

Candidate Mini-Project: Physics

Please complete this assignment and submit your work within 48 hours to:

Data

You will receive a [link] to a Google Sheet containing 25 physics prompts from one of our evaluation sets.

- Each prompt includes a multiple-choice answer (A to D) listed in a column, which may be explicit options or labels referring to corresponding images.
- The correct answer is in a separate column, **but some answers have been intentionally removed**.
- Each question/answer set may include 1 to 5 images, provided in a column to complete the prompt's context.

Task #1: Gemini Response and Evaluation

Write a Python Colab notebook to process each prompt and its related images by submitting them to Gemini.

- Use Gemini 2.5 Flash via the free-tier API key from Google API Studio.
- Write the Gemini response to the empty column in the Google Sheet. The response must include the **reason** for the answer and the answer **label (A to D)**.
- Evaluate the response's correctness:
 - If no correct answer was provided, supply the correct answer yourself, based on your own knowledge and the explanation.
 - Determine whether the Gemini response was correct for all prompts
 - **CRITICAL:** Do not use an LLM or any other aid to determine the correct answer; rely only on your own knowledge. It is acceptable to be incorrect, but not to use an aid.
- Summarize Gemini's performance and include your observations and considerations in your report.

Task #2: Complexity Classification and Analysis

In the same Colab, implement a method to classify the complexity of each prompt. This can be based on an appropriate grade level, or a difficulty scale of your choice.

- If possible, include this classification in the same prompt that produces the answer.
- Write the complexity level to the Google Sheet.
- Analyze the correct response rate by level and summarize your findings in the report (e.g., how does Gemini perform across different complexity levels?).

Task #3: Error Analysis and Prompt Refinement

Select a few prompts for which Gemini did not produce a correct answer. Explore ways to improve Gemini's accuracy on these specific prompts.

- Add a hint to the prompt that helps achieve the correct answer and include this hint in the spreadsheet. *Provide a link to your chat as well*
- Compare the Gemini response with and without Google Search enabled.
- Include commentary in your report on your success in getting Gemini to produce a different answer for these specific prompts.

Task #4: Role of AI in Education

Write an essay on the following: Can AI democratize education and make it universally and economically accessible? To make this happen, what should a company like Google do, and what role can you play in that effort?

Deliverable

Write a report of 5 pages or less to summarize the output of your analysis from Tasks #1, #2, and #3, as well as your essay for Task #4, duly providing links to the colab notebook & the output sheet Tasks #1, #2 & #3

Guidelines

1. Please ensure that your complete submission does not exceed 5 pages
2. Do not resort any unfair practices for submitting this report
3. Ensure that all your work (including any LLM conversations) is linked at appropriate places within the output sheet/report
4. **Your submissions will be evaluated on the depth of your analysis, ingenuity and comprehensiveness**