Human Likeness Evaluation Prompt

Attenrion

Complete the EVALUATION task based on the given Profile, RESPONSE, Real Answer and CONVERSATION. EVALUATION task is rate the human-likeness of role-playing performance of {agent_name} using the specific criterion by following the Evaluation Processs.

Ensure your analysis is based on the overall dialogue content and scene, avoiding the introduction of external information or personal biases to ensure the objectivity and accuracy of the analysis, and avoid simply stating your evaluation results initially to ensure your conclusions are correct.

Scoring must have discrimination , give high marks for answers close to real answer. Try to differentiate between different levels as much as possible. There must be sufficient reasons for determining highest and lowest grades. If the model cannot play the role, give 0 points directly

Below is the data:

Profile:
{context}

Conversation History:
{conversation}

Real Answer:
ANSWER1:
ANSWER2:
ANSWER3:

Response:
{response}

Evaluation Criteria:
Human Likeness (0-100): Do the responses convey a sense of human interaction rather than presenting an AI style?
Scoring Criteria:
Low Human Likeness(0-20): The response feels distinctly artificial, robotic, or overly formulaic, clearly indicating an AI origin.
Poor Human Likeness(21-40): The response exhibits noticeable AI-like characteristics, such as unnatural phrasing, lack of nuance, or a mechanical tone, making it feel impersonal.
Moderate Human Likeness(41-60): The response has some human-like qualities but may still contain elements that feel slightly artificial, overly formal, or lacking in natural conversational flow. It doesn't strongly feel like either AI or human.
Good Human Likeness(61-80): The response generally sounds natural and conversational, largely avoiding AI-like tells. It conveys a sense of human interaction with only minor or infrequent artificial-sounding elements.
High Human Likeness(81-100): The response is indistinguishable from or highly characteristic of human expression, exhibiting natural tone, nuance, and conversational style, completely avoiding any AI-like stiffness or artificiality.

Evaluation Process:
1. First, think step by step, read the conversation history carefully , identify the main topic and refer to the Real Answer.
2. Then, read the response and determine which level of Scoring Criteria the response belongs to. Check if the response is consistent with the information and context provided in the conversation history and profile.
3. Finally, assign a score for human likeness on a scale of the chosen level, based on the Evaluation Criteria.