Appendix

Sample QA pairs in English:


```
{
        "Book name": "Applied-Fluid-Mechanics-Lab-Manual
        -1730157003",
        "Length": 114,
        "Recognition level": "1",
        "Question": "What is the formula for calculating the flow
        rate if the volumetric measuring tank was used?",
        "Correct Answer": "Q = \\frac{V}{t}",
        "Evidence page": "105",
        "Evidence source": "Text",
        "Evidence validation": true,
        "Alignment with the guidelines": "This is a
        straightforward read-off question from the text, fitting
        the Level 1 recognition level and factual recall type."
    },


{
        "Book name": "Applied-Fluid-Mechanics-Lab-Manual
        -1730157003",
        "Length": 114,
        "Recognition level": "2",
        "Question": "How would you compare the results of two
        pumps in series and two pumps in parallel at 60 rev/s?",
        "Correct Answer": "Compare the results by analyzing the
        flow rate, head, and efficiency data for both
        configurations at 60 rev/s.",
        "Evidence page": "100, 101, 104",
        "Evidence source": "Text, Diagram, Table",
        "Accuracy score": 0,
        "Evidence validation": true,
        "Alignment with the guidelines": "This question involves
        reasoning and inference, fitting the Level 2 recognition
        level and comparative & prediction analysis type."
    }
```

```
{
    "Book name": "Applied-Fluid-Mechanics-Lab-Manual-1730157003",
    "Length": 114,
    "Recognition level": "3",
    "Question": "What would be the impact on the experimental
    results if the water density used in the hydrostatic pressure
    experiment was different from the standard value?",
    "Correct Answer": "Not Answerable",
    "Evidence page": "NA",
    "Evidence source": "NA",
    "Evidence validation": true,
    "Alignment with the guidelines": "This question involves
    hypothetical reasoning and cannot be answered based on the
    available information."
}
```

```
{
        "Book name": "DCElectricalCircuitAnalysis",
        "Length": 374,
        "Recognition level": "1",
        "Question": "What is the initial voltage of the capacitor
        at t = 0 in Example 8.6?",
        "Correct Answer": "V_c(0) = 0 volts",
        "Evidence page": "276",
        "Evidence source": "Text",
        "Evidence validation": true,
        "Alignment with the guidelines": "This question is a
        straightforward read-off from the text, fitting the Level
        1 recognition level."
    }

{
        "Book name": "DCElectricalCircuitAnalysis",
        "Length": 374,
        "Recognition level": "2",
        "Question": "How does the Monte Carlo analysis in Figure
        7.23 demonstrate the effect of component tolerance on the
        circuit of Figure 7.20, and what are the key takeaways
```

```
        from the analysis?",
        "Correct Answer": "The Monte Carlo analysis in Figure 7.23
         shows the effect of component tolerance by running ten
        randomized trials for the value of V_b. The results show a
         spread of values around the nominal simulation result,
        indicating that component tolerances can cause variations
        in the circuit's behavior. The key takeaway is that even
        small tolerances can lead to noticeable differences in the
         circuit's performance.",
        "Evidence page": "231, 232, 233",
        "Evidence source": "Text, Image (Figure 7.23)",
        "Evidence validation": true,
        "Alignment with the guidelines": "This question involves
        reasoning and inference, asking for an explanation of how
        an analysis demonstrates the effect of component tolerance
         and the key takeaways."
    }

{
        "Book name": "DCElectricalCircuitAnalysis",
        "Length": 374,
        "Recognition level": "3",
        "Question": "What would be the impact on the secondary
        current (I2) in Figure 10.3 if the primary current (I1)
        was increased to 2 amps?",
        "Correct Answer": "The secondary current (I2) would
        increase proportionally because the transformer operates
        on the principle of conservation of energy, and the
        current ratio is inversely proportional to the turns ratio
        .",
        "Evidence page": "NA",
        "Evidence source": "NA",
        "Evidence validation": true,
        "Alignment with the guidelines": "This question originates
         from the content of the images but cannot be answered
        based on the available information, fitting the Level 3
        recognition level and hypothetical reasoning type."
    }
```

Sample QA pairs in Arabic:


```
{
        "Book name": "6010",
        "Length": 140,
        "Recognition level": "1",
        "Question": "ما هو عنوان الكتاب المدرسي المعروض في الصفحة
الأولى؟",
        "Correct Answer": "الرياضيات، الصف الخامس الابتدائي، الفصل
الدراسي الأول",
        "Evidence page": "1",
        "Evidence source": "Text",
        "Evidence validation": true,
        "Alignment with the guidelines": "This question is a
straightforward read-off from the context, fitting the Level 1
recognition level and single-image content guideline."

}

{
        "Book name": "6010",
        "Length": 140,
        "Recognition level": "2",
        "Question": "كيف يتم تمثيل الكسور العشرية في الكتاب؟",
        "Correct Answer": "يتم تمثيل الكسور العشرية باستخدام الشبكات
والصور التوضيحية.",
        "Evidence page": "20, 22",
        "Evidence source": "Text, Diagrams",
        "Evidence validation": true,
        "Alignment with the guidelines": "This question involves
reasoning and inference, fitting the Level 2 recognition level and
cross-two-image content guideline."

}

{
        "Book name": "6010",
        "Length": 140,
        "Recognition level": "3",
        "Question": "ما هي التحديات التي قد يواجهها الطالب عند استخدام
التقنيات الحديثة في المواقف الرياضية المختلفة؟",
        "Correct Answer": "Not Answerable",
        "Evidence page": "NA",
        "Evidence source": "NA",
        "Evidence validation": true,
        "Alignment with the guidelines": "This question originates
from the content of the images but cannot be answered based on the
available information, fitting the Level 3 recognition level and
cross-multi-image content guideline."

}
```

```
{
        "Book name": "5977",
        "Length": 108,
        "Recognition level": "1",
        "Question": "كم عدد العظام في جسم الطفل؟",
        "Correct Answer": "٣٠٠ عظمة",
        "Evidence page": "100",
        "Evidence source": "Text",
        "Evidence validation": true,
        "Alignment with the guidelines": "This is a straightforward
factual recall question based on a single image."

}

{
        "Book name": "5977",
        "Length": 108,
        "Recognition level": "2",
        "Question": "كيف يمكن أن تساعد مقارنة الأعداد الطالب في اتخاذ
قرارات حياتية؟",
        "Correct Answer": "مقارنة الأعداد تساعد الطالب في تقييم
الخيارات واتخاذ قرارات مستنيرة بناءً على القيم العددية.",
        "Evidence page": "29",
        "Evidence source": "Text",
        "Evidence validation": true,
        "Alignment with the guidelines": "This question involves
reasoning and inference based on the provided context, fitting the
Level 2 recognition level."

}

{
        "Book name": "5977",
        "Length": 108,
        "Recognition level": "3",
        "Question": "ما هي الطرق الأخرى التي يمكن استخدامها لإيجاد ناتج
جمع 7 + 9؟",
        "Correct Answer": "Not Answerable",
        "Evidence page": "NA",
        "Evidence source": "NA",
        "Evidence validation": false,
        "Alignment with the guidelines": "This question originates
from the content but cannot be answered based on the available
information, fitting the Level 3 recognition level and hypothetical
reasoning type."

}
```


Question 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 together with their OCR results and precisely understand the details of each one.

2. Given the images with their one-to-one OCR results and your profound under-

standings 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 20 percent of the questions based on single-image, strictly 30 percent of the questions based one 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:

* Factual Recall
* Conceptual Understanding
* Step-by-step explanation
* Math or reasoning & problem solving
* Comparative & prediction analysis
* Hypothetical reasoning
* What-IF
* Multi-hop reasoning
* Data retrieval
* Image-based question (diagram, table, graph)
* chat-style question & follow ups
* Experimental design
* Argumentation
* Debugging error

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

(3) Use the page number from the OCR for evidence page numbers asked in (2), which is provided as 'Content of page:***' in the OCR results.

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

(5) 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.

(6) Ask the question in Arabic if the OCR results contain Arabic

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

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

(9) 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.

OCR Prompt

You are an excellent and professional English and Arabic optical character recognition (OCR) agent.

Perform clear and sharp OCR scans for the provided image and return the scanned characters in well-formatted text with the following guidelines:

1. Aligned with the layout of the image.
2. Identify the type of graphics
3. Scan and understand the associated contents including but not limited to text and numbers
4. Place the scanned results in the correct placeholder in line with the format of the provided image layout together with the previous scanned results
5. Provide detailed description (in Arabic if the image contains Arabic) such as objects, colors, actions, purposes, etc of the content of the graphics
6. Add the well-formatted description to the same placeholder from (4)
7. The procedure should be recursive if there are sub-figures in the graphics
8. In case of math formula and equations, strictly use latex expressions.

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 together with their OCR results and precisely understand 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 and their one-to-one OCR results received. If the answers to the questions cannot be found on the images and OCR results, 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 60%, 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 60%, 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

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 here. If the images
contain Arabic, answer should be in Arabic.)
```