

Functional Roles

Confirmation of Photo: The sentence confirms what the user uploaded. When the user asks an identification question (e.g., *what is this?*), this sentence may also be annotated as an answer.

*This usually comes at the beginning of the sentence, to provide the overview of the image. This sentence often looks similar to typical image captions.

- Example: *The image you've provided appears to be of a SodaStream Raspberry flavor syrup bottle.*

Answer: The sentence directly addresses the question. If the answer is provided in multiple sentences, they can all be labeled as “answer” as in the example below. Incorrect answers are still labeled as answers.

- Example: (question: what color is this?)
 - Answer: *I would describe the shirt as a reddish-brown color.*
 - Answer: *This also has a white logo.*

Answer Failure: The sentence states the inability to address the question, often accompanied by an “Explanation of Reasoning” explaining the reason.

- Example: (question: what is this box?) *The "box" you're referring to is not clearly visible or identifiable in the picture, as the focus seems to be on the paper towel roll in the foreground.*
- Example: *I cannot provide information such as the details on this globe.*

Auxiliary Information: The sentence provides additional information not directly related to the question but observed in the photo, or general knowledge or facts related to the query.

- Example: (What is the color of the pants?)
Answer: *You are wearing a black pair of pants.*
Auxiliary Information: *You also seem to be wearing a red shirt with something brown underneath.*
- Example: (question: Read this label)
Answer: *The label on the container appears to read "RealFruit Instant Pectin" and is likely used for making homemade jams or preserves.*
Auxiliary Information - Non-Visual: *Pectin is a natural thickening agent that's extracted from fruits and used commonly in cooking to gel liquids.*

Auxiliary Information Failure: The sentence states the inability to provide auxiliary information not directly related to the question.

- Example: (What is the color of the pants?)
- Answer: *You are wearing a black pair of pants.*
Auxiliary Information Failure: *However, I cannot identify its brand name because the image is blurry.*

Explanation of Reasoning: The sentence explains the reasoning for the information it gives by describing its thought process or providing the information source. It can also explain the reasoning behind failure (e.g., low quality of the photo).

- Example: (question: what is this?) *Given the color of the liquid visible through the bottle, it's likely to be a type of mustard sauce, possibly a honey mustard dressing or similar condiment based on its golden-brown hue.*
- Example: (question: what is this box?) *The "box" you're referring to is not visible or identifiable in the picture, as the focus seems to be on the paper towel roll in the foreground. If you're referring to something specific in the background, the details are too obscured to make out clearly.*

Suggestion: The sentence suggests users retake the photo, provide additional information, or try a different method or strategy to find the answer. It may also advise caution or exceptions.

- Example: *If you need the exact instructions, you might try using a smartphone to take a clearer picture of the instructions on the box.*
- Example: *If you need information about a specific item, a more detailed description might help.*
- Example: *If you need the exact instructions, you might try looking up the product online.*
- Example: *Keep in mind that these instructions are general and may not be specific to the Marie Callender's pie you have. It is always best to follow the specific instructions provided by the manufacturer for the best results.*

Misc: The sentence does not provide new information. Examples can be sentences that confirm what the user asked, organizational sentences (explaining how the answer will be structured before elaborating), and final remarks. Any sentences that do not belong in any of the categories above.

- Example: *You mentioned looking for the input or mode button.*
- Example: *Here, I list all the items that I see on the image.*
- Example: *I'm happy to assist you, please let me know if you have any other requests!*

Information Modality

Image Content: The sentence provides the visual information about the image content (e.g., setting, object, text, color, atmosphere, etc)

- Example: *The vehicle in the image appears to be a utility service truck, commonly used by contractors or maintenance departments.*
- Example: *You are holding a pink bottle. Only some text on the bottle is visible but it is an Avon product. Some of the text on the bottle is in French.*

Image Quality: The sentence provides the information about the image quality.

- Example: *The image you've provided is quite blurry and the details are not clear, which makes it difficult to identify the jar's contents with certainty.*

External Information: The sentence offers general knowledge or facts that are not found from the image. It enriches the response with broader contextual or background information.

- Example: *The vehicle in the image appears to be a utility service truck, commonly used by contractors or maintenance departments.*
- Example: *Expiration date is typically located near the top or side of the bottle or on the label itself.*

*Multiple types of information can be provided in a single sentence