LLaVA-Instruct-150K: Language-only In-context



In-context:

Instruction: What type of traffic
device can be seen in the image?
Response: There is a street sign and

a traffic light above a city road.

Query:

Instruction: What is the current color of the traffic
lights in the image?

Response: All the traffic lights in the image are showing the color red at the same time.

MIMIC-IT: Multi-modal In-context
Up->Down: LLaVA-Interleaved->TV Show Captions



Instruction:What the reason for the
dog's current expression and outfit?
Response:... wearing a festive
Christmas hat and possibly panting, ...,
making a cute and memorable moment
with their pet during the holiday
season.



Instruction: Why teddy bear ornament on the Christmas tree?
Response: The teddy bear ornament hanging on the Christmas tree conveys a sense of ... children and adults during the festive season.



Instruction: What task is the woman performing?
Response: The woman is vacuuming the rug area between chairs and couches.



Instruction: At what time of day is?

Response: As the bedside lam is opening and the light in this room is dark, it is in the night.

MIMIC-IT: Language-only In-context Up->Down: Ego4D->Spot The Difference











Instruction: What color of jersey does my teammates wear?
Response: Black, as you pass the ball to a player
wearing black jersey.

Instruction: Does our team goal?

Response: Unfortunately, does not. The The goalkeeper saves the ball at the end



Instruction: What is the difference between these two images?
Response: The first one shows an outhouse with a toilet, while the second one shows...

Instruction: Are there any objects in image two but
not in image one?

Response: Yes, there are additional objects in image two that are not present in image one. Image two has a white sink, a bathtub, a mirror, and other bathroom items, which are not mentioned in image one.