PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH **19,** 020163 (2023)

Toward AI grading of student problem solutions in introductory physics: A feasibility study

Gerd Kortemeyer[*]

Educational Development and Technology, ETH Zurich, Zurich, Switzerland

(Received 26 April 2023; accepted 31 October 2023; published 29 November 2023)

Solving problems is crucial for learning physics, and not only final solutions but also their derivations are important. Grading these derivations is labor intensive, as it generally involves human evaluation of handwritten work. AI tools have not been an alternative, since even for short answers, they needed specific training for each problem or set of problems. Extensively pretrained AI systems offer a potentially universal grading solution without this specific training. This feasibility study explores an AI-assisted workflow to grade handwritten physics derivations using MathPix and GPT-4. We were able to successfully scan handwritten solution paths and achieved an R-squared of 0.84 compared to human graders on a synthetic dataset. The proposed workflow appears promising for formative feedback, but for final evaluations, it would best be used to assist human graders.

DOI: 10.1103/PhysRevPhysEducRes.19.020163

I. INTRODUCTION

A. Generative pretrained transformer

In Fall 2022, generative pretrained transformer (GPT) [1] rapidly gained the World stage as a publicly available AI tool with surprising capabilities. Not only would this large-language model (LLM) likely pass the Turing test for many situations [2], but the massive training effort behind GPT resulted in the emergence of capabilities that are not necessarily expected from a language model. For example, GPT-4 has been found to pass several standardized exams in the upper percentiles [3–5]. A previous version, GPT-3, could already pass an introductory physics course at a nominal level [6], and there are indications that GPT-4 performs even better on physics concepts [7].

GPT is probabilistic and essentially predicts the next so-called token in sequences, where tokens are similar to syllables in words. It does so based on probabilities "learned" from a huge text corpus consisting of materials gathered from the web, unpublished books, papers, etc. While being used, apart from continued human fine-tuning by OpenAI, the company behind GPT, the system does not learn anymore as a whole; it is pretrained. There are preliminary indications, though, that in spite of continued fine-tuning, GPT does not consistently get better over time [8]. However, GPT does learn within the confines of a particular dialogue, so it can refer to statements made earlier in the same dialogue.

GPT-4 itself is multimodal, so it can accept image input [5]. Relevant to the proposed workflow, OpenAI demonstrated how handwritten mathematical formulas are turned into machine-readable documents. Using this feature, however, is still restricted by OpenAI and thus unavailable for this feasibility study; instead, preliminary results using Mathpix [9] for the same functionality are presented.

GPT is limited by memory restrictions: at some point, it hits its internal so-called "token limit;" it can only keep a limited number of tokens in memory, comparable to the working memory limitations of a human. Inputs thus cannot become too extensive before either an error message occurs or the system simply appears to forget earlier input. which will turn out to be a limitation for the proposed workflow.

B. Mechanisms for grading

Strategically solving physics problems requires logical, conceptual, and mathematical competencies [10,11], and hardly a topic in physics education research has been investigated more extensively [12]. Feedback on problem solving is one of the most powerful tools for learning [13] but providing it frequently on a large scale poses challenges. While the final solution to physics problems, as well as predetermined, scaffolded steps along the way, can be assessed to varying degrees using computer systems [14–21], a deeper analysis of the complex problem-solving competencies of learners requires an evaluation of the solution path and derivation [22,23]. However, grading

---

[*] kgerd@ethz.ch
Also at Michigan State University, East Lansing, USA.

more than just the final answer generally requires human effort (either from faculty or from peers), as learners may take very different paths toward the same solution; there may even be errors along the way that cancel each other, such as the famous "even number of sign mistakes." As opposed to solution-focused systems, grading the journey instead of the goal requires intelligence.

Intelligence is also required to even communicate mathematical content between the learner and the graders. Short of extensive typesetting, which is inappropriate during exam situations, the best way of communicating derivations or proofs is still handwriting. Making the solutions available to machine-assisted grading requires optical character recognition (OCR). While OCR has been around and successfully used for decades [24], the recognition of handwritten mathematical formulas poses challenges due to their two-dimensional structures (superscripts, indices, fractions, matrices, etc.) [25,26].

For the longest time, the only intelligence available was human, which due to the involved cost and limited availability of expert graders resulted in learners having few assessment opportunities and time delay before feedback is returned [27]. The idea of instead using artificial intelligence (AI) has been around for a while [28] but was oftentimes hampered by the effort and cumbersomeness of training and validation. Even for short answers, AI systems (including LLMs like BERT [29]) needed to be specifically trained for each class of problems [30–33], with very few exceptions [34].

In contrast, extensively pretrained systems hold the promise of *ad hoc* being able to grade problems that they have not encountered before [34,35]. Such a turnkey solution seems within reach since these same systems successfully grade short student-generated answers from standard benchmark datasets [36] without even the need to provide reference answers [35].

Table I illustrates the benefits and drawbacks of different mechanisms for grading. These categories, of course, are rather general, and mixed forms also exist; for example, short answers can be processed without the use of artificial intelligence [37] or pretrained artificial intelligence can be used to adaptively and automatically support peer feedback among humans [38].

This feasibility study explores using pretrained LLMs (rightmost column in Table I) as standalone grading engines, formative feedback tool, or as support mechanism for human graders. In particular, it is investigated if meaningful feedback on the derivations and solution paths in free-form written homework and exam problems can be obtained, thus affordably and pragmatically providing more assessment opportunities.

II. SCENARIO

The proposed scenario has multiple steps, illustrated in Fig. 1. Learners are writing problem solutions on paper; while this might seem quaint, paper and pencil are far more efficient and intuitive than any method of typesetting (there are even arguments that this materiality is essential [66]). The scenario is particularly well suited for exams, which can then take place in traditional, supervised, "offline" modes.

1. The first step is simply scanning the assignment papers into a PDF or image format; this could be accomplished not only with most copy machines but also with smartphone apps.
2. The PDF is then processed by optical character recognition (OCR) and translated into a machine-readable format, for example, LaTeX. Scanning errors can be corrected using AI tools.
3. The LaTeX documents are then graded by an extensively pretrained AI tool.

These steps would be automated in a web-based platform: A PDF with student solutions would be uploaded, and the system would internally and automatically manage the workflow using API access to the tools.

III. METHODOLOGY

A. Machine readability of handwritten solutions

To explore the feasibility of the second step in the workflow illustrated in Fig. 1, two handwritten documents with embedded physics formulas were scanned to PDF using the smartphone app Scanner Pro [67] and transcribed into LaTeX using MathPix [9]. The output of MathPix was subsequently cleaned up using GPT-4. Had direct access to the multimodal capabilities of GPT-4 been available, the intermediate step of using MathPix could likely have been skipped.

B. Generation of sample solutions

To generate a dataset for this feasibility study, we chose to avoid the privacy and data security concerns associated with processing identifiable information such as handwriting in cloud-based systems. Instead, GPT was used to synthetically generate solutions for a simple time-dependent RC-circuit problem, which is shown as the prompt in Fig. 2. The problem involves some conceptual, strategic, and mathematical challenges but is likely in one form or the other a part of most calculus-based introductory physics courses. The initial potential difference across the capacitor is given, even though it is not needed; this superfluous information makes it tempting to immediately calculate the initial current through the capacitor, leading to a less than straightforward solution since that current later drops out. Particularly when not working symbolically, but immediately "plugging and chugging," this involves unnecessary steps [68].

Since GPT uses a probabilistic algorithm, presenting the same prompt twice will lead to different responses.

TABLE I. Comparison of assessment methods and mechanisms. This study explores the rightmost mechanism.

	Human graders	Traditional computerized systems	Traditional AI-based systems	Pretrained large language model
Benefits	• Traditional, thus generally accepted	• Highly scalable [15,27]	• Highly scalable	• Highly scalable
	• Can provide empathic feedback [39]	• Robust [40]	• Can evaluate solution paths and derivations [41]	• Can evaluate solution paths and derivations
	• Can deal with scenarios that have not been anticipated [42]	• Immediate feedback [43,44]	• Immediate, adaptive feedback [45]	• Immediate, adaptive feedback
	• Can use “common sense” [42,46]	• Can provide meaningful feedback on mastery of concepts [47,48]	• In formative scenarios, could engage in limited dialogue [49]	• In formative scenarios, could engage in dialogue
		• Can be used in adaptive testing [50,51]		• Can deal with scenarios that have not been anticipated
		• Generally less complaints about unfair grading [47]		• Does not require problem-specific training [34,35]
		• Deterministic		
Drawbacks	• Expensive [52]	• Most frequently limited to examining the final solution [53,54]	• Requires training for each problem [28]	• Reliability a possible issue
	• Does not scale well [52]	• Adaptive feedback limited to anticipated, preprogrammed solution scenarios [53]	• Reliability an issue [55]	• Feedback might be strongly rubric-specific [22,56]
	• Reliability an issue [57–59]	• Does not assess the ability to communicate concepts [42]	• May fail to deal with scenarios that have not been anticipated [60]	• May hallucinate, which students might be unable to detect [61]
	• Might be influenced by presentation [62]	• Can be perceived as inflexible [42,48]	• Probabilistic	• Probabilistic
	• Feedback might be strongly rubric-specific [22,56]	• Might be particularly vulnerable to cheating and guessing [63,64]		
	• Time-delayed feedback can have influence on learning success, depending on course [65]			
	• Subjective			

This property was used to generate a set of 25 unique sample solutions for this study. GPT-4 has better reasoning capabilities than its predecessors, so it is more likely to produce correct solutions. To also have plausible, but incorrect solutions in the sample, the majority of the solutions were generated by earlier releases.

At the time of this study (mid-April 2023), three different models of GPT were available through ChatGTP: GPT-4

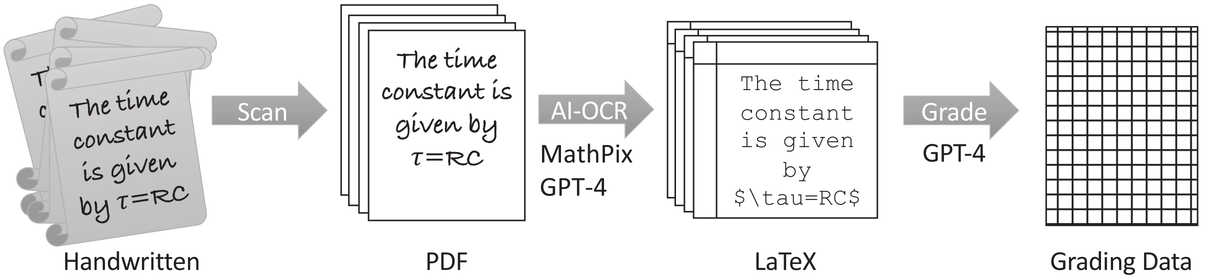


FIG. 1. Possible workflow for AI-assisted grading of handwritten work.

Prompt: At $t = 0$, a resistor (resistance R=10 Ohms) is connected across a fully charged capacitor (voltage V=5V, capacitance C=100mF). At what time t_2 does half the initial current flow through the resistor? Do not use display LaTeX, only inline LaTeX.

Solution 1: We have a resistor (R = 10 Ohms) connected across a fully charged capacitor (V = 5V, C = 100mF). To find the time t_2 at which half the initial current flows through the resistor, we need to analyze an RC circuit.
In an RC circuit, the voltage across the capacitor as a function of time is given by:
$V_C(t) = V_0 e^{-t/RC}$,
where V_0 is the initial voltage across the capacitor (5V), R is the resistance (10 Ohms), C is the capacitance (100mF), and t is the time. The current flowing through the resistor as a function of time is given by:
$I(t) = \frac{V_0}{R} e^{-t/RC}$.
Now, we need to find the time t_2 when the current is half of the initial current, so:
$\frac{1}{2} I(0) = I(t_2)$,
Substituting $I(0)$ and $I(t_2)$ with the given expressions, we get:
$\frac{1}{2} \frac{V_0}{R} = \frac{V_0}{R} e^{-t_2/RC}$.
Dividing both sides by (V_0/R), we get:
$\frac{1}{2} = e^{-t_2/RC}$.
Taking the natural logarithm of both sides, we get:
$-\ln 2 = -\frac{t_2}{RC}$.
Now, we can solve for t_2:
$t_2 = RC \ln 2$.
Substituting the given values of R and C:
$t_2 = (10, \mathrm{Ohms})(100 \times 10^{-3}, \mathrm{F}) \ln 2$.
$t_2 = 1, \mathrm{s} \ln 2$.
Therefore, the time t_2 when half the initial current flows through the resistor is approximately $0.693, \mathrm{s}$.

FIG. 2. Prompt and solution 1 for this study.

(March 23, 2023 release [69]), Default GPT-3.5, and Legacy GPT-3.5. In particular, based on the prompt in Fig. 2, solutions 1–5 were generated by GPT-4, solutions 6–13 by default GPT-3.5, and solutions 14–25 by legacy GPT-3.5; these solutions are listed in Figs. 2–7.

As an output format, LaTeX was chosen. During the output phase, ChatGPT converts display equations into a symbol font, so upon copying these expressions into a text file, the LaTeX source code would be lost; thus inline expressions were requested. Workarounds like these would not be necessary when accessing GPT directly through an API.

GPT has been successfully used for educational data augmentation in other physics-related contexts [70]; using a completely synthetic dataset is arguably less representative. Certainly, the solutions are more verbose than what students typically write [62], and they are adopting the Royal "We" of instructors and textbooks (from where the text corpus of GPT was likely gathered). However, to explore the feasibility of the workflow in Fig. 1, the most important issue will be agreement on the grading of solutions between AI and humans. To that end, the dataset spans the gamut from solutions 1 and 12, which are almost perfect, to solutions 9 and 19, which are completely missing the mark, and it includes the expected unnecessary calculations and transfer of numerical values from one formula to the next.

C. Grading of solutions

The solutions were scored by AI and humans on a scale of 0 (worst) to 4 (best) on a rubric of correctness of approach, correctness of symbolic derivations, correctness of the numerical result, and straightforwardness. While a more specific rubric, for example, addressing particular steps along the solution path, would likely have led to more consistent grading results, the rubric items were intentionally left general to explore the affordances of pretrained systems that do not need to be tuned toward particular problems.

In addition, for each solution, a one-sentence feedback was requested. The rubric scores were combined to a total score with a stronger weight on the final, numerical result. Finally, the system was prompted to generate a correlation table between the solutions based on similarity-of-approach; Fig. 8 shows an example of the associated prompt.

Since GPT is probabilistic, it will not arrive at one deterministic score combination for each solution. Similar to having more than one human grader look at the same solution, each solution was scored several times. Due to the limitations of GPT's token limit, not all 25 solutions could be fed into the system at the same time. They were thus processed in 75 dialogues of randomly compiled batches of 5 solutions, leading to an average of 15 grading rounds for each solution. Unfortunately, more often than not, ChatGPT did not provide the tables in the requested CSV format, which necessitated some subsequent manual reformatting of the responses in a text editor and Excel. The narrative one-sentence feedback messages from on average 15 "graders" were summarized using GPT-4 into one longer statement reflecting the majority, using the prompt shown in Fig. 9.

The same solutions were independently graded by the author and two of his colleagues on the same rubric. Human grading decisions are hopefully not probabilistic, but they are somewhat arbitrary, so the results of the human graders were averaged in the same way as the AI-generated grades. An alternative to this averaging approach would have been for the graders to first grade individually, but then come together, discuss possible discrepancies, and come to a consensus; however, for the purposes of this study, it was decided to use the same approach for the AI and human grades. The reader can form his or her own judgment for each one of the solutions, which can be found in Figs. 2–7.

Several measures of similarity between solutions were considered as a basis for clustering: similarity between the solution scores on the rubric for both human and AI grades, as well as the similarity-of-approach provided in the response of the AI tool; due to the token limit, only five solutions could be graded at a time, the overall 25×25-matrix was compiled by averaging the 5×5 matrices. Clustering was accomplished using the gplots package [71] within R [72].

Solution 2: We can use the RC circuit equations to find the time when the current is half the initial value. The current in the circuit can be described by the equation:
$I(t) = I_0 e^{-t/RC}$
Where $I(t)$ is the current at time t, I_0 is the initial current, and RC is the time constant of the circuit. The initial current, I_0, can be found using Ohm's law:
$I_0 = \frac{V}{R}$
We're looking for the time t_2 when the current is half the initial current:
$I(t_2) = \frac{1}{2} I_0$
Now we substitute the values for R, V, and C:
$I_0 = \frac{5V}{10\Omega} = 0.5A$
Plugging in the value of I_0 into the equation for $I(t_2)$:
$0.25A = \frac{1}{2}(0.5A)e^{-t_2/(10\Omega \cdot 100 \times 10^{-3} F)}$
We want to find t_2. First, we can simplify the equation by dividing both sides by $0.5A$:
$0.5 = e^{-t_2/(10\Omega \cdot 100 \times 10^{-3} F)}$
Now, take the natural logarithm of both sides:
$\ln(0.5) = -t_2/(10\Omega \cdot 100 \times 10^{-3} F)$
Then, multiply both sides by the time constant, RC:
$t_2 = -\ln(0.5)(10\Omega \cdot 100 \times 10^{-3} F)$
And finally, calculate the value of t_2:
$t_2 \approx 0.693 \cdot (10\Omega \cdot 100 \times 10^{-3} F) \approx 0.693 \cdot 1s \approx 0.693s$
So, at $t_2 \approx 0.693$ seconds, half the initial current flows through the resistor.

Solution 3: To solve this problem, we will first find the equation for the current in the circuit as a function of time, and then we will find the time when the current is half of its initial value.
The current flowing through the resistor at any given time t can be determined using the capacitor discharge equation:
$I(t) = I_0 e^{-t/RC}$
Where $I(t)$ is the current at time t, I_0 is the initial current, R is the resistance, C is the capacitance, and t is the time elapsed.
We need to find I_0 first, which is the initial current flowing through the resistor at $t = 0$. We can find it using Ohm's law:
$I_0 = \frac{V}{R}$
Given values are $R = 10, \text{Ohms}$ and $V = 5, \text{V}$.
$I_0 = \frac{5,\text{V}}{10,\text{Ohms}} = 0.5, \text{A}$
Now we have the equation for the current:
$I(t) = 0.5e^{-t/RC}$
We want to find the time t_2 when the current is half of its initial value:
$I(t_2) = \frac{1}{2} I_0 = 0.25, \text{A}$
So,
$0.25 = 0.5e^{-t_2/RC}$
Divide both sides by 0.5:
$0.5 = e^{-t_2/RC}$
Now take the natural logarithm of both sides:
$\ln 0.5 = -\frac{t_2}{RC}$
Given capacitance $C = 100, \text{mF} = 100 \times 10^{-3}, \text{F}$, we can now solve for t_2:
$t_2 = -RC \ln 0.5$
$t_2 = -(10, \text{Ohms})(100 \times 10^{-3}, \text{F}) \ln 0.5$
$t_2 \approx 0.693 \times 10^{-3}, \text{s}$
So, at $t_2 \approx 0.693, \text{ms}$, half the initial current flows through the resistor.

Solution 4: To find the time when half the initial current flows through the resistor, we first need to find the initial current and the equation for the current through the resistor as a function of time.
The initial current (I_0) can be found using Ohm's Law:
$I_0 = \frac{V}{R}$
where V is the initial voltage across the capacitor (5V) and R is the resistance (10 Ohms).
$I_0 = \frac{5\text{V}}{10 \text{ Ohms}} = 0.5\text{A}$
The current through the resistor as a function of time can be found using the formula for the current in an RC circuit:
$I(t) = \frac{V}{R} e^{-\frac{t}{RC}}$
where $I(t)$ is the current at time t, V is the initial voltage across the capacitor, R is the resistance, C is the capacitance, and t is the time. In our case, $V = 5\text{V}$, $R = 10$ Ohms, and $C = 100\text{mF} = 0.1\text{F}$.
Now we need to find the time t_2 when half the initial current flows through the resistor, i.e., when $I(t_2) = 0.5I_0 = 0.25\text{A}$. We can set up the equation:
$0.25\text{A} = \frac{5\text{V}}{10 \text{ Ohms}} e^{-\frac{t_2}{(10 \text{ Ohms})(0.1\text{F})}}$
Now we need to solve for t_2. First, we can simplify the equation:
$0.5 = e^{-\frac{t_2}{1\text{s}}}$
Next, take the natural logarithm of both sides:
$\ln(0.5) = -\frac{t_2}{1\text{s}}$
Finally, solve for t_2:
$t_2 = -1\text{s} \cdot \ln(0.5) \approx 0.6931\text{s}$
So, at $t_2 \approx 0.6931\text{s}$, half the initial current flows through the resistor.

Solution 5: To solve this problem, we'll first find the initial current and then determine the time it takes for the current to decrease to half that value.
The initial current (I_0) flowing through the resistor can be found using Ohm's law:
$I_0 = \frac{V}{R}$
With $V = 5V$ and $R = 10, \text{Ohms}$:
$I_0 = \frac{5}{10} = 0.5, \text{A}$
Now, let's find the time constant (τ) of the RC circuit:
$\tau = RC$
With $R = 10, \text{Ohms}$ and $C = 100, \text{mF} = 0.1, \text{F}$:
$\tau = 10 \times 0.1 = 1, \text{s}$
The current in the circuit at any time t can be expressed as:
$I(t) = I_0 e^{-t/\tau}$
We want to find the time t_2 when the current is half the initial current:
$\frac{1}{2} I_0 = I_0 e^{-t_2/\tau}$
Divide both sides by I_0:
$\frac{1}{2} = e^{-t_2/\tau}$
Now, take the natural logarithm of both sides:
$\ln\left(\frac{1}{2}\right) = -t_2/\tau$
$t_2 = -\tau \ln\left(\frac{1}{2}\right)$
Since $\tau = 1, \text{s}$:
$t_2 = -1 \ln\left(\frac{1}{2}\right)$
$t_2 \approx 0.693, \text{s}$
So, at $t_2 \approx 0.693, \text{s}$, half the initial current flows through the resistor.

FIG. 3. Solutions 2 through 5 to the prompt in Fig. 2.

Solution 6: The current in the circuit at time t is given by the equation: $i(t) = \frac{V}{R}e^{-\frac{t}{RC}}$, where RC is the time constant of the circuit.
At time t_2, half of the initial current flows through the resistor, so:
$\frac{1}{2}i(0) = \frac{V}{R}e^{-\frac{t_2}{RC}}$
$\Rightarrow \frac{1}{2}\cdot\frac{V}{R} = \frac{V}{R}e^{-\frac{t_2}{RC}}$
$\Rightarrow \frac{1}{2} = e^{-\frac{t_2}{RC}}$
Taking the natural logarithm of both sides, we get:
$\ln\frac{1}{2} = -\frac{t_2}{RC}$
$\Rightarrow t_2 = -RC\ln\frac{1}{2}$
Substituting the values given in the problem, we get:
$t_2 = -10\cdot 0.1\cdot\ln\frac{1}{2} \approx 0.69$ seconds
Therefore, half the initial current flows through the resistor at $t_2 \approx 0.69$ seconds.

Solution 7: The current through a capacitor is given by $I(t) = C\frac{dV}{dt}$, and the voltage across a capacitor is given by $V(t) = V_0e^{-\frac{t}{RC}}$ where V_0 is the initial voltage across the capacitor, R is the resistance of the circuit, and C is the capacitance of the capacitor.
At $t = 0$, the voltage across the capacitor is $V_0 = 5$ V, so we have:

$$I(0) = C\frac{dV}{dt}\Bigg|t = 0 = C\frac{d}{dt}\left[V_0e^{-\frac{t}{RC}}\right]\Bigg|t = 0 = -\frac{V_0}{R} = -0.5\text{ A}$$

where the negative sign indicates that the current is flowing in the opposite direction to the assumed direction.
To find the time t_2 at which half the initial current flows through the resistor, we need to solve the equation:
$I(t_2) = -\frac{0.5}{2} = -0.25$ A
We can use the equation for the voltage across the capacitor to solve for t_2:
$V(t_2) = V_0e^{-\frac{t_2}{RC}} = 2.5$ V
Taking the natural logarithm of both sides and rearranging, we get:
$\ln\left(\frac{V_0}{2.5}\right) = \frac{t_2}{RC}$
Substituting in the values of V_0, C, and R, we get:
$\ln\left(\frac{5}{2.5}\right) = \frac{t_2}{10\times10^{-3}}$
Simplifying, we get:
$t_2 = 10\times10^{-3}\ln(2) \approx 6.93$ ms
Therefore, half the initial current flows through the resistor at a time of $t_2 \approx 6.93$ ms.

Solution 8: The current through the circuit is given by the equation:
$i(t) = \frac{V}{R}e^{-\frac{t}{RC}}$
where $i(t)$ is the current at time t, V is the initial voltage across the capacitor, R is the resistance, C is the capacitance, and e is the base of the natural logarithm.
To find the time t_2 when half the initial current flows through the resistor, we need to solve for t in the equation:
$\frac{i(t_2)}{i(0)} = \frac{1}{2}$
Substituting the equation for $i(t)$ and simplifying, we get:
$e^{-\frac{t_2}{RC}} = \frac{1}{2}$
Taking the natural logarithm of both sides, we get:
$-\frac{t_2}{RC} = \ln\left(\frac{1}{2}\right)$
Solving for t_2, we get:
$t_2 = -RC\ln\left(\frac{1}{2}\right) = RC\ln(2) \approx 0.69$ s
Therefore, half the initial current flows through the resistor at approximately $t_2 = RC\ln(2)$.

Solution 9: The current through a capacitor in a circuit can be calculated as $i(t) = C\frac{dv(t)}{dt}$, where $v(t)$ is the voltage across the capacitor at time t.
In this case, the voltage across the capacitor is given as $V = 5V$, and the capacitance is $C = 100mF = 0.1F$. Therefore, the initial current flowing through the resistor is:
$i_0 = C\frac{dV}{dt} = 0$
This is because at time $t = 0$, the capacitor is fully charged, so there is no current flowing through the circuit.
The current at any later time t can be calculated using:
$i(t) = C\frac{dV(t)}{dt}$
The voltage across the capacitor at any time t can be found using:
$V(t) = Ve^{-\frac{t}{RC}}$
where RC is the time constant of the circuit, which is equal to the product of resistance R and capacitance C. In this case, $RC = 1s$.
Therefore, the current at time t is:
$i(t) = C\frac{dV(t)}{dt} = \frac{V}{R}e^{-\frac{t}{RC}}$
To find the time t_2 at which half the initial current flows through the resistor, we can set $i(t_2) = \frac{i_0}{2}$ and solve for t_2:
$\frac{V}{R}e^{-\frac{t_2}{RC}} = \frac{i_0}{2} = 0$
$e^{-\frac{t_2}{RC}} = 0$
This equation has no solution, which means that half the initial current never flows through the resistor.
The reason for this is that a capacitor does not allow sudden changes in voltage or current, but instead resists them by storing or releasing charge. When the circuit is first connected, the capacitor is fully charged and there is no current flowing. As the voltage across the capacitor decreases, the current flowing through the resistor increases, but it never reaches half the initial current. The current continues to decrease exponentially over time, approaching zero as the voltage across the capacitor approaches zero.

Solution 10: The current flowing through the resistor and capacitor is given by the equation:
$i(t) = \frac{V}{R}e^{-\frac{t}{RC}}$
where RC is the time constant of the circuit, which is equal to $RC = 10^{-1}$ seconds in this case.
The initial current $i(0)$ can be found by plugging in $t = 0$:
$i(0) = \frac{V}{R} = 0.5$ A
To find the time t_2 when half the initial current is flowing through the resistor, we need to solve for t in the equation:
$\frac{i(t)}{i(0)} = \frac{1}{2}$
Substituting in the expression for $i(t)$ and solving for t, we get:
$\frac{V}{R}e^{-\frac{t_2}{RC}} = 0.25$
$e^{-\frac{t_2}{RC}} = \frac{0.25R}{V}$
$-\frac{t_2}{RC} = \ln\frac{0.25R}{V}$
$t_2 = -RC\ln\frac{0.25R}{V}$
Plugging in the values for R, C, and V, we get:
$t_2 = -10^{-1}\ln\frac{0.25\times10}{5} \approx 0.138$ seconds
Therefore, half the initial current flows through the resistor at time $t_2 \approx 0.138$ seconds.

FIG. 4. Solutions 6 through 10 to the prompt in Fig. 2.

IV. RESULTS

A. Optical character recognition

The left panels of Figs. 10 and 11 show handwritten copies of the solutions in Figs. 2 and 3. There were several small scanning errors made by MathPix, such as translating "…" into "000". Using the prompt

> The following LaTeX text was OCR-scanned. Correct any errors that may have occurred during scanning, no other errors:

in GPT-4 resulted in the corrected documents shown in the right panels of Figs. 10 and 11; indicated in red is the original text, and indicated in green the corrected version.

Solution 11: The current through a capacitor is given by:
$i(t) = C\frac{dv(t)}{dt}$
where $i(t)$ is the current at time t and $v(t)$ is the voltage across the capacitor at time t.
The voltage across the capacitor decreases exponentially over time according to:
$v(t) = V_0 e^{-\frac{t}{RC}}$
where V_0 is the initial voltage across the capacitor, R is the resistance of the resistor, C is the capacitance of the capacitor, and t is the time elapsed since the capacitor was fully charged.
The current flowing through the resistor is given by Ohm's Law:
$i_R(t) = \frac{v(t)}{R}$
Substituting the equation for $v(t)$ into this equation yields:
$i_R(t) = \frac{V_0}{R} e^{-\frac{t}{RC}}$
To find the time t_2 when half of the initial current flows through the resistor, we need to solve the following equation for t:
$i_R(t_2) = \frac{1}{2} i_R(0)$
Substituting the expressions for $i_R(t_2)$ and $i_R(0)$, we get:
$\frac{V_0}{R} e^{-\frac{t_2}{RC}} = \frac{1}{2}\frac{V_0}{R}$
Simplifying this equation yields:
$e^{-\frac{t_2}{RC}} = \frac{1}{2}$
Taking the natural logarithm of both sides, we get:
$-\frac{t_2}{RC} = \ln\frac{1}{2}$
Solving for t_2, we get:
$t_2 = -RC\ln\frac{1}{2}$
Substituting the values for R and C, we get:
$t_2 = -10\ \Omega \times 100\ \mu\mathrm{F} \times \ln\frac{1}{2}$
Calculating this expression yields:
$t_2 \approx 0.69$ s
Therefore, half of the initial current flows through the resistor at approximately $t_2 = 0.69$ s.

Solution 12: The current flowing through the resistor and capacitor at time t is given by the equation:
$i(t) = \frac{V}{R} e^{-\frac{t}{RC}}$
where RC is the time constant of the circuit.
At time $t = 0$, the current is:
$i(0) = \frac{V}{R} = \frac{5}{10} = 0.5$ A
To find the time t_2 when half the initial current flows through the resistor, we need to solve the equation:
$i(t_2) = \frac{1}{2} i(0)$
Substituting the expression for $i(t)$ and simplifying, we get:
$\frac{V}{R} e^{-\frac{t_2}{RC}} = \frac{1}{2}\frac{V}{R}$
$e^{-\frac{t_2}{RC}} = \frac{1}{2}$
Taking the natural logarithm of both sides, we get:
$-\frac{t_2}{RC} = \ln\frac{1}{2} = -\ln 2$
$t_2 = RC\ln 2$
Substituting the values of R and C, we get:
$t_2 = (10\ \Omega)(100\ \mu\mathrm{F})\ln 2 = 0.693$ s
Therefore, half the initial current flows through the resistor after $t_2 = 0.693$ s.

Solution 13: The current flowing through the circuit can be described by the equation:
$I = \frac{V}{R} \cdot e^{-\frac{t}{RC}}$
where I is the current, V is the voltage, R is the resistance, C is the capacitance, and t is time.
To find the time t_2 when half the initial current flows through the resistor, we need to set I equal to half of its initial value and solve for t:
$\frac{V}{R} \cdot e^{-\frac{t_2}{RC}} = \frac{1}{2} \cdot \frac{V}{R} \cdot e^{-\frac{0}{RC}}$
Canceling out the terms, we get:
$e^{-\frac{t_2}{RC}} = \frac{1}{2}$
Taking the natural logarithm of both sides, we get:
$-\frac{t_2}{RC} = \ln\left(\frac{1}{2}\right)$
Solving for t_2, we get:
$t_2 = -RC \cdot \ln\left(\frac{1}{2}\right)$
Plugging in the values for R and C, we get:
$t_2 = -10\ \mathrm{Ohms} \cdot 100\ \mathrm{mF} \cdot \ln\left(\frac{1}{2}\right)$
$t_2 \approx 0.69$ s
Therefore, half the initial current flows through the resistor after approximately 0.69 s.

Solution 14: The voltage across the capacitor as a function of time is given by:
$V(t) = V_0 e^{-t/RC}$
Where V_0 is the initial voltage and RC is the time constant of the circuit. The time constant is defined as $RC = \frac{R}{C}$. The current through the resistor as a function of time is given by:
$I(t) = \frac{dV(t)}{dt} = -\frac{V_0}{RC} e^{-t/RC}$
At time $t = 0$, the current is at its maximum value, which is given by:
$I_0 = -\frac{V_0}{RC}$
Half the initial current is equal to $\frac{1}{2} I_0$. The time t_2 at which this occurs can be found by setting $I(t_2) = \frac{1}{2} I_0$ and solving for t_2:
$\frac{1}{2} I_0 = -\frac{V_0}{RC} e^{-t_2/RC}$
$e^{-t_2/RC} = \frac{1}{2}$
$-\frac{t_2}{RC} = \ln\frac{1}{2}$
$t_2 = -RC\ln\frac{1}{2}$
Substituting the values for R and C:
$t_2 = -10\ \mathrm{Ohms} \cdot 100\ \mathrm{mF}\ln\frac{1}{2} = -10 \cdot 10^{-3}\ln\frac{1}{2} \approx 6.93$ s
So the time t_2 at which half the initial current flows through the resistor is approximately 6.93 seconds.

FIG. 5. Solutions 11 through 14 to the prompt in Fig. 2.

The procedure fixed all scanning errors but introduced one new error: milli-Farad was erroneously changed to micro-Farad in the first solution. Overall, though, the output of this step seems appropriate for grading purposes, and the feasibility of this step has been demonstrated.

B. Grading results

Tables II–IV show the rubric-grading results by GPT-4, based on an average of 15 rounds of grading within different batches of five problems each.

Striking are the scores that have more than one point of standard deviations to their averages, which are

- The correctness of the argument in solutions 10, 15, 21, 23, 24, and 25: All of these solutions in the end arrive at the wrong numerical answer. Solutions 10 and 15 also needlessly calculate a numerical value for the initial current while solutions 21, 23, and 25 take a slight detour via the capacitative current $I(t) = C \cdot dV/dt$. Solution 24 starts with the expression for charging instead of a discharging capacitor, which would still have arrived at the same answer had it not been for numerical errors.
- The correctness of the symbolic calculations in solutions 15 and 23: Solution 15 is plug-and-chug, i.e., it plugs numbers into the first available formula and then transfers the numerical result to the next formula. Thus, the score for symbolic correctness of solution 15 should have been consistently low for all grading rounds. Solution 23 makes an error in the very

Solution 15: The voltage across the capacitor V_C and the current through the resistor I are related by the equation:

$V_C = IR$

Since the voltage across the capacitor and the capacitance are constant, the current through the resistor can be expressed as:

$I = \frac{dQ}{dt} = C\frac{dV_C}{dt} = \frac{V}{R} = \frac{5V}{10\Omega} = 0.5A$

where Q is the charge on the capacitor.

The current through the resistor decreases as the voltage across the capacitor decreases, so at time t_2, the current through the resistor will be half of its initial value:

$I(t_2) = 0.5 \cdot 0.5A = 0.25A$

The voltage across the capacitor can be expressed as:

$V_C = V_0 e^{-\frac{t}{RC}} = 5V \cdot e^{-\frac{t}{100m\Omega \cdot 10^{-3}F}}$

Setting $I(t_2) = 0.25A$ and solving for t gives:

$0.25A = 0.5A \cdot e^{-\frac{t_2}{100m\Omega \cdot 10^{-3}F}}$

$e^{-\frac{t_2}{100m\Omega \cdot 10^{-3}F}} = \frac{1}{2}$

$-\frac{t_2}{100m\Omega \cdot 10^{-3}F} = \ln\frac{1}{2}$

$t_2 = 100m\Omega \cdot 10^{-3}F \cdot \ln\frac{1}{2} = 100 \cdot 10^{-3} \cdot \ln\frac{1}{2} \approx 0.693 \cdot 100 \cdot 10^{-3} \approx 69.3 \cdot 10^{-3}s = 69.3ms$

So the time t_2 at which half the initial current flows through the resistor is approximately 69.3ms.

Solution 16: The voltage across the capacitor and the resistor can be expressed as:

$V(t) = V_0 e^{-t/(RC)}$

where V_0 is the initial voltage across the capacitor.

The current flowing through the resistor can be expressed as:

$I(t) = \frac{dV}{dt} = -\frac{V_0}{RC} e^{-t/(RC)}$

The time t_2 at which half the initial current flows through the resistor can be found by setting $I(t_2) = \frac{I_0}{2}$ and solving for t_2:

$\frac{I_0}{2} = -\frac{V_0}{RC} e^{-t_2/(RC)}$

$\frac{I_0}{2} RC = -V_0 e^{-t_2/(RC)}$

$\frac{I_0}{2V_0} = e^{-t_2/(RC)}$

$\ln\left(\frac{I_0}{2V_0}\right) = -\frac{t_2}{RC}$

$t_2 = -RC\ln\left(\frac{I_0}{2V_0}\right)$

Substituting the values $I_0 = \frac{V_0}{R} = \frac{5}{10} = 0.5$ A, $R = 10$ Ohms, and $C = 100$ mF, we get:

$t_2 = -10 \cdot 100 \cdot 10^{-3} \ln\left(\frac{0.5}{5}\right) = -10 \cdot 100 \cdot 10^{-3} \ln 0.1 \approx 2.302$ s

Solution 17: The initial current $i(0)$ through the resistor at time $t = 0$ can be calculated using Ohm's law:

$i(0) = \frac{V}{R} = \frac{5V}{10\Omega} = 0.5A$

Let t_2 be the time when half the initial current, $i(0)/2$, is flowing through the resistor.

The current through a resistor in an RC circuit is given by:

$i(t) = i(0) \cdot e^{-\frac{t}{RC}}$

Setting $i(t_2) = i(0)/2$ and solving for t_2:

$\frac{i(0)}{2} = i(0) \cdot e^{-\frac{t_2}{RC}}$

$e^{-\frac{t_2}{RC}} = \frac{1}{2}$

$-\frac{t_2}{RC} = \ln\frac{1}{2}$

$t_2 = -RC\ln\frac{1}{2}$

Substituting the values of R, C, and $i(0)$:

$t_2 = -10\Omega \cdot 100 \cdot 10^{-3}F \cdot \ln\frac{1}{2} \approx 11.43\text{ms}$

So, half the initial current is flowing through the resistor after approximately 11.43 milliseconds.

Solution 18: The initial current $I(0)$ through the resistor at time $t = 0$ can be calculated using Ohm's Law:

$I(0) = \frac{V}{R} = \frac{5}{10} = 0.5$ A

The current through the resistor at any time t can be calculated using the exponential decay equation for an RC circuit:

$I(t) = I(0) * e^{-\frac{t}{RC}}$

Where RC is the time constant of the circuit, which is equal to $RC = R * C = 10 * 10^{-5} = 10^{-4}$ seconds.

To find the time t_2 when half the initial current is flowing through the resistor, we can set $I(t_2) = \frac{I(0)}{2}$ and solve for t_2:

$\frac{I(0)}{2} = I(0) * e^{-\frac{t_2}{RC}}$

$\frac{1}{2} = e^{-\frac{t_2}{10^{-4}}}$

$\ln(\frac{1}{2}) = -\frac{t_2}{10^{-4}}$

$t_2 = -10^{-4} * \ln(\frac{1}{2})$

Using a calculator, we can find that $t_2 \approx 0.693/10^{-4} \approx 693$ ms. So, half the initial current flows through the resistor after approximately 693 milliseconds.

Solution 19: The current through the resistor at any time t can be given by the equation $I(t) = C\frac{dV}{dt}$. At time $t = 0$, the voltage across the capacitor is 5V and the current is given by $I(0) = C\frac{dV}{dt} = \frac{5}{10^{-6}} = 5 \cdot 10^6$ Amps.

The half-life of the current can be calculated as the time at which the current through the resistor is half of its initial value, or $I(t_2) = \frac{1}{2}I(0)$.

Using Ohm's law $V = IR$, the voltage across the resistor at time t_2 can be calculated as $V(t_2) = I(t_2)R = \frac{1}{2}I(0)R = \frac{1}{2} \cdot 5 \cdot 10^6 \cdot 10 = 2.5 \cdot 10^7$ Volts.

Finally, the time t_2 can be calculated using the equation for the current, $I(t) = C\frac{dV}{dt}$, and the voltage across the capacitor, $V(t) = V_0 e^{-\frac{t}{RC}}$, as follows:

$I(t_2) = C\frac{dV}{dt} = C\frac{d}{dt}(V_0 e^{-\frac{t}{RC}}) = -CV_0\frac{1}{RC}e^{-\frac{t}{RC}}$

$\frac{1}{2}I(0) = -CV_0\frac{1}{RC}e^{-\frac{t_2}{RC}}$

$t_2 = RC\ln\left(\frac{CV_0\frac{1}{R}}{\frac{1}{2}I(0)}\right) = RC\ln\left(\frac{5\cdot 10^{-6}\cdot 5}{\frac{1}{2}\cdot 5\cdot 10^6}\right) = RC\ln\left(\frac{5}{\frac{1}{2}}\right) = RC\ln 10$

$t_2 = 10 \cdot \ln 10 \approx 22.9\text{ms}$

So, the time at which half the initial current flows through the resistor is approximately 22.9ms.

Solution 20: The initial current flow through the resistor can be calculated using Ohm's law: $I(0) = \frac{V}{R} = \frac{5}{10} = 0.5$ A.

The current through the resistor at time t can be expressed as $I(t) = I(0) \cdot e^{-\frac{t}{RC}}$.

Half of the initial current would flow through the resistor when $I(t_2) = \frac{1}{2} \cdot I(0) = 0.25$ A.

Therefore, we can find t_2 by solving for t in the equation $I(t) = 0.25$ A:

$0.25 = 0.5 \cdot e^{-\frac{t_2}{RC}}$

$e^{-\frac{t_2}{RC}} = \frac{1}{2}$

$-\frac{t_2}{RC} = \ln\left(\frac{1}{2}\right)$

$t_2 = -RC \cdot \ln\left(\frac{1}{2}\right)$

$t_2 = -(10\text{ Ohms})(100\text{ mF})\ln\left(\frac{1}{2}\right) \approx 7.96$ s

So, after approximately 7.96 seconds, half of the initial current will flow through the resistor.

FIG. 6. Solutions 15 through 20 to the prompt in Fig. 2.

first line of the symbolic calculations by not eliminating the capacitance C.

- The correctness of the numerical results, in particular for solutions 2, 3, 4, 6, 7, etc.—18 out of the 25 problems altogether: 8 of these solutions actually have the correct numerical answer, 4 are wrong by one or more orders of magnitude, and 6 of the problems have a completely different numerical answers due to various reasons. The large list may be the result of GPT 4 still falling short on numerical calculations.

Solution 21: The initial current through the resistor is given by the equation $i(t) = C\frac{dv}{dt}$, where $v(t)$ is the voltage across the capacitor. When the capacitor is fully charged, the voltage across it is 5V. At time $t = 0$, the voltage across the capacitor begins to decrease as charge flows through the resistor, so $\frac{dv}{dt} < 0$.
The time t_2 at which half the initial current is flowing through the resistor can be found by solving for $i(t) = \frac{1}{2}i(0)$ and using the above equation for $i(t)$.
$i(0) = C\frac{dv}{dt} = C\frac{V}{R}$, so $i(t) = C\frac{dv}{dt} = \frac{1}{2}C\frac{V}{R} = \frac{1}{2}i(0)$.
Rearranging and solving for t gives us $t_2 = \frac{R}{2}\frac{C}{i(0)} = \frac{10\text{ Ohms}}{\frac{1}{2}\frac{5\text{ V}}{100\text{ mF}}} = 100$ ms.
So after 100 milliseconds, half the initial current will be flowing through the resistor.

Solution 22: The time constant τ of an RC circuit is given by $\tau = RC$. In this case, $\tau = RC = (10\text{ Ohms})(100\text{ mF}) = 1\text{ s}$.
The voltage across the capacitor as a function of time can be described by the equation $V_C(t) = V_0 e^{-t/\tau}$, where V_0 is the initial voltage across the capacitor.
The current through the resistor is given by $I(t) = C\frac{dV_C}{dt}$. Substituting the expression for $V_C(t)$, we get $I(t) = \frac{V_0 C}{\tau}e^{-t/\tau}$.
At time $t = 0$, the current is $I(0) = \frac{V_0 C}{\tau} = \frac{(5\text{ V})(100\text{ mF})}{1\text{ s}} = 500\text{ mA}$.
Half the initial current is $I(t) = \frac{1}{2}I(0) = 250\text{ mA}$. Setting this equal to the expression for $I(t)$ and solving for t, we get:
$\frac{1}{2}I(0) = \frac{V_0 C}{\tau}e^{-t/\tau} \implies e^{-t/\tau} = \frac{1}{2} \implies -\frac{t}{\tau} = \ln\frac{1}{2} \implies t_2 = \tau\ln 2 = 1\text{ s}\ln 2 \approx 0.693\text{ s}$
So, half the initial current flows through the resistor at approximately $t_2 = 0.693\text{ s}$

Solution 23: The current i in an RC circuit at any time t can be described by the equation $i(t) = C\frac{dV}{dt}$, where V is the voltage across the capacitor. The voltage across the capacitor can be described by the equation $V(t) = V_0 e^{-\frac{t}{RC}}$, where V_0 is the initial voltage.
Substituting the second equation into the first equation, we get:
$i(t) = C\frac{d}{dt}(V_0 e^{-\frac{t}{RC}}) = -\frac{V_0}{RC}e^{-\frac{t}{RC}}$
We want to find the time t_2 such that $i(t_2) = \frac{1}{2}i(0)$, where $i(0) = \frac{V_0}{R}$. Substituting the values for V_0, R, and C into this equation and solving for t_2 gives us:
$\frac{1}{2}i(0) = \frac{1}{2}\frac{V_0}{R} = -\frac{V_0}{2RC}e^{-\frac{t_2}{RC}}$
$e^{-\frac{t_2}{RC}} = -\frac{R}{V_0}\cdot\frac{1}{2}i(0) = -\frac{1}{10}$
$-\frac{t_2}{RC} = \ln\left(-\frac{1}{10}\right)$
$t_2 = RC\ln\left(\frac{1}{10}\right) = 10\text{s}\ln\left(\frac{1}{10}\right) \approx 16.4\text{s}$
So the time t_2 at which half the initial current flows through the resistor is approximately 16.4 seconds.

Solution 24: The voltage across the capacitor and the resistor is the same at any given time. We can use the voltage across the capacitor to determine the current flowing through the resistor.
The equation for the voltage across a capacitor is given by:
$V = V_0\left(1 - e^{-\frac{t}{RC}}\right)$
where V_0 is the initial voltage and t is the time elapsed.
The current flowing through the resistor is given by:
$I = \frac{V}{R} = \frac{V_0}{R}e^{-\frac{t}{RC}}$
We want to find the time t_2 at which the current is half of the initial current. Thus,
$\frac{V_0}{R}e^{-\frac{t_2}{RC}} = \frac{1}{2}\cdot\frac{V_0}{R}$
Solving for t_2, we get:
$t_2 = RC\ln\left(\frac{2}{e}\right) \approx 6.93RC$
Plugging in the values for R, C, and V_0, we get:
$t_2 \approx 6.93 \cdot 10 \cdot 0.0001 = 0.00693$ seconds.

Solution 25: The voltage across the capacitor as a function of time t is given by $V(t) = V_0 e^{-t/RC}$, where V_0 is the initial voltage and RC is the time constant.
The current flowing through the resistor as a function of time t is given by $I(t) = C\frac{dV}{dt} = -\frac{V_0}{RC}e^{-t/RC}$.
We want to find the time t_2 at which the current is half the initial current, $I(t_2) = \frac{1}{2}I_0$, where $I_0 = C\frac{dV}{dt}\Big|_{t=0} = \frac{V_0}{R}$.
Solving for t_2, we get:
$\frac{1}{2}I_0 = -\frac{V_0}{RC}e^{-t_2/RC}$
$\frac{1}{2}\frac{V_0}{R} = -\frac{V_0}{RC}e^{-t_2/RC}$
$e^{-t_2/RC} = -\frac{R}{2C}$
$-t_2/RC = \ln\left(-\frac{R}{2C}\right)$
$t_2 = RC\ln\left(\frac{2C}{R}\right)$
Substituting $R = 10$ Ohms and $C = 100\text{ mF} = 0.1\text{ F}$, we get:
$t_2 = 10\text{ Ohms}\cdot 0.1\text{ F}\ln\left(\frac{0.2\text{ F}}{10\text{ Ohms}}\right) \approx 13.3\text{ ms}$
So it takes approximately 13.3 ms for half the initial current to flow through the resistor.

FIG. 7. Solutions 21 through 25 to the prompt in Fig. 2.

- The straightforwardness of solutions 15, 21, 23, and 25: All of these solutions already appeared in other lists of ambiguous scores.

Overall, there does not appear to be a discernible pattern as to why these solutions are leading to widely spread scores. The one-sentence feedback given by GPT-4 also does not provide helpful hints for determining which features of these solutions the algorithm may have latched on to.

The summaries of the one-sentence feedback message sound very plausible, but many of them include incorrect or at least misleading statements. Examples of incorrect statements include:

- "The majority opinion indicates that there is an error in the calculation of the time constant, leading to an incorrect numerical result for t_2" for solution 2, when in fact the numerical result is correct.
- "The majority of graders agree that solution 24 uses the correct approach, equations, and symbolic representation," when in fact the equation for a charging capacitor is used.

Examples of misleading statements are

- "One grader notes that the final answer should be in milliseconds, not seconds" for solution 1; while it is true that microseconds would have been more elegant, this is not required.
- "Solution 17 demonstrates the correct approach," when in fact it includes unnecessary calculations.

Often this includes remarks about numerical inaccuracies, even if the result is correct. However, particularly for the solution attempts that were completely incorrect, the feedback can be useful, for example, "Solution 9 demonstrates an incorrect understanding of the initial current, mistakenly assuming it to be zero."

C. Agreement with manual grading

The scatterplot in Fig. 12 shows the correlation between the rubric scores resulting from manual and AI grading. Each symbol marks the average of the AI versus the average of the human grades for one particular rubric item

Prompt: Consider the following numbered solutions to this problem:

"At $t=0$, a resistor (resistance R=10 Ohms) is connected across a fully charged capacitor (voltage V=5V, capacitance C=100mF). At what time t_2 does half the initial current flow through the resistor?"

Grade them from 0 (worst) to 4 (best) in terms of each correctness of argumentation, symbolic correctness, numerical correctness, and straightforwardness. Provide these grades in a comma-separated table with an additional column that gives a one-sentence feedback. Also, provide a comma-separated matrix of similarity between the solutions based on approaches, with the solution number as header row.

======== Solution 1:
We have a resistor (R = 10 Ohms) connected across a fully charged capacitor (V = 5V, C = 100mF).
...
Therefore, the time t_2 when half the initial current flows through the resistor is approximately $0.693,\text{s}$.
...
======== Solution 22:
The time constant τ of an RC circuit is given by $\tau = RC$. In this case, $\tau = RC = (10\text{ Ohms})(100\text{ mF}) = 1\text{ s}$.
...

FIG. 8. Example of a grading prompt.

Prompt: The following are feedbacks on solutions to problem 1 on a test, prepared by different graders. Summarize these statements into one feedback reflecting the majority of the graders.
* Solution 1 is correct and straightforward.
* Correct and detailed solution.
* The solution is correct and straightforward, using the correct formula for the current in an RC circuit and solving for the required time.
* Correct and straightforward approach with correct calculations and explanation.
* A complete and accurate solution using the correct approach.
...

FIG. 9. Example of a prompt to summarize the grading feedback.

for one of the solutions, while the lines are linear interpolations for each rubric item.

The scores are clearly positively correlated, with a particularly high R^2 for the correctness of the argument ($R^2 \approx 0.81$) and the correctness of the numerical answer ($R^2 \approx 0.76$). This finding is surprising since one would surmise that the correctness of the argument is more subjective than, for example, the correctness of the symbolic operations; however, it turns out that symbolic correctness has the lowest R^2 ($R^2 \approx 0.54$).

The positive offsets show that the AI grade is generally higher than the human-assigned grade; also here, symbolic correctness turns out to have the largest deviation from human grading, as the AI assigns on average almost 1.5 more grade points than humans. There are few exceptions from this AI-induced grade inflation; most notably, all AI "graders" assigned zero points for the numerical correctness of solution 9, while one human grader gave credit arguing that having no numerical result is the correct result as a follow-up error to the wrong approach.

Not indicated in the plot are the standard deviations of the rubric grades. While for AI, the average of the numerical scores has the largest average standard deviations ($\bar{\sigma}_{\text{AI,num}} \approx 1.1$), for the human graders, it is symbolic correctness that shows the largest deviations ($\bar{\sigma}_{\text{human,symb}} \approx 1.1$).

Figure 13 shows the total scores on the problem, where the rubric items argument correctness, symbolic correctness, and straightforwardness are weighted 20%, and the correct final numerical answer 40%, respectively. In addition, the correspondingly combined standard deviations for the AI and human gradings are given by error bars, and the data points are labeled by solution number.

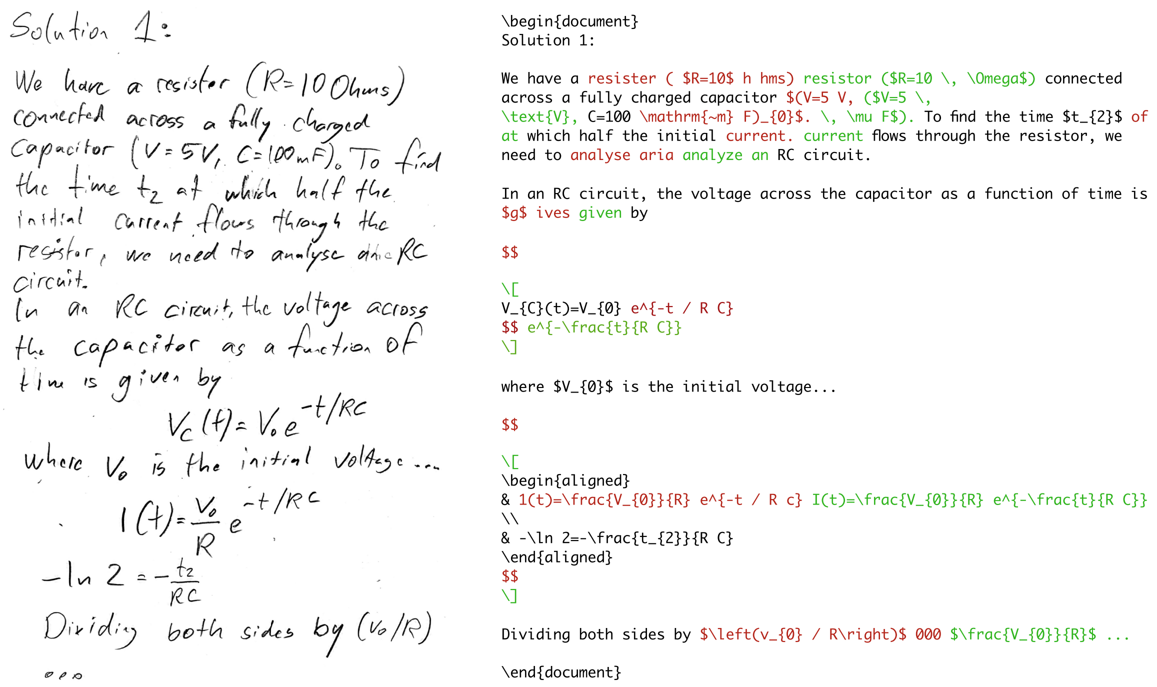
The overall agreement of $R^2 = 0.84$ is remarkably high, and the slope of the linear interpolation is close to unity. However, the AI generally scores a quarter of a grade point higher than human graders. Also, agreement between AI and human graders is better at the high end of the spectrum. The average standard deviations from the 15 AI "graders" are similar to those from the three human graders, $\bar{\sigma} \approx 0.8$. While AI and human graders overall disagreed with each other to the same degree, they did not disagree on the same solutions; for example, all AI-"graders" agreed that solution 12 is perfect ($\sigma_{\text{AI}} = 0$), but human graders applied different penalties to the unnecessary calculation of the initial current ($\sigma_{\text{human}} \approx 0.43$).

D. Clustering

Figure 14 shows dendrograms and heat maps of the similarities between the rubric grades obtained from manual grading and AI grading, as well as the AI similarity-of-approach measure, respectively. The rows and columns of this reordered similarity matrix correspond to the solution numbers indicated on the right; the dark red cells indicate high similarity, while blue fields indicate low similarity. The dendrograms on the left indicate the result of a hierarchical clustering algorithm [71], where from left to right, each branching point indicates a cut in the data. Clusters are visible as areas of high similarity along the main diagonal.

Based on these, three clusters emerge for the manual and the AI-based rubric grading. For the AI-similarity-of-approach measure (determined by GPT in response to the prompt in Fig. 8), either two clusters could be identified or the dendrogram could be cut at a deeper level (indicated by dashed lines in Fig. 14), resulting in four clusters. However, the fourth cluster only contains solution 23, so a decision to not treat it separately could be justified.

These identified clusters, indicated by the purple lines in Fig. 14, are shown in Table V; it turns out that they roughly align with the total points in Fig. 13, with the low, medium, and highly scored solutions forming the clusters.



Solution 1:

We have a resistor (R=10 Ohms) connected across a fully charged capacitor (V=5V, C=100mF). To find the time t_2 at which half the initial current flows through the resistor, we need to analyse ana RC circuit.

In an RC circuit, the voltage across the capacitor as a function of time is given by

$$V_c(t)=V_0 e^{-t/RC}$$

where V_0 is the initial voltage ...

$$I(t)=\frac{V_0}{R} e^{-t/RC}$$

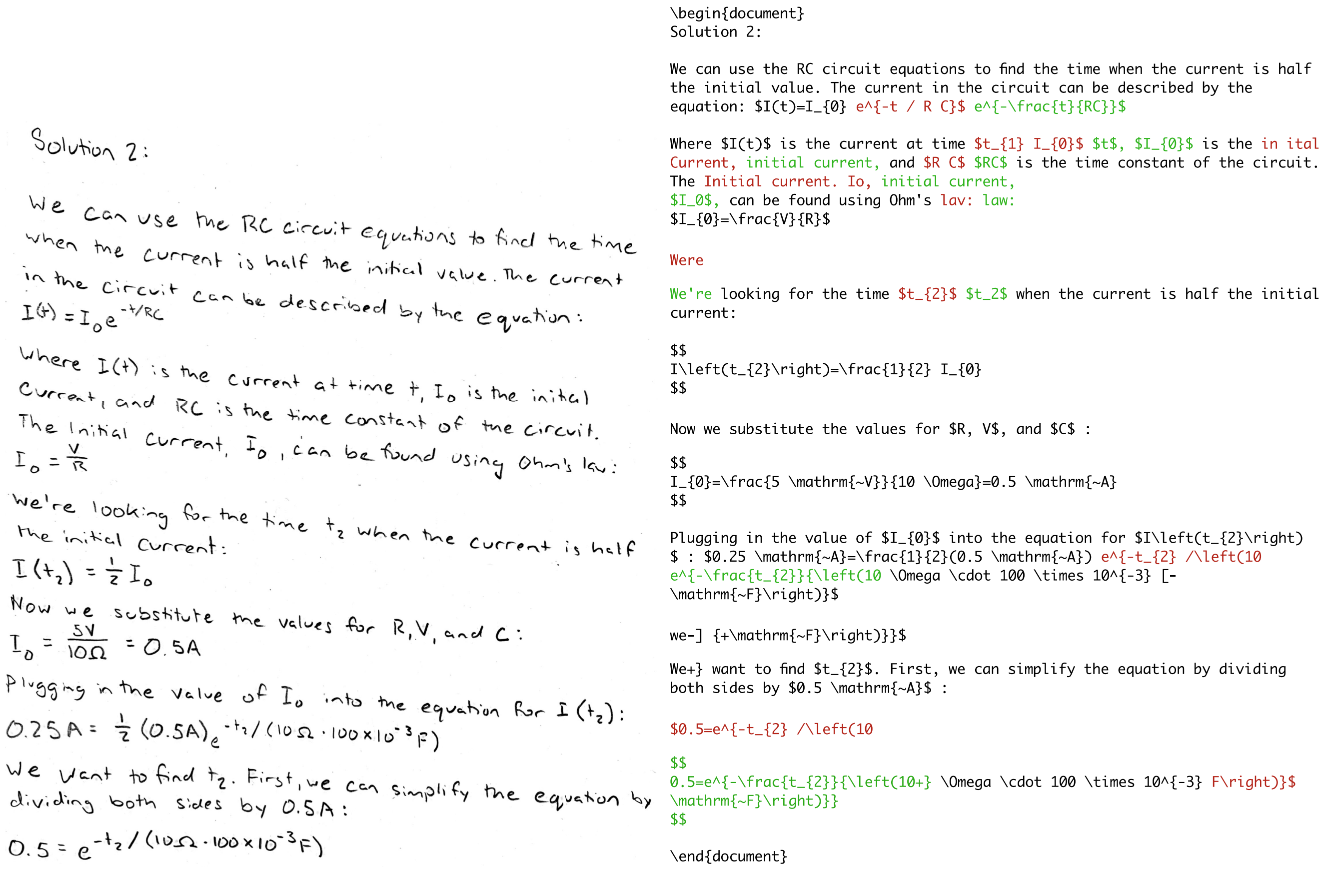
$$-\ln 2=-\frac{t_2}{RC}$$

Dividing both sides by (V_0/R)

...

```
\begin{document}
Solution 1:

We have a resister ( $R=10$ h hms) resistor ($R=10 \, \Omega$) connected
across a fully charged capacitor $(V=5 V, ($V=5 \,
\text{V}, C=100 \mathrm{~m} F)_{0}$. \, \mu F$). To find the time $t_{2}$ of
at which half the initial current. current flows through the resistor, we
need to analyse aria analyze an RC circuit.

In an RC circuit, the voltage across the capacitor as a function of time is
$g$ ives given by

$$

\[
V_{C}(t)=V_{0} e^{-t / R C}
$$ e^{-\frac{t}{R C}}
\]

where $V_{0}$ is the initial voltage...

$$

\[
\begin{aligned}
& 1(t)=\frac{V_{0}}{R} e^{-t / R c} I(t)=\frac{V_{0}}{R} e^{-\frac{t}{R C}}
\\
& -\ln 2=-\frac{t_{2}}{R C}
\end{aligned}
$$
\]

Dividing both sides by $\left(v_{0} / R\right)$ 000 $\frac{V_{0}}{R}$ ...

\end{document}
```

FIG. 10. Illustration of a possible future implementation of the first two steps in Fig. 1. A handwritten solution was scanned into PDF (left panel), translated into LaTeX using Mathpix [9], and corrected using GPT-4 (right panel).



Solution 2:

We can use the RC circuit equations to find the time when the current is half the initial value. The current in the circuit can be described by the equation:

$I(t)=I_0 e^{-t/RC}$

Where $I(t)$ is the current at time t, I_0 is the initial current, and RC is the time constant of the circuit. The Initial current, I_0, can be found using Ohm's law:

$I_0=\frac{V}{R}$

We're looking for the time t_2 when the current is half the initial current:

$I(t_2)=\frac{1}{2}I_0$

Now we substitute the values for R, V, and C:

$I_0=\frac{5V}{10\Omega}=0.5A$

Plugging in the value of I_0 into the equation for $I(t_2)$:

$0.25A=\frac{1}{2}(0.5A)e^{-t_2/(10\Omega\cdot 100\times 10^{-3}F)}$

We want to find t_2. First, we can simplify the equation by dividing both sides by 0.5A:

$0.5=e^{-t_2/(10\Omega\cdot 100\times 10^{-3}F)}$

```
\begin{document}
Solution 2:

We can use the RC circuit equations to find the time when the current is half
the initial value. The current in the circuit can be described by the
equation: $I(t)=I_{0} e^{-t / R C}$ e^{-\frac{t}{RC}}$

Where $I(t)$ is the current at time $t_{1} I_{0}$ $t$, $I_{0}$ is the in ital
Current, initial current, and $R C$ $RC$ is the time constant of the circuit.
The Initial current. Io, initial current,
$I_0$, can be found using Ohm's lav: law:
$I_{0}=\frac{V}{R}$

Were

We're looking for the time $t_{2}$ $t_2$ when the current is half the initial
current:

$$
I\left(t_{2}\right)=\frac{1}{2} I_{0}
$$

Now we substitute the values for $R, V$, and $C$ :

$$
I_{0}=\frac{5 \mathrm{~V}}{10 \Omega}=0.5 \mathrm{~A}
$$

Plugging in the value of $I_{0}$ into the equation for $I\left(t_{2}\right)
$ : $0.25 \mathrm{~A}=\frac{1}{2}(0.5 \mathrm{~A}) e^{-t_{2} /\left(10
e^{-\frac{t_{2}}{\left(10 \Omega \cdot 100 \times 10^{-3} [-
\mathrm{~F}\right)}$

we-] {+\mathrm{~F}\right)}}$

We+} want to find $t_{2}$. First, we can simplify the equation by dividing
both sides by $0.5 \mathrm{~A}$ :

$0.5=e^{-t_{2} /\left(10

$$
0.5=e^{-\frac{t_{2}}{\left(10+} \Omega \cdot 100 \times 10^{-3} F\right)}$
\mathrm{~F}\right)}}
$$

\end{document}
```

FIG. 11. Illustration of a possible future implementation of the first two steps in Fig. 1. A handwritten solution was scanned into PDF (left panel), translated into LaTeX using Mathpix [9], and corrected using GPT-4 (right panel).

TABLE II. AI-grading of solutions 1–10.

Sol.	Arg.	Symb.	Num.	Str.forw.	Feedback
1	4.0 ± 0.0	4.0 ± 0.0	3.9 ± 0.3	4.0 ± 0.0	The majority of graders agree that solution 1 is correct, straightforward, and well structured. The solution uses the correct formula for the current in an RC circuit, derives the expression for the time t_2, and solves for the required time when half the initial current flows through the resistor. One grader notes that the final answer should be in milliseconds, not seconds. Overall, the solution is clear, concise, and logically explained.
2	3.6 ± 0.9	3.7 ± 0.6	2.8 ± 1.6	3.8 ± 0.4	The majority of graders agree that the solution to problem 2 has a correct and clear approach, using the appropriate RC circuit equations and argumentation. However, there seems to be some disagreement on the numerical calculation for t_2. The majority opinion indicates that there is an error in the calculation of the time constant, leading to an incorrect numerical result for t_2. Overall, the solution is well organized and logically explained, but the numerical value should be revised.
3	3.7 ± 0.5	3.7 ± 0.5	2.2 ± 1.6	3.7 ± 0.5	The majority of graders agree that solution 3 has a correct and straightforward approach, with correct argumentation and symbolic representation. However, there is a common issue with the numerical calculation for t_2, with some graders indicating the result is off by a factor of 1000 or in the wrong unit (milliseconds instead of seconds). Despite these errors, the solution is generally well explained and follows a logical approach.
4	3.7 ± 0.5	3.8 ± 0.4	3.3 ± 1.2	3.8 ± 0.4	Solution 4 demonstrates a correct and straightforward approach to solving problem 4 using appropriate RC circuit equations and accurately deriving the time t_2. However, the majority of the graders point out a minor numerical error in the final answer, possibly due to rounding or unit conversion. Despite this, the solution is clear, well argued, and symbolically correct.
5	3.9 ± 0.4	3.9 ± 0.4	3.9 ± 0.4	3.9 ± 0.4	Solution 5 is generally correct and straightforward, successfully applying the exponential decay equation for the RC circuit to derive the result for t_2. The majority of the graders appreciate the clear argumentation, symbolic correctness, and numerical accuracy. However, there is a minor numerical error due to rounding, and some graders suggest that the clarity and straightforwardness of the argument could be improved. Overall, the solution demonstrates a good understanding of Ohm's law, the time constant, and the current equation for an RC circuit.
6	3.8 ± 0.4	3.8 ± 0.4	3.3 ± 1.1	3.8 ± 0.4	The majority of the graders agree that the solution to problem 6 is correct, well argued, and straightforward in its approach, with accurate symbolic representation and explanation. However, some graders noted an incorrect numerical value in the final result or a mistake in calculating t_2.
7	3.1 ± 0.9	3.6 ± 0.6	3.1 ± 1.1	3.1 ± 0.9	The majority of the graders found solution 7 to be correct with accurate argumentation, symbolic representation, and numerical values. However, some graders mentioned issues such as incorrect current direction, voltage calculation, and final result, as well as an incorrect manipulation of the equation. Despite these discrepancies, most graders agreed that the solution was straightforward and used the correct approach.
8	3.8 ± 0.4	3.8 ± 0.4	3.4 ± 1.1	3.8 ± 0.4	The majority of graders agree that solution 8 has correct argumentation, symbolic representation, and a straightforward explanation. The solution correctly derives the formula for the time at which half the initial current flows and uses the appropriate approach. However, there is disagreement on the numerical value, with some graders noting that it is incorrect or off by a factor due to not considering units. Overall, the solution is mostly correct, but the final numerical answer may need to be revised.
9	0.2 ± 0.4	1.0 ± 0.9	0.0 ± 0.0	0.8 ± 0.7	Solution 9 demonstrates an incorrect understanding of the initial current, mistakenly assuming it to be zero. This leads to incorrect calculations, argumentation, and the false conclusion that half the initial current never flows through the resistor. The overall approach and reasoning are flawed, resulting in an incorrect solution.
10	2.8 ± 1.1	3.1 ± 0.8	1.3 ± 1.3	2.9 ± 0.9	The majority of the graders agree that the solution to problem 10 has correct argumentation and symbolic representation, with a clear and straightforward approach. However, there are errors in the numerical calculations, including the time constant and natural logarithm term, leading to an incorrect final result. Some graders also mentioned unnecessary complexity and incorrect derivations.

TABLE III. AI-grading of solutions 11–20.

Sol.	Arg.	Symb.	Num.	Str.forw.	Feedback
11	3.5 ± 0.8	3.5 ± 0.7	2.8 ± 1.4	3.5 ± 0.8	The majority of the graders agree that solution 11 has a correct approach by using the exponential decay equation for the RC circuit and deriving the equation for $i(t)$. However, there seems to be a consensus that there are errors in the calculations, leading to an incorrect numerical result for t_2. The argumentation, symbolic correctness, and straightforwardness could be improved. Despite these errors, some graders still consider the solution to be well structured and clear.
12	4.0 ± 0.0	4.0 ± 0.0	4.0 ± 0.0	4.0 ± 0.0	Solution 12 is a correct, well-structured, and straightforward approach to the problem. The majority of graders praised the clear argumentation, appropriate use of symbols and equations, and accurate numerical results. The solution effectively uses the RC circuit formula and provides a concise explanation.
13	3.9 ± 0.3	3.9 ± 0.3	3.5 ± 1.0	3.9 ± 0.3	The majority of the graders agree that solution 13 is correct, clear, and straightforward. The solution uses the appropriate RC circuit equations, symbols, and approach to find the time when half the initial current flows through the resistor. While there is mention of a sign error and incorrect final result by a couple of graders, the overall consensus supports the solution's correctness and organization.
14	3.2 ± 0.9	3.4 ± 0.8	2.3 ± 1.4	3.2 ± 0.9	The majority of graders agree that solution 14 has a correct approach and argumentation, with symbolic correctness and straightforwardness. However, there are inconsistencies in the numerical results, likely due to errors in calculations or sign errors.
15	2.4 ± 1.4	2.4 ± 1.4	1.9 ± 1.6	2.3 ± 1.5	The majority of graders agree that the solution to problem 15 is mostly accurate in terms of approach and equations used, but there are some minor mistakes in argumentation, symbolic correctness, and numerical calculations. The solution is considered to be somewhat straightforward and well explained but with a few errors in derivation and numerical results.
16	3.3 ± 0.8	3.6 ± 0.5	1.9 ± 1.5	3.6 ± 0.5	The majority of the graders agree that solution 16 has a correct approach and demonstrates accurate symbolic representation, argumentation, and use of formulas. However, a calculation mistake in determining the initial current and subsequent errors in numerical calculations lead to an incorrect numerical answer for t_2. Despite these numerical issues, the solution is well organized and clear.
17	3.3 ± 1.0	3.8 ± 0.6	2.2 ± 1.4	3.6 ± 0.6	Solution 17 demonstrates the correct approach, argumentation, and symbolic representation but makes an error in the numerical calculation, specifically a sign error in the time constant, which leads to an incorrect final result. Despite this, the solution is considered straightforward and directly addresses the problem.
18	3.1 ± 0.8	3.4 ± 0.6	2.6 ± 1.2	3.2 ± 0.7	Solution 18 has the correct approach and argumentation, but the majority of graders point out an incorrect time constant calculation, which leads to an incorrect final numerical result. Despite this, the solution is well-structured and uses proper symbols and numerical values.
19	0.3 ± 0.6	0.9 ± 0.9	0.2 ± 0.4	0.5 ± 0.7	The majority of the graders agree that solution 19 contains multiple errors, including incorrect initial current and voltage calculations, as well as an incorrect approach for finding the time t_2 when half of the initial current flows through the resistor. These errors lead to an incorrect final result. The solution also has issues with argumentation, symbolic representation, and straightforwardness.
20	3.1 ± 0.7	3.5 ± 0.5	2.0 ± 1.3	3.3 ± 0.6	Solution 20 generally employs the correct approach and uses the exponential decay equation for solving the problem, but there are errors in the numerical calculations, specifically in the time constant and initial current. These inaccuracies lead to an incorrect final result. While some graders appreciate the clear and straightforward steps, the majority point out the numerical errors affecting the outcome.

All similarity measures identify the almost entirely wrong solutions 9, 19, and 21 as members of the "low" cluster. The manual grading has solution 23 in the middle cluster, as the only mistake is $\frac{CV_0}{-RC} = \frac{V_0}{RC}$, that is, an error in the calculation rather than a fundamental error; this solution was singled out in the clustering according to similarity-of-approach. Solution 25 not once but twice includes the same kind of error in calculating fractions. Solution 11, which arrived at the correct solution, made it into the highest cluster for manual grading but is not found in the corresponding cluster for AI grading.

Overall, clustering the solutions based on the similarity of the scores within the grading rubric provided similar

TABLE IV. AI grading of solutions 20–25.

Sol.	Arg.	Symb.	Num.	Str.forw.	Feedback
21	1.6 ± 1.1	1.9 ± 1.0	0.7 ± 0.8	1.8 ± 1.1	The majority of the graders agree that solution 21 has an incorrect approach, incorrect or incomplete argumentation, and incorrect numerical calculations. The relationship between $i(t)$ and dv/dt is also incorrect, leading to a wrong calculation for t_2. Some graders acknowledge clear explanation, symbolism, or straightforwardness in parts of the solution, but overall, the solution is deemed incorrect and weak.
22	3.7 ± 0.6	3.9 ± 0.4	3.4 ± 1.2	3.7 ± 0.6	The majority of graders agree that solution 22 is correct and straightforward in its approach to finding the time when half the initial current flows through the resistor. The solution uses the time constant and the voltage across the capacitor as a function of time. However, a few graders noted some minor issues with the numerical calculation and representation of the time constant. Overall, the solution is well explained and clear, with accurate derivation and symbolic correctness.
23	2.1 ± 1.5	2.6 ± 1.1	1.0 ± 1.6	2.1 ± 1.4	The majority of the graders find solution 23 to have a correct approach and clear step-by-step argumentation. However, there are concerns about the incorrect final numerical result and potential errors in the calculations involving the exponential term and natural logarithm. Some graders also mention an incorrect use of the time constant, incorrect equation, and wrong numerical value.
24	2.9 ± 1.2	3.5 ± 0.7	1.7 ± 1.4	3.0 ± 0.9	The majority of graders agree that solution 24 uses the correct approach, equations, and symbolic representation to find the time at which half the initial current flows through the resistor. However, there are numerical errors in the calculation, particularly regarding the time constant and the voltage across the capacitor. Additionally, some graders noted unnecessary complexity in the solution. Overall, the solution is partially correct but suffers from incorrect numerical values and minor inconsistencies.
25	1.7 ± 1.3	2.7 ± 0.8	0.7 ± 1.1	2.1 ± 1.1	The majority of the graders agree that solution 25 has an incorrect numerical result for t_2 due to inconsistencies, errors, and incorrect values in the derivation and equation. The approach is mostly correct and some steps are symbolically accurate, but there are sign errors and the method is not straightforward.

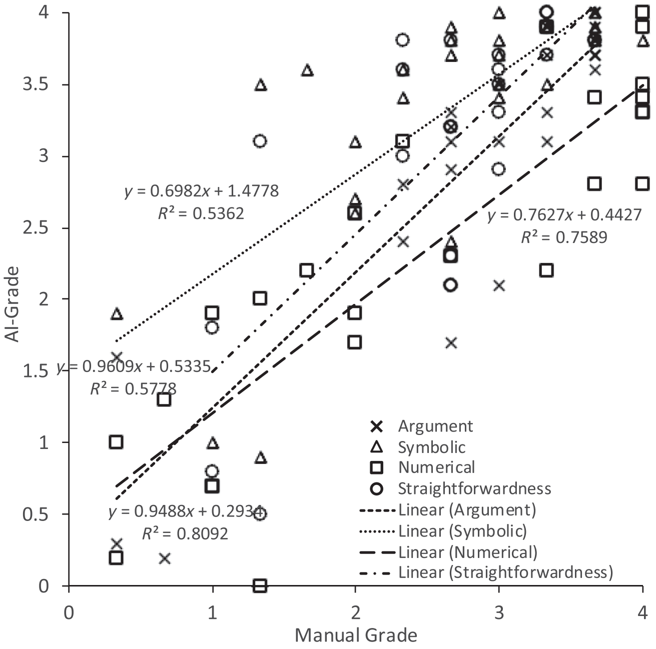


FIG. 12. Correlation between manual and AI-grading rubric scores.

information for manual and AI grading, while clustering based on the similarity of approach is less informative. The automatically formed clusters based on AI could be used as a triage mechanism in preparation for human grading.

V. DISCUSSION

A. Potential for assisting in grading

GPT and MathPix show great promise as a tool for grading handwritten solutions to physics problems, as evidenced by the scanning results in Figs. 10 and 11, the solid correlation results in Fig. 13 and the clusters in Fig. 14. The system performs best at the high end of the grading spectrum: clearly, correct solutions are generally reliably recognized, but overall, the system currently falls short of the stringency required for standalone, unsupervised exam grading. Also, the system generally assigns more points than the author and his colleagues would. This could possibly be mitigated by a more detailed grading prompt than the one in Fig. 8, however, it is unclear how to generate that prompt, and sometimes, rubrics evolve as grading proceeds, when the same unanticipated mistake occurs in multiple solutions.

While there are proposals to mitigate these shortcomings by providing "high transparency" assessments with multiple

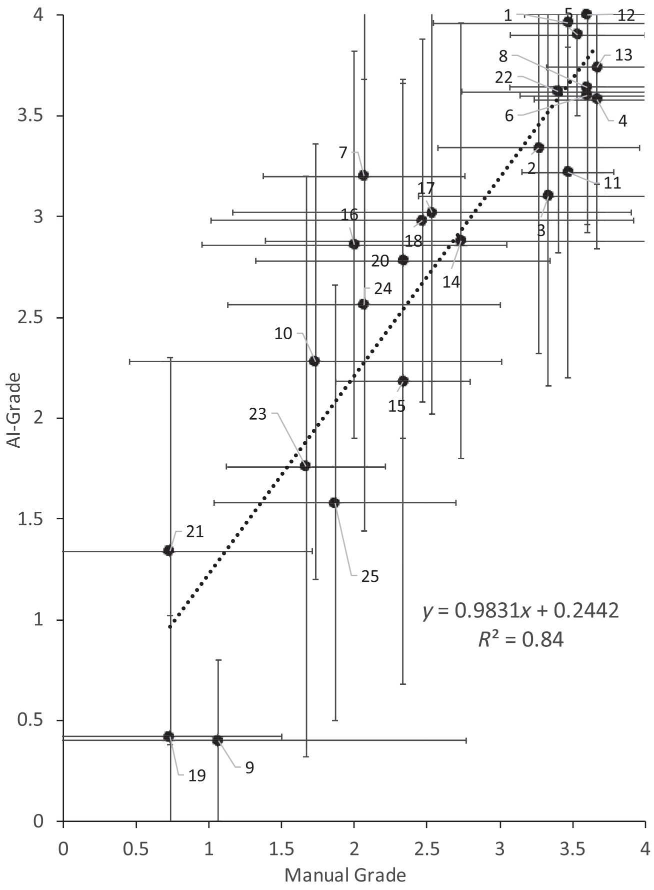


FIG. 13. Correlation between manual and AI-grading total scores. For each data point, the solution number and the standard deviations of the AI grading (vertical error bars) and human grading (horizontal error bars) are indicated.

attempts and regrade requests [55], students will likely only invoke these for false negatives, i.e., scores that are lower than what they expected. This could lead to grade inflation. Instead of waiting for student complaints, instructional staff will need to verify the scores before the exam is returned.

As typical for the tool, feedback sounds plausible, but might be misleading. In grading problems, GPT-4 still exhibits some of the same weaknesses with symbolic and numerical calculations that its predecessor GPT-3 did in solving problems [6]. Some of these shortcomings may be addressed by the Wolfram Alpha [73] plugin, which is available in ChatGPT in connection with GPT-4; at the time of writing (August 2023), preliminary experiments by the author suggest that GPT-4 is making callouts to Wolfram Alpha at appropriate times along mathematical derivations to evaluate particular expressions. Unfortunately, this plugin had not yet been available when this study was conducted in April 2023.

B. Potential for assisting in learning

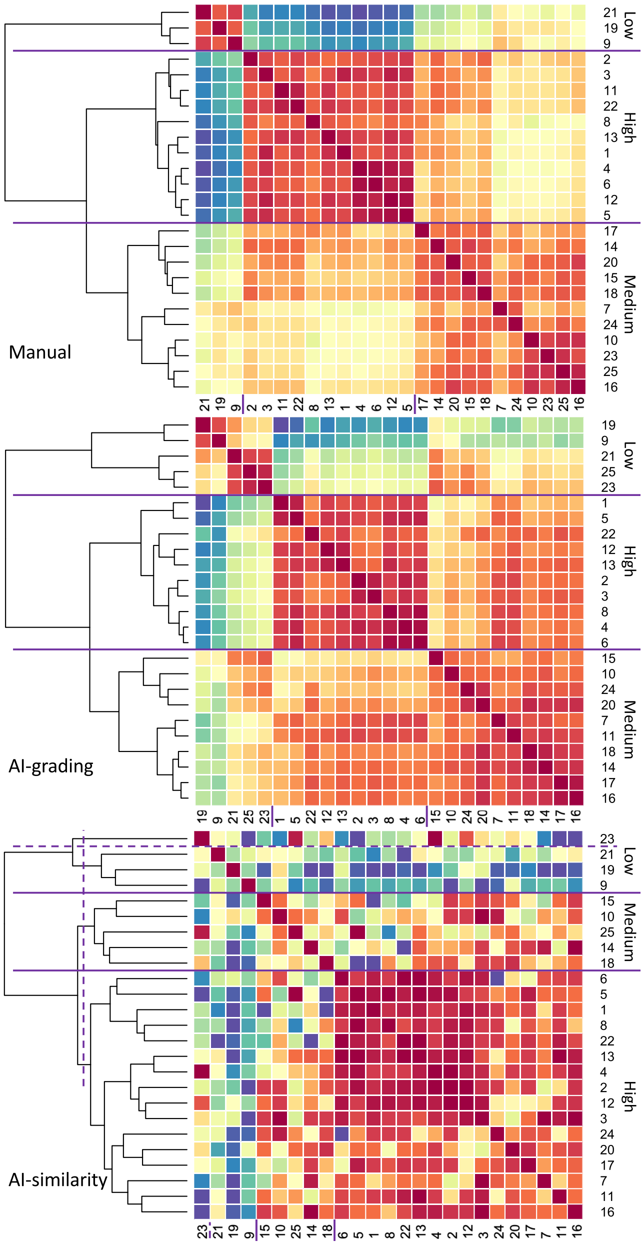


FIG. 14. Dendrograms and heat maps of similarities between the rubric grades obtained from manual grading (top panel) and AI grading (middle panel), as well as the AI similarity-of-approach measure (bottom panel). The cluster cuts are indicated by purple lines; the dashed line indicates possible cuts.

Especially in connection with smartphone-based scanning, GPT or similar tools could be used to assist students in learning. This is a far more classic domain for AI in physics education [45,74].

In that context, it is important if the system could pinpoint particular errors. At the moment, though, this is still somewhat doubtful. For example, when prompted to find the mistake in solution 14, the system responds:[1]

---

[1] The GPT output is reproduced as provided by the system, including confusing commas before physical units and unnecessary multiplication signs.

TABLE V. Clustering based on the dendrograms and heat maps in Fig. 14.

	Low	Medium	High
Manual	9 19 21	7 10 14 15 16 17 18 20 23 24 25	1 2 3 4 5 6 8 11 12 13 22
AI grading	9 19 21 23 25	7 10 11 14 15 16 17 18 20 24	1 2 3 4 5 6 8 12 13 22
AI similarity	9 19 21 (23)	10 14 15 18 25	1 2 3 4 5 6 7 8 11 12 13 16 17 20 22 24

> The mistake in the solution is in the definition of the time constant. The correct definition is $RC = R \cdot C$. The rest of the solution is correct. We need to correct the time constant and then find the correct value of t_2.
>
> The time constant is defined as $RC = R \cdot C$. In this case, $R = 10,$ Ohms and $C = 100,$ mF $= 100{\cdot}10^{-3},$ F, so the time constant is:
>
> $$RC = R \cdot C = 10, \text{Ohms} \cdot 100 \cdot 10^{-3}, \text{F} = 1, \text{s}.$$
>
> Now we can substitute the corrected time constant into the equation for t_2:
>
> $$t_2 = -RC \ln\frac{1}{2} = -1\,, \text{s} \ \ln\frac{1}{2} \approx 0.693, \text{s}.$$
>
> So the correct time t_2 at which half the initial current flows through the resistor is approximately 0.693 s.

The error is not really in the "definition" of the time constant but simply in the calculation of the powers of ten.

When prompted to pinpoint the error in solution 23, the response is

> The mistake in the solution is in the derivation of the equation $i(t_2) = \frac{1}{2} i(0)$. The correct equation should be
>
> $$i(t_2) = \frac{1}{2} i(0) = \frac{1}{2}\frac{V_0}{R}.$$
>
> However, the solution incorrectly equates this expression to the equation for $i(t)$:
>
> $$-\frac{V_0}{2RC} e^{-\frac{t_2}{RC}}.$$
>
> The correct equation for $i(t)$ should be
>
> $$i(t) = \frac{V_0}{R} e^{-\frac{t}{RC}}.$$
>
> Now, we can set up the correct equation for $i(t_2)$:
>
> $$\frac{1}{2} i(0) = \frac{1}{2}\frac{V_0}{R} = \frac{V_0}{R} e^{-\frac{t_2}{RC}}.$$
>
> Now, we solve for t_2:
>
> $$e^{-\frac{t_2}{RC}} = \frac{1}{2},$$
>
> $$-\frac{t_2}{RC} = \ln\left(\frac{1}{2}\right),$$
>
> $$t_2 = RC \ln 2 = 10 \text{ s} \quad \ln 2 \approx 6.93 \text{ s}.$$
>
> So the time t_2 at which half the initial current flows through the resistor is approximately 6.93 s.

The system correctly found that the C was not canceled but then actually proceeded to wrongly calculate the answer. The feedback given by the system can be helpful, but needs to be evaluated critically by the learner—which in itself is an increasingly valuable competency, as AI will almost invariably become ubiquitous in everyday life.

In any case, the system should not attempt to give away the solution; unfortunately, using the prompt "Act as a Socratic teacher and assist a learner with this solution attempt" made the reply more personable and supportive ("You're on the right track, but there is a small mistake …"), but did not stop the system from providing a solution, either.

VI. LIMITATIONS

This study is limited to one particular tool, GPT-4, at one particular point in time, mid-April 2023. Due to the rapid development of AI tools, it can thus provide nothing more than a snapshot of what is minimally possible.

The usage of a synthetic dataset may not fully represent the spectrum of solution approaches that students in introductory physics courses might take, and the solutions are also more verbose than typical student answers. The validity of the results is thus limited to a comparison between AI and human grades on a dataset that is not fully representative of the target scenario.

The author had to make choices when providing prompts to GPT. While some experimentation went into formulating the prompts, due to the probabilistic nature of the responses, some decisions were made based on anecdotal evidence, and better prompts framing the solutions could well have resulted in more reliable results.

GPT is only one of the rapidly evolving tools becoming broadly available. There are competing solutions by

Google [75], NVIDIA [76], Microsoft [77], and Meta [78], which may perform better or worse but could not be evaluated here.

VII. OUTLOOK

The next step to this feasibility study would be the usage of actual student work, once hurdles of human-subject research and cloud-security assessments have been cleared. These concerns need to be addressed, particularly when dealing with mandatory assignments [79]. As a welcome side effect, earlier steps in Fig. 1 could provide better workflow management even for human graders, as less paper would be shuffled around and grades could more easily be adjusted.

In addition to character and formula recognition, diagrams play an important role in problem solving [80]. A translation of sketches to vector graphics would expand the boundaries of what can be automatically preassessed; GPT-4 is capable of “making sense” out of vector graphics [81].

A possible second scenario for the first step in Fig. 1 could be using pen computing, where learners are directly drawing on a screen. Some modern tablets and laptops mimic the paper feel with the appropriate friction, and students appear to be comfortable with this technology, seeing how many of them are using this for lecture notes. A third scenario for the first step would be to photograph or scan the solutions with a smartphone and upload them to an online system. These latter two scenarios are less fit for exam settings, as being online means having access to all kinds of resources, communication channels, and online tools (including AI tools!). These additional affordances of being online would either need to be incorporated into the tasks themselves (possibly making them more demanding) or attempted to be blocked by lock-down technologies [82].

VIII. CONCLUSION

In this feasibility study, using sample handwriting and a synthetic dataset of problem solutions, GPT has shown considerable potential for grading freeform student work in physics. While AI-assigned grades have a strong correlation to manually assigned grades ($R^2 = 0.84$), they are currently not reliable enough for summative assessments, such as high-stake exams.

The system, however, is reliable enough to assist human graders by presorting or clustering solutions and by providing preliminary scores. GPT still remains hampered by its limited capabilities and inconsistencies in carrying out symbolic and numerical calculations, so currently, several independent grading rounds are needed. The narrative feedback provided by the system seems plausible but currently still frequently falls short of being reliable. The system can be helpful in formative assessment, but also in that scenario, learners need to critically evaluate its responses.

ACKNOWLEDGMENTS

The author would like to thank Christine Kortemeyer for helpful feedback. He would like to particularly thank Christian Kautz and Peter Riegler for also manually grading all solutions.

---

[1] OpenAI, ChatGPT, https://chat.openai.com/chat [accessed April 2023].
[2] A. M. Turing, Computing machinery and intelligence, Mind **LIX**, 433 (1950).
[3] T. H. Kung, M. Cheatham, A. Medinilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, and V. Tseng, Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models, PLOS Digit. Health **2**, e0000198 (2023).
[4] Samantha Murphy Kelly, ChatGPT passes exams from law and business schools, https://edition.cnn.com/2023/01/26/tech/chatgpt-passes-exams/index.html [accessed January 2023].
[5] OpenAI, ChatGPT, https://openai.com/research/gpt-4 [accessed April 2023].
[6] G. Kortemeyer, Could an artificial-intelligence agent pass an introductory physics course?, Phys. Rev. Phys. Educ. Res. **19**, 010132 (2023).
[7] C. G. West, AI and the FCI: Can ChatGPT project an understanding of introductory physics?, arXiv:2303.01067.
[8] L. Chen, M. Zaharia, and J. Zou, How is ChatGPT’s behavior changing over time?, arXiv:2307.09009.
[9] Mathpix, Mathpix OCR API for STEM, https://mathpix.com/ocr [accessed August 2023].
[10] F. Reif, J. H. Larkin, and G. C. Brackett, Teaching general learning and problem-solving skills, Am. J. Phys. **44**, 212 (1976).
[11] F. Reif, Millikan lecture 1994: Understanding and teaching important scientific thought processes, Am. J. Phys. **63**, 17 (1995).
[12] L. Hsu, E. Brewe, T. M. Foster, and K. A. Harper, Resource letter RPS-1: Research in problem solving, Am. J. Phys. **72**, 1147 (2004).
[13] J. Hattie, *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement* (Routledge, London, 2008).
[14] D. A. Kashy, G. Albertelli, G. Ashkenazi, E. Kashy, H.-K. Ng, and M. Thoennessen, Individualized interactive exercises: A promising role for network technology, in *Proceedings of Frontiers in Education* (IEEE Computer

Society, Washington, DC, 2001), Vol. 31, pp. 1073–1078, ISBN 978-0-7803-6669-5.

[15] G. Kortemeyer, E. Kashy, W. Benenson, and W. Bauer, Experiences using the open-source learning content management and assessment system LON-CAPA in introductory physics courses, Am. J. Phys. **76**, 438 (2008).

[16] J. Risley, Motivating students to learn physics using an online homework system, Newsl. APS Forum on Education Fall 2001, 3 (2001). https://higherlogicdownload.s3.amazonaws.com/APS/379bd548-10a1-4054-aec2-911470db8df9/UploadedImages/Documents/fall01.pdf.

[17] T. Stelzer and G. Gladding, The evolution of web-based activities in physics at Illinois, Newsl. APS Forum on Education Fall 2001, 7 (2001). https://higherlogicdownload.s3.amazonaws.com/APS/379bd548-10a1-4054-aec2-911470db8df9/UploadedImages/Documents/fall01.pdf.

[18] J. Mestre, D. M. Hart, K. A. Rath, and R. Dufresne, The effect of web-based homework on test performance in large enrollment introductory physics courses, J. Comput. Math. Sci. Teach. **21**, 229 (2002).

[19] C. Fredericks, Patterns of behavior in online homework for introductory physics, Ph.D. thesis, University of Massachusetts, 2007.

[20] M. Richards-Babb, J. Drelick, Z. Henry, and J. Robertson-Honecker, Online homework, help or hindrance? What students think and how they perform, J. Coll. Sci. Teach. **40**, 81 (2011).

[21] D. C. Perdian, Early identification of student performance and effort using an online homework system: A pilot study, J. Sci. Educ. Technol. **22**, 697 (2013).

[22] J. L. Docktor, J. Dornfeld, E. Frodermann, K. Heller, L. Hsu, K. A. Jackson, A. Mason, Q. X. Ryan, and J. Yang, Assessing student written problem solutions: A problem-solving rubric with application to introductory physics, Phys. Rev. Phys. Educ. Res. **12**, 010130 (2016).

[23] E. Burkholder, J. Miles, T. Layden, K. Wang, A. Fritz, and C. Wieman, Template for teaching and assessment of problem solving in introductory physics, Phys. Rev. Phys. Educ. Res. **16**, 010123 (2020).

[24] S. Mori, C. Y. Suen, and K. Yamamoto, Historical review of OCR research and development, Proc. IEEE **80**, 1029 (1992).

[25] H. Okamura, T. Kanahori, M. Suzuki, R. Fukuda, W. Cong, and F. Tamari, Handwriting interface for computer algebra systems, in *Proceedings of the Fourth Asian Technology Conference on Mathematics* (Guangzhou Normal University (GNU), Guangzhou, P.R. China, 1999), pp. 291–300, http://inftyproject.org/articles/1999_ATCM_Okamura.pdf.

[26] H. Wang, C. Pan, X. Guo, C. Ji, and K. Deng, From object detection to text detection and recognition: A brief evolution history of optical character recognition, Wiley Interdiscip. Rev. **13**, e1547 (2021).

[27] S. W. Bonham, D. L. Deardorff, and R. J. Beichner, Comparison of student performance using web and paper-based homework in college-level physics, J. Res. Sci. Teach. **40**, 1050 (2003).

[28] P. Mitros, V. Paruchuri, J. Rogosic, and D. Huang, An integrated framework for the grading of freeform responses, in *Proceedings of the Sixth Conference of MIT's Learning International Networks Consortium* (Learning International Networks Consortium, MIT, Cambridge, MA, 2013), https://linc.mit.edu/linc2013/proceedings/Session3/Session3Mit-Par.pdf.

[29] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, arXiv:1810.04805.

[30] S. Burrows, I. Gurevych, and B. Stein, The eras and trends of automatic short answer grading, Int. J. Artif. Intell. Educ. **25**, 60 (2015).

[31] M. Fowler, B. Chen, S. Azad, M. West, and C. Zilles, Autograding "explain in plain English" questions using NLP, in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (Association for Computing Machinery, New York, NY, 2021), pp. 1163–1169, https://doi.org/10.1145/3408877.3432539.

[32] R. Stanyon, E. Martello, M. Kainth, and N. K. Wilkin, Demo of Graide: AI powered assistive grading engine, in *Proceedings of the 9th ACM Conference on Learning@ Scale* (Association for Computing Machinery, New York, NY, 2022), pp. 466–468, https://doi.org/10.1145/3491140.3528263.

[33] S. Haller, A. Aldea, C. Seifert, and N. Strisciuglio, Survey on automated short answer grading with deep learning: From word embeddings to transformers, arXiv:2204.03503.

[34] C. Sung, T. Dhamecha, S. Saha, T. Ma, V. Reddy, and R. Arora, Pre-training BERT on domain resources for short answer grading, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (Association for Computational Linguistics, Hong Kong, China, 2019), pp. 6071–6075, https://dx.doi.org/10.18653/v1/D19-1628.

[35] G. Kortemeyer, Performance of the pre-trained large language model GPT-4 on automated short answer grading, arXiv:2309.09338.

[36] M. O. Dzikovska, R. Nielsen, C. Brew, C. Leacock, D. Giampiccolo, L. Bentivogli, P. Clark, I. Dagan, and H. T. Dang, Semeval-2013 task 7: The joint student response analysis and 8th recognizing textual entailment challenge, in *.Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)* (Association for Computational Linguistics, Hong Kong, China, 2013), pp. 263–274.

[37] C. M. Nakamura, S. K. Murphy, M. G. Christel, S. M. Stevens, and D. A. Zollman, Automated analysis of short responses in an interactive synthetic tutoring system for introductory physics, Phys. Rev. Phys. Educ. Res. **12**, 010122 (2016).

[38] E. Bauer, M. Greisel, I. Kuznetsov, M. Berndt, I. Kollar, M. Dresel, M. R. Fischer, and F. Fischer, Using natural language processing to support peer-feedback in the age of artificial intelligence: A cross-disciplinary framework and a research agenda, Br. J. Educ. Technol. **54**, 1222 (2023).

[39] L. Strange and F. Muzny, Ungrading with empathy: An experiment in ungrading for intermediate data science, in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2* (Association for

Computing Machinery, New York, NY, 2022), pp. 1274–1274, https://doi.org/10.1145/3545947.3573258.

[40] G. Kortemeyer, Extending item response theory to online homework, Phys. Rev. ST Phys. Educ. Res. **10**, 010118 (2014).

[41] J. Wilson, B. Pollard, J. M. Aiken, M. D. Caballero, and H. J. Lewandowski, Classification of open-ended responses to a research-based assessment using natural language processing, Phys. Rev. Phys. Educ. Res. **18**, 010141 (2022).

[42] J. Davis and T. McDonald, Online, handwritten or hybrid homework: What's best for our students in the long run?, J. Online Eng. Educ. **7**, 1 (2016), https://www.onlineengineeringeducation.com/index.php/joee/article/view/25.

[43] G. Gladding, B. Gutmann, N. Schroeder, and T. Stelzer, Clinical study of student learning using mastery style versus immediate feedback online activities, Phys. Rev. ST Phys. Educ. Res. **11**, 010114 (2015).

[44] K. Miller, K. Callaghan, L. S. McCarty, and L. Deslauriers, Increasing the effectiveness of active learning using deliberate practice: A homework transformation, Phys. Rev. Phys. Educ. Res. **17**, 010129 (2021).

[45] K. VanLehn, C. Lynch, K. Schulze, J. A. Shapiro, R. Shelby, L. Taylor, D. Treacy, A. Weinstein, and M. Wintersgill, The Andes physics tutoring system: Lessons learned, Int. J. Artif. Intell. Educ. **15**, 147 (2005).

[46] R. G. Tobin, S. J. Lacy, and S. Crissman, Does the dog in the car have kinetic energy? A multiage case study in the challenges of conceptual change, Phys. Rev. Phys. Educ. Res. **19**, 010133 (2023).

[47] M. Scott, T. Stelzer, and G. Gladding, Evaluating multiple-choice exams in large introductory physics courses, Phys. Rev. ST Phys. Educ. Res. **2**, 020102 (2006).

[48] B. Gutmann, G. Gladding, M. Lundsgaard, and T. Stelzer, Mastery-style homework exercises in introductory physics courses: Implementation matters, Phys. Rev. Phys. Educ. Res. **14**, 010128 (2018).

[49] K. Shridhar, J. Macina, M. El-Assady, T. Sinha, M. Kapur, and M. Sachan, Automatic generation of Socratic subquestions for teaching math word problems, arXiv:2211.12835.

[50] F. M. Van der Kleij, R. C. Feskens, and T. J. Eggen, Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis, Rev. Educ. Res. **85**, 475 (2015).

[51] J. W. Morphew, J. P. Mestre, H.-A. Kang, H.-H. Chang, and G. Fabry, Using computer adaptive testing to assess physics proficiency and improve exam performance in an introductory physics course, Phys. Rev. Phys. Educ. Res. **14**, 020110 (2018).

[52] D. A. Kashy, G. Albertelli, E. Kashy, and M. Thoennessen, Teaching with ALN technology: Benefits and costs*, J. Eng. Educ. **90**, 499 (2001).

[53] G. Kortemeyer and W. Bauer, Multimedia collaborative content creation (MC3): The MSU lecture online system, J. Eng. Educ. **88**, 421 (1999).

[54] A. J. Veale and T. S. Craig, Design principles for final answer assessment in linear algebra: Implications for digital testing, Teach. Math. Appl. **41**, 280 (2022).

[55] S. Azad, B. Chen, M. Fowler, M. West, and C. Zilles, Strategies for deploying unreliable ai graders in high-transparency high-stakes exams, in *Proceedings of the Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, 2020* (Springer, New York, 2020), pp. 16–28.

[56] M. M. Hull, E. Kuo, A. Gupta, and A. Elby, Problem-solving rubrics revisited: Attending to the blending of informal conceptual and formal mathematical reasoning, Phys. Rev. ST Phys. Educ. Res. **9**, 010105 (2013).

[57] H. Kruglak, An experimental study of multiple-choice and open-ended tests in physics, J. Res. Sci. Teach. **4**, 247 (1966).

[58] J. Docktor and K. Heller, Robust assessment instrument for student problem solving, in *Proceedings of the NARST 2009 Annual Meeting, Garden Grove, CA* (NARST, Minneapolis, MN, 2009), Vol. 1, pp. 1–19.

[59] H. L. Hammer and L. Habib, A quantitative analysis of the uncertainty in grading of written exams in mathematics and physics, Eurasia J. Math. Sci. Technol. Educ. **12**, 975 (2016).

[60] S. Hsu, T. W. Li, Z. Zhang, M. Fowler, C. Zilles, and K. Karahalios, Attitudes surrounding an imperfect AI autograder, in *Proceedings of the 2021 CHI conference on human factors in computing systems* (Association for Computing Machinery, New York, NY, 2021), pp. 1–15, https://doi.org/10.1145/3411764.3445424.

[61] M. N. Dahlkemper, S. Z. Lahme, and P. Klein, How do physics students evaluate artificial intelligence responses on comprehension questions? A study on the perceived scientific accuracy and linguistic quality of ChatGPT, Phys. Rev. Phys. Educ. Res. **19**, 010142 (2023).

[62] J. Stewart and S. Ballard, Effect of written presentation on performance in introductory physics, Phys. Rev. ST Phys. Educ. Res. **6**, 020120 (2010).

[63] D. J. Palazzo, Y.-J. Lee, R. Warnakulasooriya, and D. E. Pritchard, Patterns, correlates, and reduction of homework copying, Phys. Rev. ST Phys. Educ. Res. **6**, 010104 (2010).

[64] G. Kortemeyer, An empirical study of the effect of granting multiple tries for online homework, Am. J. Phys. **83**, 646 (2015).

[65] E. R. Fyfe, J. R. de Leeuw, P. F. Carvalho, R. L. Goldstone, J. Sherman, D. Admiraal, L. K. Alford, A. Bonner, C. E. Brassil, C. A. Brooks *et al.*, ManyClasses 1: Assessing the generalizable effect of immediate feedback versus delayed feedback across many college classes, Adv. Methods Pract. Psychol. Sci. **4**, 1 (2021).

[66] C. Greiffenhagen, The materiality of mathematics: Presenting mathematics at the blackboard, Br. J. Sociol. Educ. **65**, 502 (2014).

[67] Readdle Technologies Limited, Scanner Pro, https://apps.apple.com/ch/app/scanner-pro-dokumente-scannen/id333710667 [accessed September 2023].

[68] G. Kortemeyer, The losing battle against plug-and-chug, Phys. Teach. **54**, 14 (2016).

[69] OpenAI, ChatGPT Release Notes, https://help.openai.com/en/articles/6825453-chatgpt-release-notes [accessed April 2023].

[70] F. Kieser, P. Wulff, J. Kuhn, and S. Küchemann, Educational data augmentation in physics education research using ChatGPT, arXiv:2307.14475.

[71] G. R. Warnes, B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeller, M. Schwartz, and B. Venables, gplots: Various R programming tools for plotting data (2015), R package version 2.16.0, https://CRAN.R-project.org/package=gplots.

[72] R-Project, The R project for statistical computing, https://www.r-project.org (retrieved August 2019).

[73] Wolfram Research, WolframAlpha computational intelligence, https://www.wolframalpha.com [accessed March 2023].

[74] A. C. Graesser, K. VanLehn, C. P. Rosé, P. W. Jordan, and D. Harter, Intelligent tutoring systems with conversational dialogue, AI Magazine **22**, 39 (2001).

[75] Google, Google Bard, https://bard.google.com/ [accessed April 2023].

[76] NVIDIA, NVIDIA AI, https://www.nvidia.com/en-us/ai-data-science/ [accessed April 2023].

[77] Microsoft, Microsoft Copilot, https://www.microsoft.com/en-us/ai [accessed April 2023].

[78] Meta, Llama 2, https://ai.meta.com/llama/.

[79] A. Alharthi, F. Yahya, R. J. Walters, and G. B. Wills, An overview of cloud services adoption challenges in higher education institutions, in *Workshop on Emerging Software as a Service and Analytics*, Vol. 2 (SCITEPress, Setúbal, Portugal, 2015), pp. 102–109, https://dx.doi.org/10.5220/0005529701020109

[80] D. E. Meltzer, Relation between students' problem-solving performance and representational format, Am. J. Phys. **73**, 463 (2005).

[81] G. Kortemeyer, Using artificial-intelligence tools to make LaTeX content accessible to blind readers, arXiv:2306.02480.

[82] SEB Alliance, ETH Zurich, Save exam browser, https://www.safeexambrowser.org/ [accessed April 2023].