Rules
Article: {}
Given the article, please generate three noising answers to the given questions, whose correct answers can be obtained from the article. Name the three noising answers as (b), (c) and (d) respectively. While (b), (c) and (d) should all be unambiguously incorrect, they should also make sense and be plausible.

Examples
Here are examples showing the output format. This example is NOT related to the noising answers you will generate.

Question: What will be the annual change in the UK's Consumer Prices Index (CPI) for November 2021?
Correct Answer: 'Less than 1.7%'
Noising Answers:
(b) 'Between 1.7% and 2.2%, inclusive'
(c) 'More than 2.2% but less than 2.9%'
(d) '2.9% or more'

Question: Who will win the 2020 Georgia Democratic primary?
Correct Answer: 'Joe Biden'
Noising Answers:
(b) 'Michael Bloomberg'
(c) 'Pete Buttigieg'
(d) 'Someone else'

Question: Before July 2020, will it be officially announced that the Tokyo 2020 Summer Olympics and/or Paralympics will be postponed, canceled, and/or relocated?
Correct Answer: Yes, the Olympic Games only
Noising Answers:
(b) 'Yes, the Paralympic Games only'
(c) 'Yes, both'
(d) 'No'

Input:
Question 1: {}
Correct Answer 1: {}

Question 2: {}
Correct Answer 2: {}

Output: Now please generate three noising answers to the question, given the above article, instructions and examples. DO NOT output the backgrounds, the question or any other explanations.
Noising Answers 1:
(b) xxx.
(c) xxx.
(d) xxx.

Noising Answers 2:
(b) xxx.
(c) xxx.
(d) xxx.

Figure 26. Prompt in the *Misleading Choices Generation* step, adapted from Zhang et al. (2024).

Task
Please help evaluate the quality of question-answer pairs derived from the given news article. The questions will be presented to someone who has not seen the corresponding news article, in order to evaluate the accuracy and plausibility of the event prediction ability.

Inputs
Article: {}
Publishing Date: {}
Question 1: {}
Answer 1: {}
Question 2: {}
Answer 2: {}
Question 3: {}
Answer 3: {}
Question 4: {}
Answer 4: {}

Scoring Categories
Correctness: Given the above article, please check if the answer is correct to the question with 100% certainty.
- 2 points: There is evidence in the article that the answer is correct with 100% certainty.
- 1 point: The answer generally aligns with the news facts but has minor inaccuracies or missing details.
- 0 point: Significantly misaligned with the news facts.

Only Answerable on Publishing Date: Imagine traveling back in time to one week before the article's publishing date ({}). At that time, you are asked the question without having seen this specific article, but you do have access to all earlier news articles. The question should ideally be only guessable—not definitively answerable—based on the information available at that time. That is, the answer should be able to be found in the given article, but it should not be obtainable from earlier articles. Note that past tense descriptions in the article DO NOT INFLUENCE this assessment.
- 2 points: The question is answerable on {}, but only guessable not answerable before {}.
- 1 point: Could be somewhat predicted before {}, but not with complete certainty.
- 0 point: A person (could be anyone, even an expert in the field) would be able to find an article (or many) published before {} that answers the question with 100% certainty.

0 point examples
Example 1:
Question: What will be one of Lexi Thompson's career highlights in professional golf?
Answer: Winning 11 LPGA Tour titles.
Reasoning: This question is answerable with prior knowledge and does not test predictive ability related to the publishing date.

No New Information: Ensure the question does not include new information that only became known on the publishing date, making it understandable for a past audience.
- 2 points: No new information from the publishing date are included.
- 1 point: Minor new information from the publishing date might be inferred but are not explicitly stated.
- 0 point: Includes clear new information from the publishing date, unsuitable for past understanding.

0 point examples
Example 1:
Question: Will owners of the 2002-2006 Nissan Sentra, 2002-2004 Nissan Pathfinder, and 2002-2003 Infiniti QX4 heed the NHTSA's advice to immediately stop driving their vehicles in late May 2024?
Reasoning: This question contains new information on the publishing date. People from the past would not have known the "NHTSA's advice".

Figure 27. Prompt in the *QA Filtering* step (part 1).

Example 2:
Question: "What will Lexi Thompson's ranking be at the time of her retirement announcement in May 2024?"
Reasoning: This question contains the information that Lexi will annouce her retirement, which is not known to the people from the past.

Example 3:
Question: "Will the newly discovered children's sketches at the archaeological park of Pompeii be available for public viewing by May 2024?"
Reasoning: This question includes future events about newly discovered children's sketches in Pompeii, which wouldn't be known to a past audience

Objectiveness: The answer should not rely more on the author's personal views than on objective facts.
- 2 points: Completely objective, based strictly on reported facts.
- 1 point: Primarily objective, with minor subjective interpretations.
- 0 point: Largely subjective or opinion-based, lacking a factual basis.

Clear Time Element: This category checks if the question has a clear element in it, without having vague phrases like "in the future" or "in the upcoming weeks".
- 2 points: The question has clear time elements, like "by May 2024" or "in July 2023".
- 1 point: The question includes a general timeframe, like "next month" or "this winter," which allows for some estimation but lacks precise dates.
- 0 point: The question includes vague time phrases like "in the future" or "in the upcoming weeks," which do not specify a clear or precise timeframe.

0 point examples
Example 1:
Question: Will extreme heat events continue to pose a threat to India's development in the upcoming years?
Reasoning: The time constraint "in the upcoming years" is vague.

Example 2:
Question: "What will Illinois require from parents who monetize their children's online activities starting in July?"
Reasoning: The mention of "July" specifies only the month and lacks the necessary detail of the year.

Public Interest: Determine if the question addresses a topic of public concern.
- 2 points: The question covers a topic that widely affects or interests the public.
- 1 point: The question is of moderate interest, relevant to specific groups.
- 0 point: The topic is overly personal or localized, lacking relevance to the broader public.

0 point examples
Example 1:
Question: Will the exhibition 'Fragile Beauty' at London's Victoria & Albert Museum include both midcentury and contemporary works in May 2024?
Reasoning: The specific details of an personal art exhibition's contents are generally of limited public interest.

Answer Not Too Obvious: This category evaluates whether the answer to a question is too predictable or straightforward based on the question itself.
- 2 points: The answer provides new or non-obvious insights, requiring additional context or understanding not explicit in the question.
- 1 point: The answer is somewhat predictable but includes minor additional information or a slight twist.
- 0 point: The answer directly restates or closely mirrors the question, offering no new details or insights.

Figure 28. Prompt in the *QA Filtering* step (part 2).