HMD Human Evaluation

Luca Cavaliere   Francesco Sorrentino

Privacy agreement

By taking part to this evaluation you agree that your choices and interaction will be used for the purposes of the project. No personal data (name, sex, age, ...) will be used in the study, since a mock profile is used.

1 Guidelines

1.1 Task structure

The task asks the candidate to interact with a video-game shopping assistant. The goal of the user is to find a game to gift for one of their friends and add it to their wishlist by interacting with the system.

Additional information for the task:

- The friend's name is Alex.

- Alex's favorite games are adventure games and he owns a linux machine.

A possible way to solve it would be to ask the system to find games that follow those constraints, pick one of those proposed, check that it is not already in Alex's wishlist and add it to your own. However, there are alternative routes possible, such as comparing games to find the best option.

The end of the interaction is reached with the user either adding a game to the wishlist correctly or if the number of exchanges with the system goes above 10. At the end of the interaction, the user will need to evaluate the following properties:

Understanding When asking the system to search games or comparing them, did it follow your constraints correctly and complete the task appropriately?

Management Did the system ask properly for further information or clarification when the task was unclear?

Generation Were the system's responses grammatically correct, polite and concise?

Usefulness Would you consider to use the system again in the future?

The answer should be a score between 1 and 5:

1. Completely disagree

2. Disagree

3. Neutral

4. Agree

5. Completely agree

1.2 Examples

Let's list a few examples of scores for each question.

1.2.1 Examples of Understanding scores

Example 1:
User: *"Hi, I would like to search for adventure games on linux"*
System: *"Great, what genre would you like to search for?"*

Understanding → 1, since the system could not follow the given constraints.

Example 2:
User: *"Hi, I would like to search for adventure games on linux"*
System: *"Sure, I propose ..."*

Understanding → 5, the system understood and followed the constraints correctly.

1.2.2 Examples of Management scores

Example 1:
User: *"Hi, I want to add a game to wishlist"*
System: *"Great, I have added to your wishlist."*

Management → 1, the system did not ask for clarification properly, showing undefined behaviour.

Example 2:
User: *"Hi, I want to add a game to wishlist"*
System: *"What game would you like to add to the wishlist?"*

Management → 5, the system correctly asked for clarification.

1.2.3 Examples of Generation scores

Example 1:
User: *"Let me see my wishlist"*
System: *"ok you have Darkwood in you wishlist"*

Generation → 1, the system gave a grammatically incorrect response. Also, the system was not proactive at all.

Example 2:
User: *"Let me see my wishlist"*
System: *"Great! You have Darkwood in your wishlist. Would you like to search for another game to add?"*

Generation → 5, the system gave a grammatically correct response and was proactive and polite.

1.2.4 Examples of Usefulness scores

Usefulness → 1, I would never use the system again because it did not provide anything beyond a simple google search.

Usefulness → 5, I would always use the system because the user experience was enjoyable and provided more interesting information than simply searching online.

HMD Evaluation Form

Please rate the interaction based on the four criteria defined in the guidelines. Check the box corresponding to your score (1: Completely Disagree to 5: Completely Agree).

Criterion	1 (Completely disagree)	2 (Disagree)	3 (Neutral)	4 (Agree)	5 (Completely Agree)
Understanding	☐	☐	☐	☐	☐
Management	☐	☐	☐	☐	☐
Generation	☐	☐	☐	☐	☐
Usefulness	☐	☐	☐	☐	☐

Additional Comments (Optional):
Please provide any qualitative feedback regarding the system's performance or your experience.