General findings in german documents:

54 correct, 7 wrong

Questions are always based on the given text.

Repeating the process multiple times for the same page strangely reproduces the same questions over and over, even on pages where there is more than enough text to generate 10 different MCQs. Or even on pages with tables, it will repeatedly ask a question about the exact same row, although the same question could be generated for every other row (see example at end of document).

Correct Answer in base MCQ is 'B' or 'C' in like 90% of MCQs. (Correct answer counts: {'B': 34, 'C': 15, 'A': 1, 'D': 3, 'E': 1})

Answer options often come in some increasing order, meaning that one can consistently reproduce the original MCQ from the re-ordered versions.

MCQ with more than 5 options, tend to include very unrealistic options.

Could not spot a pattern in where in the text the necessary information to answer the question is located. (Not always at the end of page for example)

Mistakes found:

-Inaccuracy, but question is still correctly solvable:

Question asks about order to the hospital, but the text is actually about orders from the hospital.

Page contains 2 big images and no usable text, thus a question is produced that has the solution within the question itself.

Using different wording than in the document that allows for multiple interpretations

-Multiple correct options:

Text contains an or statement and both are given as options, so both are true.

Text mentions that A is part of B, then lists both as answer options making both of them technically correct.

-Correct answer is wrong:

Text mentions usual timewindow which is taken as correct answer, but the maximum timewindow is longer.

Text requires an option to be there, but in the question it is changed to it being the only option making the correct answer false.

2nd iteration:

Added following criteria to the prompt:

Randomized Option Placement: The correct answer should appear in a randomly determined position among the options, ensuring no predictable pattern or bias.

Reasoning: Provide step-by-step reasoning for why each option is correct or incorrect, referencing specific details from the text.

Self-Consistency: Generate multiple versions of the question and reasoning, then compare results to ensure the question is clear, consistent, and accurate.

53 correct, 8 wrong

Correct answer distribution: {'B': 37, 'A': 1, 'C': 13, 'D': 2}

-Inaccuracy, but question is still correctly solvable:

Rephrased the text incorrectly in the question, changing the meaning.

Not self-contained

Using different wording than in the document that allows for multiple interpretations

text contains "at least" which is removed in the question, making the question inpresise and technically incorrect options could be correct.

Question asks where X is according to Y. According to should be replaced with in context of, as this way it doesn't really make sense.

-Multiple correct options:

Text mentions that A is part of B, then lists both as answer options making both of them technically correct.

-Correct answer is wrong:

Text mentions usual timewindow which is taken as correct answer, but the maximum timewindow is longer.

Text contains a questionnaire, LLM made a MCQ taking information from follow up questions as definitive answer to first question. But the information to answer this question is simply not available in the document, as it should be answered by the people filling out the questionnaire.

Ein Bild, das Text, Zahl, parallel, Screenshot enthält.

Automatisch generierte Beschreibung