
You are writing critiques for a submitted response on a given user’s query. Here is the data:

```
[BEGIN DATA]
***
[Query]: {query}
***
[Response]: {response}
***
[END DATA]
```

Here are the instructions you should follow:

1. You only need to write critiques on its shortcomings; there’s no need to comment on its strengths. The critiques should be as specific as possible by quoting details from the response and query, and don’t include the criterion name.
-

Table 12: Prompt used when collecting raw output for single-response evaluation from GPT-4.

Write a meta-critique by combining the following two critiques for a submitted response on a given user’s query, and grade the response:

```
[BEGIN DATA]
***
[Query]: {query}
***
[Response]: {response}
***
[Critique 1]: {critique 1}
***
[Critique 2]: {critique 2}
***
[END DATA]
```

You should give a meta-critique by merging the two critiques into a more comprehensive critique for the response in fluent language. After that, you should give a final rating for the response on a scale of 1 to 10 by strictly following this format: "[[rating]]", for example: "Rating: [[5]]".

Table 13: Prompt for asking GPT-4 to combine two critiques to a comprehensive critique and give out final rating.

G COMPLETE RESULTS AND CASES

Tab. 23 contains the complete comparison results of Fig. 3 (§6.2).

Tab. 24-28 provide the comprehensive details of §6.4. Tab. 24 shows the detailed ranking of Fig. 5. The complete cases of §6.4 are shown in Tab. 25-28.