LFVQA COLM 2024 **[Spotlight]**

Long-Form Answers to Visual Questions from Blind and Low Vision People

Mina Huh, Fangyuan Xu, Yi-Hao Peng, Chongyan Chen, Hansika Murugu, Danna Gurari, Eunsol Choi, Amy Pavel

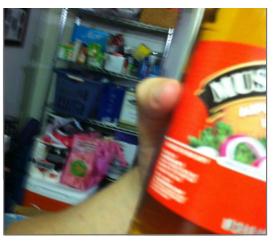
VizWiz-LF Dataset

Our dataset *VizWiz-LF* contains **4.2k** long-form answers (*avg.* **30 words**) to visual questions, collected from **human expert** describers and **6 VQA models**.

Annotation

Visual Question

What is in this bottle?

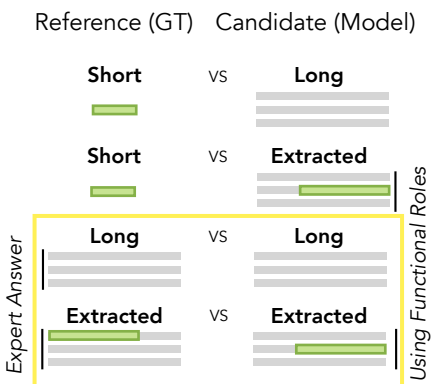


Short Answer

Tomato sauce

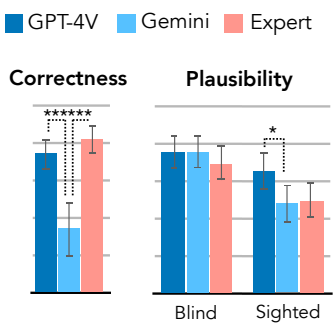
	Long-Form Answers	Functional Roles	Information Source
GPT-4V	The image is blurry, but it shows a bottle with a label that includes the word MUSTARD.	confirmation	image content, image quality
	Given the color of the liquid visible through the bottle, it's likely to be a type of mustard sauce [...]	answer, explanation	image content
	Mustard sauces are commonly used as a condiment or dressing for sandwiches, salads, and many other dishes.	auxiliary information	external
	If you have more information or a clearer picture, I could attempt to provide a more accurate identification.	suggestion	
Gemini	This bottle contains Heinz Distilled White Vinegar.	answer	image content
Human	I can't tell what is in the bottle.	answer failure	image content
	If I had to guess, I would say it was some sort of cooking oil because it has an amber color and pictures of vegetables [...]	answer, explanation	image content
	It could also be an apple juice or another type of juice.	answer	image content

Automatic Evaluation



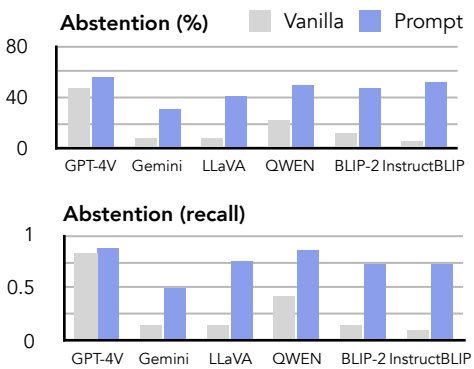
- *Evaluation using **long reference answers** shows a better correlation with human*
- ***LLM metric** > ROUGE/METEOR/BERTScore*

Human Evaluation



- *Model answers hallucinate with incorrect visual details, but **blind people often perceive them as plausible***
- *Sighted people's evaluation is not a strong proxy for blind people*

Abstention Experiment



- *We can use **functional roles** to check if a model's long-form answer abstained.*
- ***Prompting models with abstain instructions** can reduce hallucinations.*