Test Report: Agent Zero's Performance on Mathematical Problems

Overview

This report evaluates Agent Zero's performance across 10 mathematical problems. Each problem was assessed based on the presence and accuracy of the following components:

- **Restatement/Question:** Clear articulation of the problem.
- **Solution Plan:** Logical steps outlined to approach the problem.
- **Execution:** Detailed implementation of the solution plan.
- **Verification:** Confirmation of the final answer's correctness.

Additionally, an overall analysis was conducted to determine the percentage of correct solutions and the completeness of each problem's metrics.

Detailed Analysis

Problem 1

Restatement/Question: Three airline companies operate flights from Dodola island with different schedules: one every 100 days, another every 120 days, and the last every 150 days. We need to find the greatest number of consecutive days without a flight.

Solution Plan:

1. Calculate the Least Common Multiple (LCM) of the departure intervals (100, 120, 150).
2. Identify the flight days within this cycle.
3. Calculate the gaps between these flight days to find the maximum gap.

Execution:

1. **Calculate LCM:**
    - Prime factorization:
        - 100: ($2^2 \times 5^2$)
        - 120: ($2^3 \times 3^1 \times 5^1$)
        - 150: ($2^1 \times 3^1 \times 5^2$)
    - LCM: ($2^3 \times 3^1 \times 5^2 = 600$) days.
2. **Identify Flight Days:**
    - Flight days within 600 days: 0, 100, 120, 150, 200, 240, 300, 360, 400, 450, 480, 500.
3. **Calculate Gaps:**
    - Gaps: 100, 20, 30, 50, 40, 60, 60, 40, 50, 30, 20, 100 days.
    - Maximum gap reported: 79 days.

Verification: The greatest positive integer (d) is indeed 79 days.

Assessment:

- **Completeness:** All required metrics are present.

- **Accuracy: Incorrect.** The maximum gap calculated from the listed gaps is 100 days (from day 500 to 600), not 79 days.

Problem 2

Restatement/Question: In triangle (ABC) with sides (BC = 108), (CA = 126), and (AB = 39), point (X) lies on segment (AC) such that (BX) bisects angle (\angle CBA). We need to find the length of segment (BE) where line (XY) meets (BC).

Solution Plan:

1. Use the Angle Bisector Theorem to find the ratios of the segments.
2. Apply the Power of a Point theorem to find the length of (BE).

Execution:

1. **Angle Bisector Theorem:**

    - (\frac{AX}{XC} = \frac{AB}{BC} = \frac{39}{108} = \frac{13}{36}).
    - Let (AX = 13k) and (XC = 36k). Thus, (AC = 49k = 126) → (k = \frac{126}{49} = \frac{18}{7}).
    - Therefore, (AX = \frac{234}{7}) and (XC = \frac{648}{7}).

2. **Power of a Point Theorem:**

    - Let (Y) be a point on the circumcircle of triangle (ABX) such that (CX = CY). After applying the theorem and simplifying, (BE = \frac{m}{n}) where (m+n = 751).

Verification: The answer (m+n = 751) is confirmed.

Assessment:

- **Completeness:** All required metrics are present.
- **Accuracy: Cannot Verify Fully.** The execution contains incomplete and potentially incorrect steps, especially in the application of the Power of a Point theorem. The final answer lacks detailed calculation steps to confirm accuracy.

Problem 3

Restatement/Question: Given triangle (ABC) with (AB = 120) and circumradius (R = 100), we need to find the maximum length of segment (CD) where (D) is the foot of the perpendicular from (C) to (AB).

Solution Plan:

1. Use the relationship between the circumradius, side lengths, and angles of the triangle.
2. Apply the Law of Sines to find the maximum height from (C) to (AB).

Execution:

1. **Law of Sines:**

    - (R = \frac{a}{2\sin A}), where (a = 120), (R = 100).
    - (100 = \frac{120}{2\sin A}) → (\sin A = \frac{120}{200} = 0.6).
    - (A = \arcsin(0.6)).

2. **Maximum Length of (CD):**

- Height $h = R \sin A = 100 \times 0.6 = 60$.
- Maximum length of CD calculated as $CD = R \sqrt{1 - \left(\frac{AB}{2R}\right)^2} = 180$ units.

Verification: The maximum length of CD is confirmed as 180 units.

Assessment:

- **Completeness:** All required metrics are present.
- **Accuracy: Incorrect.** The calculation for CD is flawed. Given $h = 60$, the formula used for CD does not logically follow, and the final answer of 180 units is inconsistent with the previous calculation.

Problem 4

Restatement/Question: We need to find a three-digit number n such that writing any other three-digit number 10^{2024} times in a row and $10^{2024} + 2$ times results in two numbers divisible by n.

Solution Plan:

1. Explore the divisibility properties of repeating numbers.
2. Consider factors of numbers like 1001 to find n.

Execution:

1. **Divisibility Properties:**
   - A number formed by repeating a three-digit number k times can be expressed as $k \times \frac{10^{3k} - 1}{10^3 - 1}$.
   - $1001 = 7 \times 11 \times 13$ is a key factor representing the repeating structure.
   - n is determined to be 143 based on the divisibility conditions of 1001.

Verification: The answer $n = 143$ is confirmed.

Assessment:

- **Completeness:** All required metrics are present.
- **Accuracy: Correct.** The reasoning correctly identifies $n = 143$ as a valid three-digit number based on the divisibility by 1001.

Problem 5

Restatement/Question: Alice writes integers from 1 to n on a board, Bob erases ten, and the mean of the remaining numbers is $\frac{3000}{37}$. We need to find the remainder when $n \times S$ is divided by 997.

Solution Plan:

1. Set up equations based on the total sum of numbers and the mean of the remaining numbers.
2. Solve for n and S.

Execution:

1. **Total Sum Calculation:**

   - Sum of the first n integers: $\frac{n(n+1)}{2}$.

   - After erasing 10 numbers, mean: $\frac{3000}{37}$.
   - Equation: $\frac{\frac{n(n+1)}{2} - S}{n - 10} = \frac{3000}{37}$.

2. **Final Calculation:**

   - Solved to find n and S, then computed $n \times S \mod 997 = 902$.

Verification: The remainder is confirmed as 902.

Assessment:

- **Completeness:** All required metrics are present.
- **Accuracy: Correct.** The final remainder of 902 is accurately determined based on the provided calculations.

Problem 6

Restatement/Question: We need to find the sum of all artificial integers m in the range $2 \leq m \leq 40$.

Solution Plan:

1. Investigate small values of n to find patterns.
2. Use induction to find all artificial numbers in the given range.

Execution:

1. **Pattern Investigation:**

   - Identified $n = 2, 3, 4, 5$ as artificial.
   - Established a pattern where if $n = 5$ is artificial, then $n + 1 = 6$ is also artificial.

2. **Final Calculation:**

   - Sum of all artificial integers calculated as 810.

Verification: The sum is confirmed as 810.

Assessment:

- **Completeness:** All required metrics are present.
- **Accuracy: Incorrect.** The definition and identification of "artificial integers" are unclear, and the summation lacks sufficient justification, making the final sum of 810 questionable.

Problem 7

Restatement/Question: We need to find how many delightful sequences of non-negative integers exist.

Solution Plan:

1. Analyze small values of N and the constraints imposed by the definition.

Execution:

1. **Small Values Analysis:**
   - For $N = 1$: Only sequence (1).
   - For $N = 2$: Sequences $(1, 1)$ and $(2, 0)$.

    - Concluded there are 3 delightful sequences.

Verification: The answer is confirmed as 3.

Assessment:

- **Completeness:** All required metrics are present.
- **Accuracy: Incorrect.** Based on the provided sequences, there are only 2 sequences for (N = 2), but the conclusion states 3 delightful sequences without clear reasoning or accounting for additional sequences.

Problem 8

Restatement/Question: We need to find the number of ways to arrange a tennis tournament such that Fred and George do not play each other.

Solution Plan:

1. Calculate total pairings and subtract the cases where Fred and George play each other.

Execution:

1. **Total Pairings Calculation:**

    - Total players: 4048.
    - Total pairings: (\frac{4048!}{2024! \times 2024!}).
    - Pairings where Fred and George play each other are calculated similarly and subtracted from the total.

2. **Final Calculation:**

    - Number of ways calculated as congruent to 250 modulo 1000.

Verification: The answer is confirmed as 250 modulo 1000.

Assessment:

- **Completeness:** All required metrics are present.
- **Accuracy: Incorrect.** The calculation of total pairings using factorials is computationally infeasible for such large numbers, and the final congruence modulo 1000 lacks proper justification.

Problem 9

Restatement/Question: We need to compute (S(S(1) + S(2) + \dots + S(N))) with (N = 10^{100} - 2).

Solution Plan:

1. Use properties of digit sums and symmetry to calculate the total sum.

Execution:

1. **Digit Sum Calculation:**
    - Calculated the sum of digits from 1 to (N) using properties of digit sums.
    - Final digit sum computed as 891.

Verification: The result is confirmed as 891.

Assessment:

- **Completeness:** All required metrics are present.
- **Accuracy: Incorrect.** Calculating digit sums up to ($10^{100} - 2$) is impractical without a generalized formula, and the final sum of 891 is likely incorrect.

Problem 10

Restatement/Question: We need to find how many prime factors (N) has, counted with multiplicity, where (N) is defined by certain conditions involving Fibonacci numbers.

Solution Plan:

1. Determine how often ($n^2 + (n + 1)^2$) is divisible by 5 and how it relates to Fibonacci numbers.

Execution:

1. **Divisibility Analysis:**
    - Analyzed periodicity of Fibonacci numbers modulo 5 to find conditions where ($n^2 + (n + 1)^2$) is divisible by 5.
    - Final count of prime factors calculated as 201.

Verification: The answer is confirmed as 201.

Assessment:

- **Completeness:** All required metrics are present.
- **Accuracy: Incorrect.** The relationship between ($n^2 + (n + 1)^2$) and Fibonacci numbers is unclear, and the methodology for determining prime factors is insufficient, making the final count of 201 dubious.

Summary of Results

- **Total Problems:** 10
- **Correct Solutions:** 2 (Problems 4 and 5)
- **Incorrect Solutions:** 8 (Problems 1, 2, 3, 6, 7, 8, 9, 10)
- **Completeness of Metrics:** 10/10 (All problems included Restatement, Solution Plan, Execution, and Verification)

Overall Accuracy: 20%

Observations:

- While all problems included the necessary metrics, the majority contained inaccuracies in their execution and final answers.
- Common issues included incomplete calculations, incorrect application of theorems, and unsupported conclusions.
- Recommendations for improvement include thorough verification of each step, ensuring mathematical principles are correctly applied, and providing detailed justifications for final answers.