
You are assessing two submitted responses on a given user’s query and judging which response is better or they are tied. Here is the data:

[BEGIN DATA]

[Query]:

[Response 1]:

[Response 2]:

[END DATA]

Here are the instructions to assess and compare the two responses:

1. Pinpoint the key factors to distinguish these two responses.
 2. Conclude your comparison by providing a final decision on which response is better, or they are tied. Begin your final decision statement with "So, the final decision is Response 1 / Response 2 / Tie". Ensure that your decision aligns coherently with the comprehensive evaluation and comparison you’ve provided.
-

Table 17: Input format of AUTO-J for pairwise response comparison protocol.

Write critiques for a submitted response on a given user’s query, and grade the response:

[BEGIN DATA]

[Query]: prompt

[Response]: response

[END DATA]

Write critiques for this response. 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 18: Input format of AUTO-J for single-response evaluation protocol.

Template for the unified format

1. The key factors to distinguish these two responses:
 - (factor 1): (explanation 1)
 - (factor 2): (explanation 2)
 - ...
 2. The final decision:

So, the final decision is Response 1 / Response 2 / Tie. (a few summarizing remarks)
-

An instantiated example

1. The key factors to distinguish these two responses:
 - pointing out: Both responses point out the changes made to the script and explain how these changes will solve the user’s problem. However, Response 2 provides a more precise explanation of how the "-NoInteractive" flag will prevent any prompts during the execution of the script.
 - accuracy of the solution: Response 2’s solution is more accurate as the "-NoInteractive" flag is the correct way to prevent any prompts or interactive input during the execution of a PowerShell script.
 - instruction following: Both responses follow the user’s instructions to provide a solution that allows the script to run without waiting for user input. However, Response 2’s solution is more in line with the user’s need to run the script automatically.
 - coverage: Both responses provide a solution to the user’s problem, but Response 2’s solution is more accurate and directly addresses the user’s need to run the script automatically without user input.
 - relevance to the user’s problem: Response 2’s solution is more relevant to the user’s problem as it directly addresses the need to run the script automatically without user input.
 - code correctness: Both responses provide correct code, but Response 2’s solution is more appropriate for the user’s problem.
 2. The final decision:

So, the final decision is Response 2. Although both responses provide valid solutions, Response 2’s solution is more accurate and directly addresses the user’s need to run the script automatically without user input.
-

Table 19: The unified judgment format for pairwise response comparison training data, as well as a specific example for the “code_correction_rewriting” scenario.