

RUNNING HEAD TITLE

AI and Education: Evaluating Gemini for Physics Question Solving

NAME : Premanandh Sigamala

ROLL NO: 2201PH22

COURSE: ENGINEERING PHYSICS

INSTITUTE: INDIAN INSTITUTE OF TECHNOLOGY PATNA

PROJECT TITLE : EVALUATING AI (GEMINI 2.5 FLASH) FOR SOLVING PHYSICS-BASES MCQS

Tools Used: Python, Google Colab, Gemini 2.5 Flash API, Pandas, Excel

Dataset: 25 Physics Questions from Google Evaluation Set

Task 1: Gemini Response and Evaluation

The first part of this project was to test how well Google's Gemini 2.5 Flash model can solve basic and conceptual physics multiple choice questions. The dataset provided had 25 questions, each with four options A to D and sometimes a reference to an image. These images were mostly ray diagrams or experimental setups, but since they were only file names and not accessible links, Gemini could not actually "see" them. So, the task was done only with the text content of the questions and the answer choices. I used Google Colab to write a Python notebook for this task. The notebook imported libraries like pandas, openpyxl, and google-genai, and then connected to the Gemini 2.5 Flash API using my free-tier key. Each question and its options were read from the Excel file, combined into a single prompt, and sent to the model. Gemini's output usually contained both a short explanation and a final answer label such as Answer: B. Using regular expressions, I extracted the answer letter A to D from the text and stored it in a new column called Gemini Answer.

After that, the code compared Gemini's predicted answer with the correct one already given in the sheet. If they matched, it wrote 'Y' under the Correctness column, otherwise 'N'. For a few questions where the correct answer was blank, I filled it manually based on my own understanding as instructed in the assignment. This step ensured that every question had both Gemini's response and a ground truth answer to check accuracy.

Once all questions were processed, the updated data was saved into a new Excel file named `physics_questions_with_gemini.xlsx`. The file includes:

- the original question text and options,
- Gemini's full response and final chosen option,
- the true answer and correctness label.

While running the notebook, I printed a few sample results to check that the system worked properly. Gemini gave very detailed explanations for most questions, sometimes even outlining steps that looked logical, but the final option was not always right. In many cases, it picked the wrong letter because it could not view the diagrams or misread the physical situation described.

The model did perform reasonably well on straightforward topics like mirrors or simple motion, but struggled with conceptual or multi step reasoning problems such as electric fields Gauss's law, and rotation. Overall, the accuracy in this task was quite low only a few correct out of 25 questions but the reasoning quality was still good. It often explained why a ray would bend or *how* a force acts, even if it selected the wrong option. This shows that Gemini has potential understanding but needs multimodal input text + image for better performance in physics.

In conclusion, Task 1 helped me understand both the strengths and limits of the Gemini 2.5 Flash model. It can read and reason through textual questions but cannot replace human understanding where diagrams or visual context are required. The full working code and outputs for this task are available in the linked Colab notebook at the end of this report.

Task 2: Complexity Classification and Analysis

In this task, I tried to understand how Gemini performs on questions of different difficulty levels. The main idea was to classify each of the 25 physics questions into three categories Easy, Medium, or Hard based on the topic and the kind of thinking it required. For this, I created a small keyword-based method in Python. Words like mirror, lens, motion, and force were treated as easy since they belong to basic physics concepts. Words such as electric field, magnetic field, quantum field, or gauss field were marked as hard because they need higher level understanding and mathematical reasoning. Questions that didn't match either side were kept as medium. This simple rule-based classification helped me assign a complexity level to every question automatically.

Once the difficulty levels were assigned, I analysed how accurate Gemini was for each group. For this, I calculated the number of correct answers in each difficulty and then found the percentage. The results were interesting but expected. Gemini performed better on easy questions with around 20% accuracy, while on medium and hard ones, it dropped to about 10-15%. This shows that Gemini could handle straightforward and conceptual problems like mirror image or ray direction better than abstract or numerical ones involving electric flux or rotational motion. I saved this entire processed data into a new Excel file called **physics_questions_with_complexity.xlsx**, which contains the added column Complexity Level (Task 2) along with correctness details.

From this analysis, I observed that the model's performance clearly depends on question type and complexity. It seems to rely heavily on pattern recognition and past data rather than detailed logical reasoning. For easy questions, its training data probably helped it guess correctly, but for complex cases, it lacked step-by-step reasoning like a human would do. So, this task helped show that AI models like Gemini can answer simple physics theory questions fairly well but still struggle to generalize knowledge when the difficulty increases. The full implementation and accuracy breakdown can be found in the Colab notebook linked at the end of this report.

Task 3: Error Analysis and Prompt Refinement

In this part of the project, I focused on improving Gemini's accuracy for the questions it got wrong. After finishing Task 2, I picked five questions that Gemini had answered incorrectly. The idea was to test if adding a small hint or clue to the question could help the model think in the right direction. I wrote short hints like Hint: The question involves ray diagrams for concave lenses or Hint: Think about the law of reflection. Then I added these hints to the original question and sent them again to Gemini through the API. For each of these, I saved the new response, extracted the new predicted answer, and compared it again with the true answer to check if the accuracy improved.

While running this, I noticed that Gemini sometimes changed its explanation after reading the hint, but the new answers were not necessarily correct. In fact, in some cases, the model gave even more confident but wrong answers. This probably happened because the hints were too general and not very specific to the actual question. Since most questions were based on images and Gemini couldn't see them, the extra text confused it instead of helping. The results showed that the accuracy actually dropped from around 12% before hints to 0% after refinement. Even though the performance got worse, this experiment helped me understand that hints must be very focused and problem specific for AI models to use them properly.

At the end of this task, I saved all the new columns in a final Excel sheet named **physics_questions_final_with_hints.xlsx**. It includes the refined prompts, new responses, and updated correctness results. I also tried to test Gemini with Google Search enabled, but since the free-tier API doesn't support that, I recorded it as not available. From this task, I learned that hint-based prompting doesn't automatically improve AI performance and can sometimes confuse the model if the problem is conceptual or visual in nature. It showed that for physics questions, AI models need both text and image understanding to perform well. The complete code and experiment results for this part are available in the Colab notebook link at the end of the report.

Task 4: Role of AI in Education

I really think AI has the power to change how we learn forever. It can make education more equal and open to everyone, no matter where they live or how much money they have. For a long time, education has been limited by physical things like good schools only in cities or students who can afford coaching classes getting better chances. AI breaks that barrier because it doesn't care where you are from. All you need is a device and an internet connection. It can bring high-quality learning to the smallest towns and even to students who study on their own.

AI tools can explain difficult topics in simple ways and can adjust based on how a person learns. For example, if someone is slow in math, AI can give more basic problems and explain every step carefully. If another student is quick, AI can skip easy questions and move to complex ones. This kind of personalization was almost impossible before. I feel this makes education more human in a way because every student gets attention suited to their level, not one-size-fits-all like traditional classrooms also, AI can help break the language barrier. A lot of smart students struggle because good material is only available in English or a few big languages. But with AI translation and voice assistants, lessons can be converted into local languages. Imagine a student in a small village learning advanced physics or data science in their native language that's a huge step toward equal learning opportunities.

Still, I feel AI alone cannot solve everything. The real problem is not only knowledge but access. Many students don't have strong internet or laptops to use AI tools. Some schools don't have digital infrastructure at all. If these gaps stay, then AI might make the gap even wider because rich schools will use advanced AI systems while poor schools will still struggle. So,

democratizing education also means making sure everyone has access to devices, internet, and proper training to use technology in the right way.

Companies like Google can play a big role in this change. They already have the experience, technology, and global reach. They can make AI tools like Gemini or educational apps free for students in developing countries. They can support local governments to build digital labs in schools or provide low cost internet. Another important thing is teacher training. Many teachers still fear AI or see it as something that will replace them. But if companies like Google train teachers to use AI as an assistant, then AI can actually make teachers' work easier not replace them. For example, AI can handle grading or create personalized worksheets so teachers can focus on helping students understand.

I also think that Google and other big companies should focus more on building AI systems that are unbiased and inclusive. Many AI models are trained mostly on English or Western data, so sometimes they don't understand local content well. If AI is meant to make education equal, it has to understand different cultures, languages, and education styles. Companies should include students, teachers, and data scientists from different countries in the process of improving these tools as someone who is really interested in data and analytics, I feel excited about the role I can play. From a data enthusiast's point of view, education and AI together are like a big experiment every interaction, every question, every learning pattern can be studied to improve how we teach and learn. I would love to work on analysing educational data to see how students interact with AI tutors, which topics they struggle with most, and how the AI can adjust to make learning smoother. I think data driven improvements can make AI education tools fairer and more effective.

In the future, I want to use my skills to build or improve AI systems that make studying easier for everyone, especially for students who don't have access to expensive coaching or elite institutions. If AI can help those students, then it's doing its real job.

AI can definitely democratize education, but it has to stay focused on people, not profit. It should empower both teachers and students equally. With companies building the right tools, governments supporting infrastructure, and data enthusiasts like us working to improve them, education can truly become universal. Everyone, regardless of where they live or how much they earn, should have the same opportunity to learn. That's the kind of world I hope AI will help create someday.

Google Colab Notebook Link:

<https://colab.research.google.com/drive/16JkDIH6e6pCG25vYrVPxk4gz8ODMJyaf?usp=sharing>