May 8th, 11:00 AM - 1:00 PM

Going Down an Incline with ChatGPT

Corey R. Sissons
Portland State University

Going Down an Incline with ChatGPT

Ryan Sissons *, Justin C. Dunlap, and Ralf Widenhorn
Department of Physics, Portland State University, Portland, OR
* sissons@pdx.edu

Portland State UNIVERSITY

DEPARTMENT OF PHYSICS

INTRODUCTION

Large Language Models (LLMs) like OpenAI ChatGPT, have become popular in recent years and is easily accessible through a smartphone, tablet, or computer. We constructed a simple physics problem, with slight variations, to study how ChatGPT (LLM) answered the problem in comparison to a physics expert.

OBJECTIVES

When did the LLM treat the system in terms of rolling or sliding motion and how did it reason to include the type of motion?

How did it define and explain the conditions for the solution presented and any required assumptions made in answering the problem?

Was the final answer consistent with the rationale given, did the final answer agree with a physicist without making factually incorrect or contradictory statements?

PROCEDURE

The setup given to the LLM had the following structure:

A(n) [object] [action verb + preposition] + a(n) [incline property] inclined plane of height 1m.

The LLM was tasked with the following prompt:

Determine its speed at the bottom of the ramp.

Objects included: solid sphere, solid cylinder, hollow sphere, cube, billiard ball, AA battery, basketball, box, and book.

Action verbs included: slides down, rolls down, and is released on.

The properties of the incline plane consisted of: frictionless, no-slip, wooden, and icy.

There were 165 cases in total that the LLM had to answer. Out of the 165 different answers provided by ChatGPT, 60 cases were for the solid sphere and 20 were for the billiard ball. For the cases that included a hollow sphere, the basketball, a solid cylinder, a AA battery, and a box, 15 answers were given for each. While 10 were supplied for the book, and 5 for the cube.

RESULTS

- ChatGPT included a discussion of conservation of energy for all 165 cases. 42% of those cases DID include components needed for an expert like response.

- In 96.4% the LLM included a translational term in setting up and calculating the answer.

- For 70 (42.4%) cases, the necessary components to provide an answer that agrees with a physicist were given. Sliding motion was expected for 55 cases and 51 of them were coded that the numerical answer and the set-up agreed with an expert.

- For cases where rolling was expected, the answer would need to include both the translational and the rotational components to agree with an expert. The 75 cases that included rolling motion only had 19 (27.1%) that were coded such that the numerical answer and set-up agreed with an expert.

- 70 cases were assessed by physics experts as including rolling motion. An expert-like answer would need to include the required discussion on the moment of inertia. The LLM included a complete answer requiring a moment of inertia component in setting up and calculating the answer, the LLM described it that way in 34 cases (48.6%).

- 40 cases had conflicting verbiage and zero of them were coded where the numerical answer and set-up agreed with an expert. For this category to achieve an expert-like response, the answer needed to point of the conflicting information. The ChatGPT answers did not do that for any of them.

- ChatGPT included language that explained how or why an object was sliding or rolling for 6 cases of the total of 165. Out of those 6 cases, none met the requirements to agree with an expert.
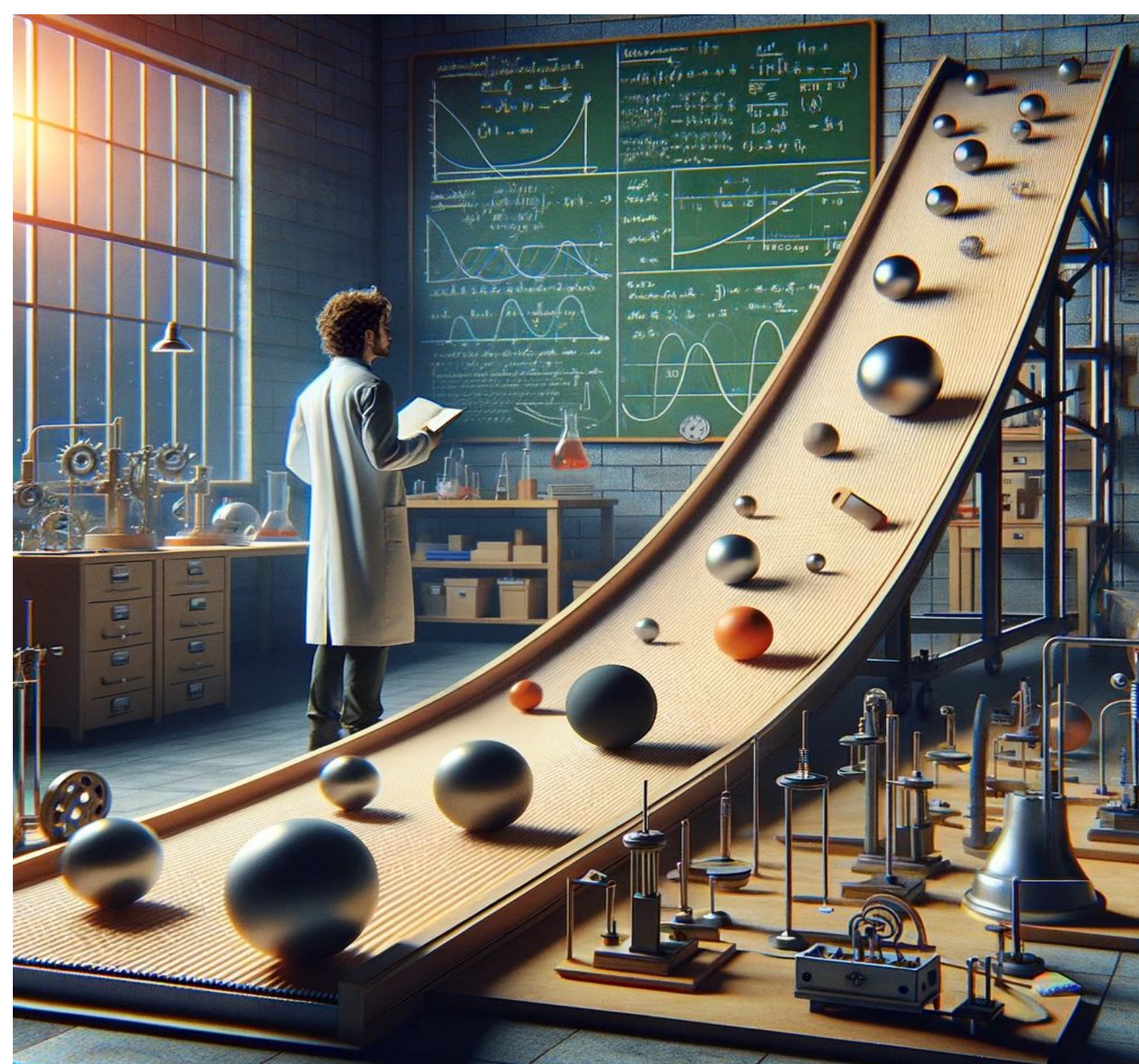
CONCLUSION

LLMs have had a large influence in education, as well as in the STEM studies. We have assessed a simple and specific physics problem with varying parameters to evaluate the proficiency of an LLM, like ChatGPT 4, in delivering expert-level responses. This approach could be used to assess the effectiveness of LLMs in addressing physics problems encountered in educational settings.

SELECT LITERATURE CITATIONS

1. OpenAI ChatGPT. https://openai.com/chatgpt, Accessed 10 July 2023.
2. Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory physics course? Physical Review Physics Education Research, 19(1), 010132.
3. Xuan-Quy, D., Ngoc-Bich, L., Xuan-Dung, P., Bac-Bien, N., & The-Duy, V. (2023). Evaluation of ChatGPT and Microsoft Bing AI Chat Performances on Physics Exams of Vietnamese National High School Graduation Examination. arXiv preprint arXiv:2306.04538.