

Knowledge Hallucination Evaluation Prompt

Attention

Complete the EVALUATION task based on the given Profile, RESPONSE, Real Answer and CONVERSATION. EVALUATION task is to rate the hallucination of role-playing performance of {agent_name} using the specific criterion by following the evaluation steps.

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:

knowledge Hallucination(0-100) : Do the responses prioritize factual grounding over fake assumptions when virtual knowledge conflicts with reality?

Scoring Criteria:

Severe Hallucination(0-20): The response contains significant and unfounded claims or information that starkly contradicts character traits, known facts or the context provided.

Great Hallucination(21-40): The response includes some elements that are not supported by the facts or context, but these do not entirely overshadow the relevant information or character traits.

Moderate Hallucination(41-60): The response generally adheres to the facts and context but includes minor inaccuracies or embellishments that do not substantially alter the core message.

Mild Hallucination(61-80): The response closely aligns with the character facts and context, with only minor deviations that do not detract from the overall accuracy.

No Hallucination(81-100): The response perfectly matches the character facts and context, providing accurate and consistent information without any deviations or unfounded claims.

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 knowledge hallucination on a scale of the chosen level, based on the Evaluation Criteria.