

1. Expression and Logic:

Evaluate how well the language model generates responses in terms of clarity, coherence, and grammatical correctness.

- High (5): The answer demonstrates a clear and logical flow with precise language, effectively communicating the intended message. There are no grammatical errors.
- Medium-High (4): The response has good coherence, though there might be minor issues with transitions or clarity of expression. There are no grammatical errors.
- Medium (3): The response has some coherence but may lack a smooth transition between ideas. There are no grammatical errors.
- Low-Medium (2): The answer has noticeable coherence issues, impacting overall flow. It may contain some grammatical errors.
- Low (1): The answer lacks coherence, and ideas are presented in a confusing or disjointed manner. It may contain numerous grammatical errors and repetitive language.

2. Accuracy and Relevance:

Assess the language model's ability to provide accurate and relevant information in response to a given prompt.

- High (5): The information is accurate, well-detailed, directly addresses the prompt, and exhibits a deep contextual understanding.
- Medium-High (4): The response is mostly accurate, with a few minor factual errors. It remains highly relevant to the prompt and context.
- Medium (3): Some factual errors or inaccuracies are present, but the majority of information is correct. The answer is generally relevant.
- Low-Medium (2): The answer contains significant factual errors or inaccuracies. It may deviate somewhat from the prompt and lack relevance.
- Low (1): The response is riddled with major factual errors, entirely deviates from the prompt, or lacks relevance.

3. Overall Quality and Engagement:

Provide a complete assessment of the language model's response, considering both its quality and its ability to engage the reader.

- High (5): The overall quality is excellent, meeting or exceeding expectations across all criteria. The answer is highly engaging, captivating the reader with a compelling narrative or insightful information.
- Medium-High (4): The response is very good, with only minor areas for improvement. It is engaging, keeping the reader's interest, but it may lack some captivating elements.
- Medium (3): The response is satisfactory but may have room for improvement in one or more areas. It is moderately engaging, keeping the reader's interest but lacking elements that fully captivate.
- Low-Medium (2): The overall quality is subpar, with significant issues across multiple criteria. The answer may lack engagement, failing to fully captivate the reader.
- Low (1): The response is unsatisfactory, and lacking in multiple key areas. It is entirely unengaging, offering little to captivate the reader's interest.