Math LLM Evaluation Suite

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

Dataset Evaluation — Visual Summary

Raw Results Table

	ID	Raw LLM Answer	Python Output	Dataset Answer	Datas
0	test/precalculus/807.json	I need to convert the rectangular coordinates (0, 3) to polar coordinates (r, θ). Given:	(3.0, 1.570796326794	\left(3, \frac{\pi}{	Match
1	test/intermediate_algebra/1994.json	I need to find a way to express $\sum_{j=1}^\infty \sum_{k=1}^\infty \frac{1}{(j+k)^3}$	-zeta(3) + pi**2/6	p - q	Mism
2	test/algebra/2584.json	I need to find the values of f(-2), f(-1), and f(0) for the function f(x) = (3x-2)/(x-2), then s	4.666666666666667	\frac{14}{3}	Let m
3	test/number_theory/572.json	I need to find the number of positive whole-number divisors of 196. First, I'll find the	9	9	Match
4	test/algebra/1349.json	Looking at the graph, I need to find which student has the greatest average speed ove	Evelyn	\text{Evelyn}	Match
5	test/prealgebra/1622.json	Let me work through this step-by-step. Given information: - A regular hexagon can be	42.0	42	Match
6	test/number_theory/515.json	I need to find the smallest positive perfect cube that can be written as the sum of thre	27	27	Match
7	test/precalculus/927.json	I need to find the angle between two lines defined by parametric equations. First, let	90.0	90^\circ	Match
8	test/algebra/2036.json	I need to find the distance between the points $(2, -6)$ and $(-4, 3)$ using the distanc	3√13	3\sqrt{13}	Match
9	test/prealgebra/1139.json	I need to find all possible values that can be obtained by inserting parentheses in the	4	4	Match

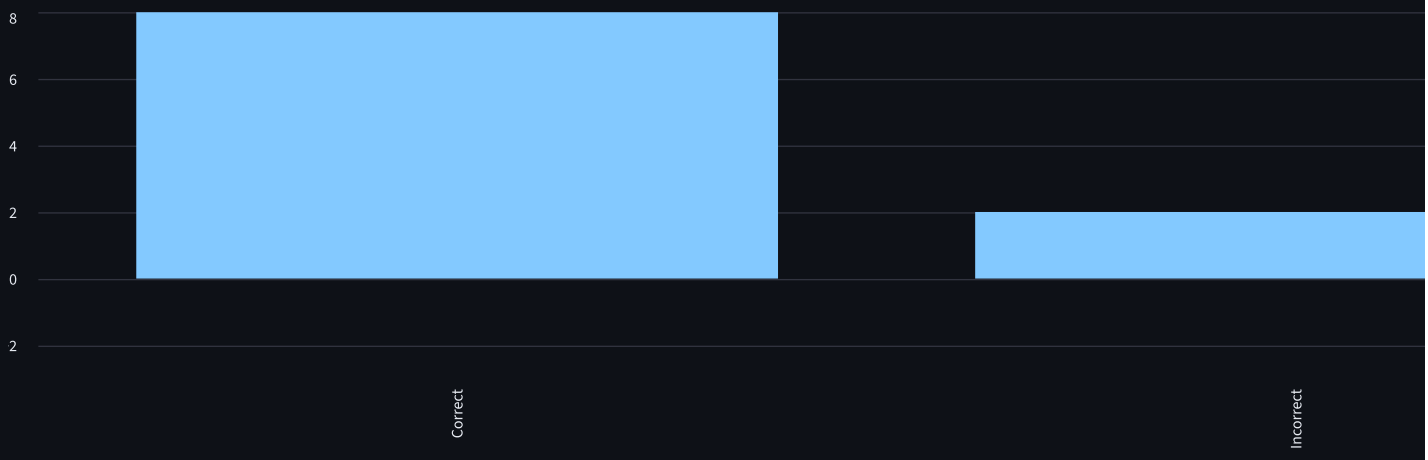
Summary Metrics

Total Questions
10

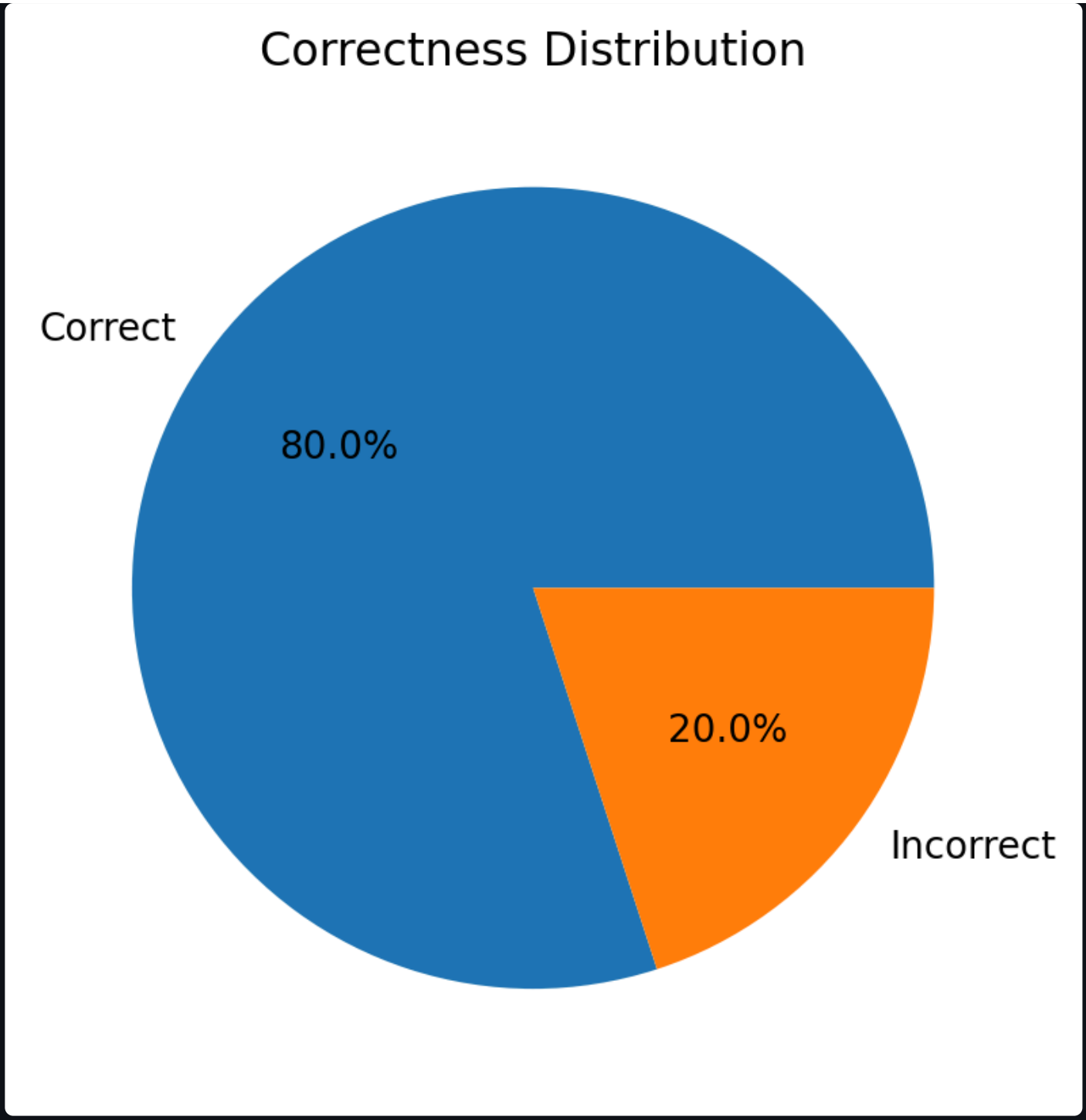
Correct Python Outputs
8

Python Accuracy (%)
80.00%

Correct vs Incorrect Predictions



Distribution



LLM vs Python Match Distribution

	LLM vs Python	count
0	Match The answer is correct. The point (0, 3) in rectangular coordinates converts to (3, π/2) in polar coordinates, where r = 3 > 0 and θ = π/2 satis	
1	Match The solution correctly derives that $\sum_{j=1}^\infty \sum_{k=1}^\infty \frac{1}{(j+k)^3} = p - q$, where $p = \sum_{k=1}^\infty \frac{1}{k	
2	Match All calculations are correct: - f(-2) = 2 - f(-1) = 5/3 - f(0) = 1 - Sum = 2 + 5/3 + 1 = 3 + 5/3 = 9/3 + 5/3 = 14/3 The decimal equivalent 4.666... ec	
3	Match The LLM answer is correct. The prime factorization of 196 is $2^2 \times 7^2$, and using the divisor formula (2+1)(2+1) = 9, which matches the manu	
4	Match The solution correctly identifies that average speed is calculated as distance/time, reads the coordinates from the graph accurately, com	
5	Match The solution is correct. The reasoning properly establishes that each equilateral triangle has sides of 7 inches (21÷3), and since the regula	
6	Match The solution is correct. The smallest positive perfect cube that can be written as the sum of three consecutive integers is indeed 27, corre	
7	Match The solution correctly finds the direction vectors for both lines by setting the common ratios equal to a parameter t, then calculates the c	
8	Match The solution is correct. The distance calculation follows the distance formula properly: - Points (2, -6) and (-4, 3) - Distance = √[(-4-2)² + (3	

Filter Incorrect Predictions

2 incorrect predictions found:

	ID	Raw LLM Answer	Python Output	Dataset Answer	Datas
1	test/intermediate_algebra/1994.json	I need to find a way to express $\sum_{j=1}^\infty \sum_{k=1}^\infty \frac{1}{(j+k)^3}$	-zeta(3) + pi**2/6	p - q	Mism
2	test/algebra/2584.json	I need to find the values of f(-2), f(-1), and f(0) for the function f(x) = (3x-2)/(x-2), then s	4.666666666666667	\frac{14}{3}	Let m