1. **Results Summary**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Test Case | Expected Output | ChatGPT Result | AgentGPT Result | ChatGPT Time (s) | AgentGPT Time (s) | Correctness | Performance |
| 1 | 22 | 12 | 8 | 0.000018 | 0.000007 | Both Failed | AgentGPT is faster |
| 2 | 4 | 9 | 11 | 0.000019 | 0.000006 | Both Failed | AgentGPT is faster |
| 3 | 48 | 0 | 0 | 0.000034 | 0.000013 | Both Failed | AgentGPT is faster |
| 4 | 48 | 24 | 8 | 0.000008 | 0.000004 | Both Failed | AgentGPT is faster |

1. **Observations**
   1. *Correctness*
      * Both ChatGPT and AgentGPT failed to produce the correct results for all four test cases. This indicates that neither solution successfully implements the problem logic as required by the specification.
   2. *Performance*
      * AgentGPT consistently outperformed ChatGPT in terms of execution time, being approximately **2-5x faster** across all test cases. This efficiency is likely due to differences in the implementation logic and computational steps.
2. **Conclusion**

While AgentGPT demonstrated superior performance in terms of execution time, both solutions failed to produce correct outputs. Future efforts should focus on refining the models to ensure correctness.