Qestion Generation Prompt:

Act as an excellent and professional practitioner for this following tasks:

　　1. High-level and in-depth image comprehension, i.e, receives a group of images and precisely understands the details of each one.

　　2. Given the images and your profound understandings of the context, come up with reasonable and definite questions where the answers are based on single-image, cross-two-image, and cross-multi(three or more)-image contents with strictly 10 percent Not Answerable, 20 percent of the questions based on single-image, 30 percent of the questions based cross-two-image, and strictly 40 percent of the questions based on cross-multi(three or more)-image. Please propose the questions according to the following three difficulty levels:

　　　　(1) It is important to note that questions should involve three recognition levels:

　　　　Level 1: Questions that are straightforward read-offs from the context

　　　　Level 2: Questions that involve reasoning and inference

　　　　Level 3: Questions that originate from the content of the images but cannot be answered based on the available information.

　　　　(Note: at least 3 Level 3 questions.)

　　　　Here are the detailed question types that can be asked:

　　　　　→ Questions types

　　　　　　　　　　*Factual Recall

　　　　　　　　　　*Conceptual Understanding

　　　　　　　　　　*Step-by-step explanation

　　　　　　　　　　*Math or reasoning & problem solving

　　　　　　　　　　* prediction analysis

　　　　　　　　　　*Hypothetical reasoning

　　　　　　　　　　*What-IF

　　　　　　　　　　*multi-hop reasoning

　　　　　　　　　　*Data retrieval

　　　　　　　　　　*Image-based question (diagram, table, graph)

　　　　　　　　　　*Experimental design

*Argumentation

(2) Detail the evidence page number, the page number or numbers where the answer comes from.

(3) Detail the evidence source, for instance does the answer come from text, tables, graphs, charts, diagrams, illustrations, mathematical formula, and so on.

(4) It is important that 30 percent of the questions asked should have answers coming more than one evidence. List all of them in this case.

(5) Questions asked should stay faithful to the major language in the context

(6) Try your best to ask very difficult, in-depth, profound, and thought-provoking questions

(7) Strictly NO latex expressions and use OCR scans instead

(8) Strictly return the questions in the following format:

Q&A pair number: #### (Number only)

Recognition level: ####

Question: ####

Evidence page: #### (For Level 3 question, directly drop NA. For other levels, it should be a number or a list of numbers containing the page number of the answer to the question. In addition, it should be a list of numbers if the answer to the question is based on cross-two-image and cross-multi-image contents)

Evidence source: #### (For Level 3 question, directly drop NA. Otherwise, follow the guideline detailed in (4) and (5). Put them in English.)

Alignment with the guidelines: #### (How does the question proposed fit into the outlined guidelines?)

3. Filter the images given, propose questions based on images with actual contents

4. Receive the feedback on the questions that you proposed from another LLM. Keep the incorrectly answered questions and Propose more profound, difficult, and thought-provoking questions to replace the correctly answered ones if the other LLM thinks that the questions are too easy.

Answer Generation Prompt:

Act as an excellent and professional practitioner for this following tasks:

1. Precise, high-level and in-depth image comprehension

2. Receive the Q&A pairs in this following format:

    Q&A pair number: ####

    Question: ####

Extract the questions for each Q&A pair

3. Answer the questions with detailed reasoning strictly based on the content of the images provided by the user, no self-expansions or hypothetical inference allowed. If the answers to the questions cannot be found on the images, you should answer "Not Answerable".

4. If you think the question is not answerable based on the available information, write your answer as not answerable

5. Strictly return the answers in the following format using bullet points in the same order as the questions asked:

    Q&A pair number: #### (Extracted)

    Question: #### (Extracted)

    Answer: #### (Write your answer with a detailed explanation of how the answer is obtained here. If the images contain Arabic, answer should be in Arabic.)

Assessment Prompt:

Act as an excellent and professional practitioner for this following tasks:

1. High-level and in-depth image comprehension, i.e, receives a group of images and precisely understands the details of each one

2. Receives the proposed questions in this following format:

    Q&A pair number: ####

    Recognition level: ####

    Question: ####

    Evidence page: ####

    Source page: ####

    Alignment with the guidelines: ####

Rebuttal: #### (If applicable)

And the attempted answers in this following format:

Q&A pair number: ####

Question: ####

Answer: ####

3. Based on your profound understanding of the image contents and the given questions and answers, strictly do the following:

*1. Answer the proposed questions yourself based on the images. If the answers to the questions cannot be found on the images, you should answer "Not Answerable".

*2. Compute the accuracy score of the answers that you received using your own answers as the correct solutions

*3. If the accuracy score is above 50%, strictly return "The questions proposed are too easy, please propose more difficult, profound, and thought-provoking ones."

*4. If the accuracy score is lower than or equal to 50%, strictly return "The questions proposed are excellent."

*5. Format your outputs like this for each Q&A pair:

Q&A pair number: #### (Extracted from context)

Recognition level: #### (Extracted from context)

Question: #### (Extracted from context)

Attempted Answer: #### (Extracted from context)

Correct Answer: #### (Your own answer)

Evidence page: #### (Extracted from context)

Evidence source: #### (Extracted from context)

Alignment with the guidelines: #### (Extracted from context)

Feedback: #### (Write your Evaluation Result here)

Accuracy score: #### (1/1 or 0/1)

*6. Return an overall accuracy score