Rating Criteria

As a senior computer vision and NLP researcher, and an impartial data quality assessor, please evaluate the quality of the generated caption in relation to the given image on a scale of 0 to 10. Each aspect has an upper bound which has been marked as follows:

- 1. Authenticity (4 points): The answer should not describe an object or aspect that doesn't exist in the image.
- 2. Correctness (2 points): Does the answer correctly identify all the objects present in the image?
- 3. Detail (2 points): How well does the answer describe each object and its attributes?
- 4. Coherence (1 points): Is the answer logically consistent and understandable, maintaining a steady flow throughout?
- 5. Completeness (1 points): Does the answer cover all aspects in the image?

Give lower points if the answer fails in any of these aspects. A score of 10 indicates that the generated answer is accurate, detailed, coherent, complete and does not hallucinate any details not present in the image. The final score for the answer is the sum of scores in each aspect. Please also offer your reason for this score, and output the dict as the answer.

Input Image



Strong LLMs





GPT-40 Gemini

Scores and Reasons

"Correctness": 1, "Authenticity": 3, "Detail": 1. "Coherence": 1.

"Final Score": 7, "Completeness": 1, "Reason": "The monster holds only one guitar, not two. It grips a small drum with one hand."

Generated Answers

The scene takes place in someone's garden, featuring a small statue of a crunchy red monster with purple limbs standing outside. The monster's hat covers one set of its big eyes as it is a garden ornament. It uniquely appears to play a musical role, holding not one but two guitars. One guitar can be seen hanging in front of the statue on its lap or in one hand, while the other one can be picked up at its side. The overall arrangement suggests the monster might be a part of a whimsical and artistic garden setup with possibly some entertainment attached.