have?	9	First prime factorize $196 = 2^2 \cdot 7^2$. The prime factorization of any divisor of 196 is of the form $2^a \cdot 7^b$ where $0 \leq a \leq 2$ and $0 \leq b \leq 2$. There are $(2+1)(2+1) = 9$ such divisors.	9
4	test/algebra/1349.json	The results of a cross-country team's training run are graphed below. Which student ran the farthest?	Evelyn	Evelyn covered more distance in less time than Briana, Debra and Angela, so her average speed was the highest.	Evelyn
5	test/prealgebra/1622.json	A regular hexagon can be divided into six equilateral triangles. If the perimeter of one triangle is 42, what is the perimeter of the hexagon?	42	The side length of the hexagon is equal to the side length of one of the equilateral triangles.	42
6	test/number_theory/515.json	What is the smallest positive perfect cube that can be written as the sum of three consecutive integers?	27	The sum of three consecutive integers takes the form $(k-1)+(k)+(k+1)=3k$ and hence must be a multiple of 3.	27
7	test/precalculus/927.json	The set of points (x, y, z) that satisfy $ 2x - 3y - z = 0$ is a line. The set of points (x, y, z) that satisfy $x^2 + y^2 + z^2 = 90$ is a sphere.	90°	For the first line, let $t = 2x - 3y - z$. Then $\begin{pmatrix} x & y & z \end{pmatrix} = \begin{pmatrix} t \\ 2t \\ 3t \end{pmatrix}$.	90°
8	test/algebra/2036.json	What is the distance, in units, between the points $(2, -6)$ and $(-4, 3)$? Express your answer as a decimal.	$3\sqrt{13}$	We use the distance formula: $\sqrt{(2 - (-4))^2 + ((-6) - 3)^2} = \sqrt{6^2 + 9^2} = \sqrt{117} = 3\sqrt{13}$	$3\sqrt{13}$
9	test/prealgebra/1139.json	The expression $2 \cdot 3 \cdot 4 \cdot 5 + 1$ is equal to 121, since multiplication is carried out from left to right. Express this reasoning using the associative property of multiplication.	4	By the associative property of multiplication, it doesn't help to insert parentheses to change the order of operations.	Error

Summary Metrics

Total Questions

100

Raw LLM Correct

87 (87.0%)

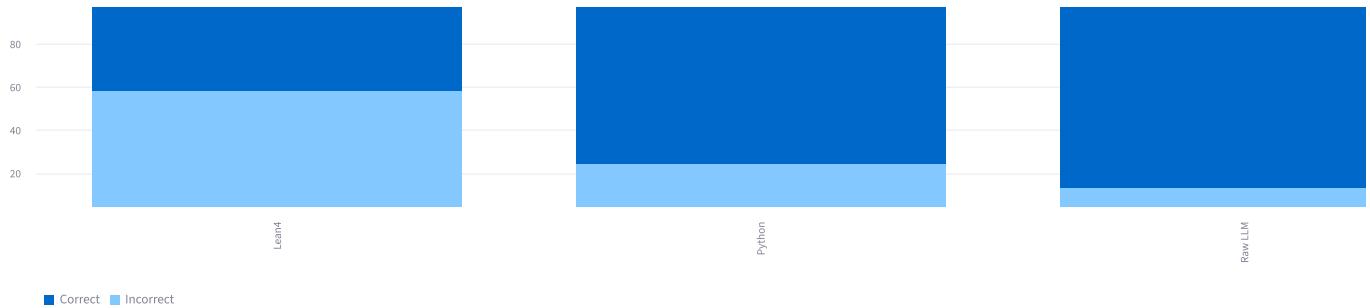
Python Correct

76 (76.0%)

Lean4 Correct

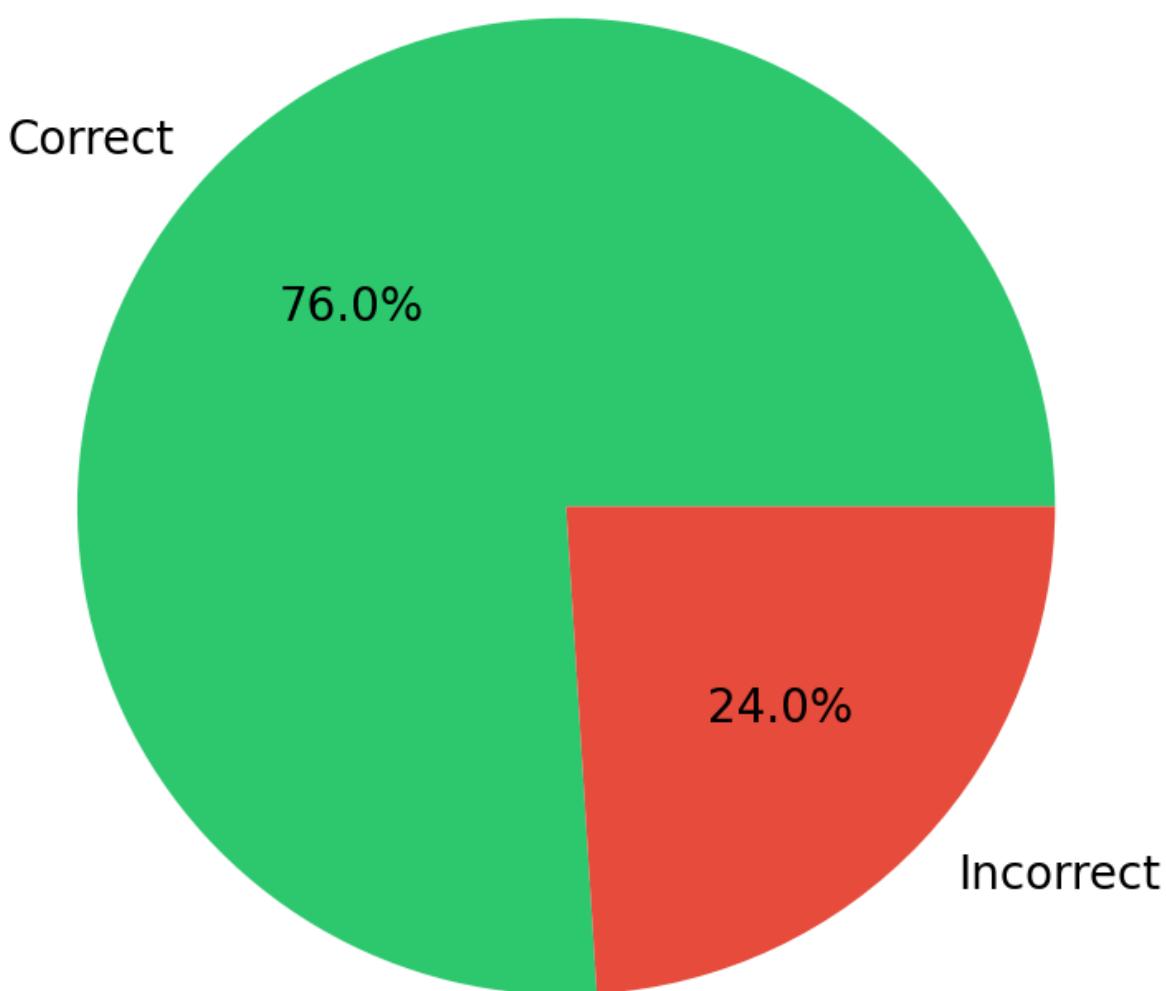
42 (42.0%)

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
0	test/calculus/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 \$(0,0)\$ to the point \$(0,3)\$, the angle between the positive x-axis and the line is \$\frac{\pi}{2}\$.	
1	test/intermediate_algebra/1994.json	Define \$p = \sum_{k=1}^{\infty} \frac{1}{k^2}\$ where \$q = \sum_{k=1}^{\infty} p - q\$	We count the number of times \$\frac{1}{n^2}\$ appears in the sum \$\sum_{j=1}^{\infty} j^{-2}\$. This is the same as the sum of the reciprocals of the squares of the natural numbers, which is known to be \$\frac{\pi^2}{6}\$.
3	test/number_theory/572.json	How many positive whole-number divisors does 196 have?	First prime factorize \$196 = 2^2 \cdot 7^2\$. The prime factorization of any divisor of 196 must be of the form \$2^a \cdot 7^b\$ where \$0 \leq a \leq 2\$ and \$0 \leq b \leq 2\$. There are \$(2+1)(2+1) = 9\$ such divisors.
4	test/algebra/1349.json	The results of a cross-country team's training run are graphed below. Which student I (text{Evelyn})	Evelyn covered more distance in less time than Briana, Debra and Angela, so her average speed was the highest.
6	test/number_theory/515.json	What is the smallest positive perfect cube that can be written as the sum of three consecutive integers?	The sum of three consecutive integers takes the form \$(k-1)+(k)+(k+1)=3k\$ and hence \$27\$ is the smallest positive perfect cube that can be written as the sum of three consecutive integers.
7	test/calculus/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} = \begin{pmatrix} t/2 \\ t/3 \\ -t \end{pmatrix} = t \begin{pmatrix} 1/2 \\ 1/3 \\ -1 \end{pmatrix}\$. The second line is \$x = 0\$.
8	test/algebra/2036.json	What is the distance, in units, between the points \$(2, -6)\$ and \$(-4, 3)\$? Express your answer as a decimal.	We use the distance formula: \$\sqrt{(2 - (-4))^2 + ((-6) - 3)^2} = \sqrt{6^2 + (-9)^2} = \sqrt{117} = 3\sqrt{13}\$.
9	test/prealgebra/1139.json	The expression \$2 \cdot 3 \cdot 4 \cdot 5 + 15\$ is equal to 121, since multiplication is carried out from left to right.	By the associative property of multiplication, it doesn't help to insert parentheses that change the order of operations.
10	test/number_theory/1032.json	What is the least positive integer multiple of 30 that can be written with only the digits 2 and 0?	Let \$M\$ be the least positive multiple of 30 that can be written with only the digits 2 and 0. Since 30 is divisible by 2 and 5, \$M\$ must end in 0. The sum of the digits of \$M\$ must be divisible by 3. The only possible values for \$M\$ are 2220 and 22220.
11	test/intermediate_algebra/1197.json	Let \$p(x)\$ be a polynomial of degree 5 such that \$ p(n) = \frac{1}{n^2 - 1}\$ for \$n = 2, 3, 4, 5, 6\$.	Let \$q(x) = (x^2 - 1)p(x) - x\$. Then \$q(x)\$ has degree 7, and \$q(n) = 0\$ for \$n = 2, 3, 4, 5, 6\$.

