

B APPENDIX B

The GPT prompt was designed based on all the ambiguities presented in [12]. The prompt used in this study is as follows:

- 1) They select an image and you will get its caption.
- 2) You will provide relevant concepts according to your guess about their potential intent. These concepts include locations, entities (including objects and living things), emotions, activities(verbs in 'ing' form), occasions, and relational roles(Human-to-Entity Relations). I need a diverse set of 10 word(phrases) for each concept. By the way, entities you are generating fall under different categories. Please provide additional entities in the ****same categories****, ****higher-level categories**** (more general, like 'furniture' for 'chair'), and ****lower-level categories**** (more specific, like 'dining chair' for 'chair'). I need 10 word(phrases) for each of them.
- 3)Next, generate 10 sentences based on the 10 activities. They must represent actions a person might want to take with entities from the image context. They must start with 'a person is ...'. For example, if the image includes fruits, you might use phrases like 'a person is slicing fruits and vegetables.'
- 4) Next, infer the person's intent for selecting that image based on these concepts and your understanding of the person. Generate five different phrases representing possible intents the person may wish to communicate. Format your response as a JSON object with the following structure:

```

"locations": [<location>; <location>],
"entities": [<entity>; <entity>],
"same-level entities": [<entity>; <entity>],
"high-level entities": [<entity>; <entity>],
"low-level entities": [<entity>; <entity>],
"activities": [<activity>; <activity>],
"occasions": [<occasion>; <occasion>],
"Relational Roles": [<role>; <role>],
"sentences": [<sentence>; <sentence>],
"intents": [<intent>; <intent>].

```

My system will use your answers to retrieve more images and show them to help them clarify their intent. Do not generate repetative concepts in the whole set. Only provide the correct format of JSON with no additional text before or after the JSON object.

The caption is:"+image caption