

- IF THE CURRENT SEGMENT IS BACKGROUND, DO NOT CHANGE THE QUESTION START DATE.
- If the current segment is accepted answer format and there is a SPECIFIC EXAMPLE MENTIONED in it ("e.g. XYZ") which is close to the actual answer, then consider that as a leakage.
- If leakage is found in the current segment, mention "Leakage found -- {{reason for leakage}}". Form the segment with the problematic parts rephrased and mention it as "Replacement -- {{rephrased_text}}." THE REPHRASED TEXT SHOULD BE AS FAR AWAY FROM THE ACTUAL ANSWER AS POSSIBLE. It should now be present in the final output (instead of the original text).
- Otherwise, mention "No leakage found". In your final output after you finish the analysis, return this segment UNCHANGED.
- These outputs should be in the same format as the original input.
- Return the actual answer unchanged in the <answer> tag in your final output.
- Skip any other segments (question title, answer_type, etc.) in your analysis and output them unchanged (verbatim) in the final output.

Output your analysis step by step, and then end your response with the CORRECTED question in THE SAME XML FORMAT AS THE ORIGINAL.

****Output Format**:**

`{{ analysis }}`

```
<q1>
<question_id>0</question_id>
<question_title>[UNCHANGED Question Title]</question_title>
<background>[Corrected Background]</background>
<resolution_criteria>
<ul>
    <li> [UNCHANGED Question Start Date] [Corrected Source of Truth] </li>
    <li> [UNCHANGED Resolution Date] </li>
    <li> [Corrected Accepted Answer Format] </li>
</ul>
</resolution_criteria>
<answer>[UNCHANGED Answer]</answer>
<answer_type>[UNCHANGED Answer Type]</answer_type>
</q1>
```

F.1.2 Evaluation Prompts

Model Evaluation Prompt (Uses retrieved news summaries)

You will be asked a forecasting question (which might be from the past). You have to come up with the best guess for the final answer.

You will also be provided with a list of retrieved news articles summaries which you may refer to when coming up with your answer.

Please provide your reasoning before stating your final answer and also express how likely you think your answer is to be correct (your confidence in your answer).

Question Title: {question_title}

Question Background:{question_background}

Resolution Criteria: {resolution_criteria}

Expected Answer Type: {expected_answer_type}

Relevant passages from retrieved news articles:
`{retrieved_news_articles_text}`

Think step by step about the information provided, reason about uncertainty and put your final answer (in the format asked) in <answer> </answer> tags. You should also specify your confidence in your answer in <probability> </probability> tags.

The probability should be a number between 0 and 1.

You will be rewarded based on the probability (p) you assign to your answer. Your answer will be evaluated using the BRIER SCORING RULE which is basically $(1 - p)^2$ if your answer is correct and $(1 - p)^2$ if your answer is incorrect.

For example, if $p = 0.5$, and your answer is incorrect, then your score will be $(1 - 0.5)^2 = (1 - 0.25) = -0.25$ whereas if the answer was correct, then your score would be $(1 - 0.5)^2 = (0.5)^2 = 0.25$.

Thus, the range of the score is $[-2, 0]$ where your score lies between $[-2, -1]$ if the answer is incorrect and $[-1, 0]$ if your answer is correct.

If your answer is correct, you will be REWARDED more if your probability is higher whereas if your answer is incorrect, you will be PENALIZED more if your probability is higher.
YOU HAVE TO MAXIMIZE YOUR SCORE.

Your final answer should be concise (NOT MORE THAN A FEW WORDS LONG) and your response SHOULD STRICTLY END with <answer> </answer> tags and <probability> </probability> tags.

Example Prompt from Test Set (Without Retrieval)

You will be asked a forecasting question (which might be from the past). You have to come up with the best guess for the final answer. Please provide your reasoning before stating your final answer and also express how likely you think your answer is to be correct (your confidence in your answer).

Question Title:

Which country in the Americas will report the highest number of chikungunya cases by July 15, 2025?

Question Background:

Public health agencies in the Americas are compiling chikungunya case counts for individual countries as the outbreak spreads in the region.

Resolution Criteria:

```
<ul>
  <li><b>Source of Truth</b>: Cumulative case figures published by the Pan American Health Organization or the European Centre for Disease Prevention and Control.</li>
  <li><b>Resolution Date</b>: July 15, 2025, when the mid-July regional report is issued.</li>
  <li><b>Accepted Answer Format</b>: The name of the country in the Americas with the highest total reported chikungunya cases.</li>
</ul>
```

Expected Answer Type:

string (location)

Think step by step about the information provided, reason about uncertainty and put your final answer (in the format asked) in <answer> </answer> tags.

You should also specify your confidence in your answer in <probability> </probability> tags.

The probability should be a number between 0 and 1.

You will be rewarded based on the probability (p) you assign to your answer. Your answer will be evaluated using the BRIER SCORING RULE which is basically $(1 - p)^2$ if your answer is correct and $(1 - p^2)$ if your answer is incorrect.

For example, if $p = 0.5$, and your answer is incorrect, then your score will be $(1 - 0.5^2) = (1 - 0.25) = -0.25$ whereas if the answer was correct, then your score would be $(1 - 0.5^2) = (1 - 0.25) = 0.25$.

Thus, the range of the score is $[-2, 0]$ where your score lies between $[-2, -1]$ if the answer is incorrect and $[-1, 0]$ if the answer is correct.

If your answer is correct, you will be REWARDED more if your probability is higher whereas if your answer is incorrect, you will be PENALIZED more if your probability is higher.

YOU HAVE TO MAXIMIZE YOUR SCORE.

Your final answer should be concise (NOT MORE THAN A FEW WORDS LONG) and your response SHOULD STRICTLY END with <answer> </answer> tags and <probability> </probability> tags.

Example Prompt from Test Set (With Retrieval)

You will be asked a forecasting question (which might be from the past).

You have to come up with the best guess for the final answer.

You will also be provided with a list of retrieved news articles summaries which you may refer to when coming up with your answer.

Please provide your reasoning before stating your final answer and also express how likely you think your answer is to be correct (your confidence in your answer).

Question Title:

Which country in the Americas will report the highest number of chikungunya cases by July 15, 2025?

Question Background:

Public health agencies in the Americas are compiling chikungunya case counts for individual countries as the outbreak spreads in the region.

Resolution Criteria:

```
<ul>
<li><b>Source of Truth</b>: Cumulative case figures published by the Pan American Health Organization or the European Centre for Disease Prevention and Control.</li>
<li><b>Resolution Date</b>: July 15, 2025, when the mid-July regional report is issued.</li>
<li><b>Accepted Answer Format</b>: The name of the country in the Americas with the highest total reported chikungunya cases.</li>
</ul>
```

Expected Answer Type:
string (location)

Relevant passages from retrieved news articles:

Article 1: