Smart Learning Environments

RESEARCH Open Access

Exploring the potential of using ChatGPT in physics education

Yicong Liang[1], Di Zou[2], Haoran Xie[3] and Fu Lee Wang[1*]

*Correspondence:
pwang@hkmu.edu.hk

[1] School of Science and Technology, Hong Kong Metropolitan University, Hong Kong, China
[2] Centre for English and Additional Languages, Lingnan University, Hong Kong, China
[3] Department of Computing and Decision Sciences, Lingnan University, Hong Kong, China

Abstract

The pretrained large language models have been widely tested for their performance on some challenging tasks including arithmetic, commonsense, and symbolic reasoning. Recently how to combine LLMs with prompting techniques has attracted lots of researchers to propose their models to automatically solve math word problems. However, most research works focus on solving math problems at the elementary school level and few works aim to solve problems in science disciplines, e.g., Physics. In this exploratory study, we discussed the potential pedagogical benefits of using ChatGPT in physics and demonstrated how to prompt ChatGPT in solving physics problems. The results suggest that ChatGPT is able to solve some physics calculation problems, explain solutions, and generate new exercises at a human level.

Keywords: ChatGPT, AI in education, Physics calculation problem, Physics learning

Introduction

In recent years, pretrained language models (PLMs) that utilize Transformer (Vaswani et al., 2017) as the fundamental architecture trained on tremendous text data have exhibited their strong capability in accomplishing various natural language processing (NLP) related tasks. Extensive research studies (Brown et al., 2020; Fedus et al., 2022; Chowdhery et al., 2023) have shown that scaling the model size, data size, and total training compute can largely improve the model performance. Hence, to discriminate language models in different parameter scales, the research community has coined the term the large language models (LLMs) for the PLMs of significant size, e.g., tens or hundreds of billions of parameters (Zhao et al., 2023). Recently, generative artificial intelligence (GAI) applications, especially the launch of ChatGPT (Chat Generative Pretrained Transformer, a powerful AI chatbot developed based on LLMs), have attracted widespread attention from both academia and industry. Meanwhile, the emergence of GAI could shift the educational objective to the highest cognitive level (i.e., creativity) (Hwang & Chen, 2023). Hwang and Chen (2023) provided some examples and guidelines for using ChatGPT in educational settings, where learners can benefit from improving their creativity, critical thinking, and problem-solving performance.

Designing an automatic solver for math word problems (MWP) is a challenging task that needs to transform the human-readable words into machine-understandable logic

Springer Open

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

representation so as to facilitate making quantitative reasoning inference. In particular, given the input text description for the math problem, the goal of the MWP solver is to map this problem into an arithmetic expression (Zhang et al., 2019). As the pretrained large language models (e.g., GPT-3) have become accessible to the public recently, researchers have been trying to explore how to improve the reasoning ability of LLMs by designing suitable prompts (Wei et al., 2022b; Wang et al., 2023; Kojima et al., 2022). Some proposed prompt techniques (Wei et al., 2022b; Wang et al., 2023) can elicit complex multi-step reasoning behavior by feeding LLMs with step-by-step reasoning examples, which helps the MWP solver achieve state-of-the-art performance in some benchmark datasets. The major advantage of these prompt-based methods is that they are free of additional training or gradient updates for the LLM, and researchers only focus on how to optimize the prompts to get better responses from the LLM in their specific downstream tasks.

Step-by-step reasoning, one of the emergent abilities of large language models (Wei et al., 2022a), can be utilized to solve mathematical word problems by using some prompting strategies, e.g., chain-of-thought (CoT) (Wei et al., 2022b). However, few research works focus on exploring how to use LLMs to solve physics calculation problems. There are some difficulties in solving physics problems based on LLMs: (1) In essence, LLMs are trained as text generators over passive plain text corpora, LLMs perform worse on the tasks that are not best expressed in the form of text, e.g., numerical computation. (2) Mathematical variables are often used in purely symbolic or algebraic contexts, while physics variables have a physical interpretation that is related to real-world phenomena. It is challenging for LLM to map problem text to corresponding physics variables. (3) Variables in math word problems only concern their quantity, but in physics problems, two vector variables have to indicate their direction inferred by LLM in vector addition. The experiment in this work shows that LLM (i.e., ChatGPT) makes errors in their final answer due to inaccurate judgment of direction. In this paper, we focus on utilizing ChatGPT in learning physics. Overall, this work aims to answer the following research questions:

RQ1 What are the pedagogical benefits of using ChatGPT for learning physics?

RQ2 How does ChatGPT perform in solving physics calculation problems?

Literature review

Language models in math word problems

Designing an automatic solver for math word problems has a long history which experiences stages from ruled-based matching to deep learning, Zhang et al. (2019) gave a comprehensive survey on math word problem solvers. The reviewed methods in Zhang et al. (2019) are mainly based on statistical learning and deep learning models where parameters are updated by training on specific datasets, so these methods failed to work in a large and diversified dataset. Recently, using few-shot prompting (Wei et al., 2022b; Wang et al., 2023; He-Yueya et al., 2023) and zero-shot prompting (Kojima et al., 2022) over large language models has emerged as a promising approach for solving MWP. Basically, these prompting methods leverage explicit intermediate reasoning steps to elicit the emergent ability (Wei et al., 2022a) in LLM for deriving the final answer. However, existing methods combined with LLM (Wei et al., 2022b; Wang et al., 2023; He-Yueya

et al., 2023) are arithmetic word problem solvers that are targeted at elementary school students. Specifically, the arithmetic expression in these problems only contains four types of fundamental operators (i.e., $\{+, -, \times, \div\}$).

Language models in STEM problems

Existing works investigated the performance of solving some math problems at higher education level (Drori et al., 2022; Frieder et al., 2023) by utilizing LLM. Frieder et al. (2023) investigated the behavior of ChatGPT on university-level math problems and found that ChatGPT understands the question description but still fails to provide a correct solution. Drori et al. (2022) demonstrate that few-show learning and program synthesis using OpenAI Codex, a neural network that is pretrained on the text and fine-tuned on code, can automatically answer 81% MIT mathematics course questions. Besides arithmetic reasoning, ChatGPT has examined its ability to perform clinical reasoning by testing its performance on the United States Medical Licensing Exam, where the results found that ChatGPT achieved near the passing threshold of 60% accuracy (Gilson et al., 2023; Kung et al., 2023).

Cognitive load theory

According to cognitive load theory (CLT), learners have a limited working memory capacity (Sweller, 1988), instructional design should aim to minimize cognitive load in terms of organizing information in a meaningful way (Sweller, 2011). As the capacity of working memory becomes effective and unlimited in the case of handling familiar material (Paas et al., 2004), learning performance is optimized under the condition that align with human cognitive architecture. It is beneficial for learners to construct instruction that can transform novel information into familiar materials. Designing instructional worked examples that demonstrate how to break down complex solutions into smaller meaningful solution elements can reduce intrinsic cognitive load (Gerjets et al., 2004). In the context of mathematical problem solving, Phan et al. (2017) explored the impact of instructional designs and found that appropriate instructional designs can serve as an optimizing agent to trigger internal personal processes.

Affordances and pedagogical benefits

In the traditional learning process, students often cannot get an instant response when they are stuck on a problem, as they need to seek help from instructors or teaching assistants. When ChatGPT is accessible to the public, it can offer instant feedback on learners' problem-solving queries, which helps learners identify mistakes and correct misconceptions in real-time. According to cognitive load theory (Paas et al., 2004), when dealing with familiar material, the limited capacity working memory will transform into the effectively unlimited capacity. In the context of solving physics problems by adapting CLT, ChatGPT can help learners avoid cognitive overload, improve their understanding of concepts, and enhance their problem-solving skills. For example, when a student is learning projectile motion, she is confused about understanding this new concept and does not know how to connect it with what she has learned before. As the response from ChatGPT is shown in Fig. 1, ChatGPT breaks down the complex motion into two simple components, which is one strategy to adapt CLT in solving problems (Paas et al., 2004).

Fig. 1 ChatGPT response to a question about how to learn projectile motion for a novice

Specifically, ChatGPT first explains the basics related to projectile motion in terms of analyzing the horizontal and vertical components separately and then combines them to illustrate how to determine the trajectory of the projectile.

ChatGPT can automatically generate an exercise related to learners' unfamiliar topics, and offer some hints instead of giving an explicit solution. Learners can solve the generated problem step by step on their own based on the given hints. For example, according to the result in Fig. 2, ChatGPT gives five hints for solving the problem of projectile motion, which breaks down the problem into smaller components step by step. In addition, the hints (Fig. 2) related to the problem are highly correlated to the basic knowledge of projectile motion (Fig. 1), suggesting that the responses from ChatGPT are consistent. ChatGPT helps reduce cognitive load for learners by presenting information in a structured format (e.g., step-by-step hints), which facilitates learners in utilizing their limited working memory in each step.

In summary, ChatGPT can offer the following pedagogical benefits in learning physics: (1) ChatGPT can offer immediate feedback and learners can seek help anytime and anywhere; (2) ChatGPT can generate accessible explanations that simplify abstract concepts, making them more understandable for learners; (3) ChatGPT can provide scaffolded learning in terms of generating step-by-step guidance and helping learners progressively build their problem-solving skills; (4) The interactive and conversational nature of ChatGPT can make the teaching-and-learning process more engaging and enjoyable, motivating learners to persist in their efforts. Kohnke et al. (2023) introduced some ways for students to improve their English by using ChatGPT. Similarly, this work offers some suggestions on how students can use ChatGPT to improve their physics study in "Appendix A".

Physics calculation problem

A math word problem can be incrementally formalized as a set of variables and equations by LLM (He-Yueya et al., 2023) and variables in MWP only take their value into consideration in the calculation process. However, physics variables are associated with specific units of measurement and unit conversion has to be done in the calculation of a physics problem. In addition, objects have to be recognized in some particular physics problems (e.g., pulley problem[1]) since analyzing the interaction of different objects helps to solve the problem.

Remove any irrelevant information or distractions that may interfere with learning, which motivates in finding a representation for a physics problem so as to minimize extraneous cognitive load (Sweller, 2011). This includes simplifying the presentation of information and focusing on the essential elements of a problem. In this paper a physics calculation problem p is defined as $< \mathcal{O}, \mathcal{V} >$, where $\mathcal{O}$ denotes a set of objects to be analyzed, and $\mathcal{V}$ denotes a set of variables contained in this problem. Furthermore, variables can be further classified into two types, namely object-related variables and environmental variables, where object-related variables should correspond to one specific object in $\mathcal{O}$. For each variable $v \in \mathcal{V}$, it consists of threefold elements $< \gamma, \phi, \lambda >$, where

[1] Pulley problem is a type of physics problem that involves two or more masses connected by a pulley system.

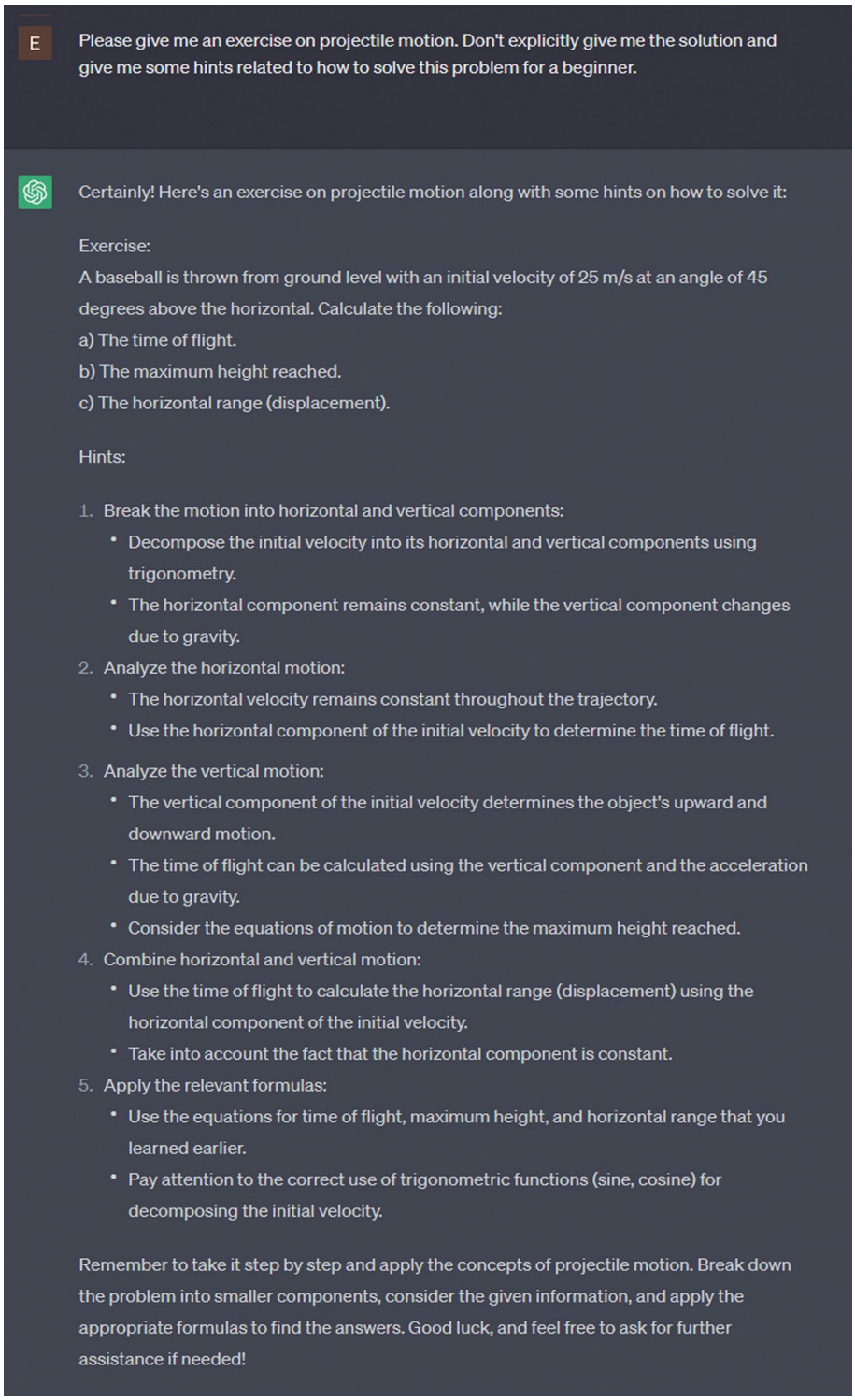

Fig. 2 ChatGPT generates a projectile motion exercise and give some hints to learners

γ denotes the lexical description of this variable v (e.g., the name of the physics variable), ϕ represents the corresponding physics symbol, and λ indicates the corresponding physical quantity having a numerical value and a unit. Intuitively, for calculation, a variable

Table 1 Illustration example

Variable	Type	Host	Description	Symbol	Physical quantity
var1	Object-related	A car	Mass	m	500 kg
var2	Object-related	A car	Speed	v	20 m/s
var3	Environmental	Environment	Time	t	10 s
var4	Object-related	A car	Breaking force	F	Unknown

can be classified into known variables and unknown variables depending on whether the physical quantity λ is given or not.

In addition, we propose some constraints on analyzed objects $\mathcal{O}$ and variables $\mathcal{V}$. First, the analyzed object should be tokens described in the problem. Second, every variable should be attached to its host, whether it is within the analyzed object or the surrounding environment. Both instructors and students could gain advantage from the AI system with the ability of generating hints automatically (Tran et al., 2021). Hence, the motivation for extracting physics variables and analyzing objects is that the extracted variables can serve as hints for learners. And hint generation as well as providing methods of step-by-step solution help students understand the problem.

In the following, we will use an example to illustrate the defined concepts regarding a computation physics problem. Suppose the text description of a physics problem is given as follows:

> *"A car of mass 500kg is travelling at 20m/s. The driver sees a red traffic light ahead and slows to a halt in 10s. Calculate the braking force provided by the car.* [2]*"*

Based on the above problem description, there is only one object to be analyzed in $\mathcal{O}$, i.e., 'a car'. The corresponding variables in this exercise are summarized in Table 1. There are three object-related variables, i.e., the 'mass' and 'speed' of this car, and 'breaking force' provided by this car, and one environmental variable, i.e., 'time'. Compared to the original text description of this problem, 'mass' and 'breaking force' are tokens in $\mathcal{T}$, while the terms 'speed' and 'time' do not explicitly appear in this text and need to be inferred by the model.

In order to solve a particular physics problem, we define the thinking path that includes the following components: (1) point out the given information, i.e., list the known variables and the unknown variable that is going to be solved (denoted as v^*); (2) identify relevant physics principles, e.g., conservation law of energy, Newton's law of motion, and their relevant equations[3]; (3) keep calculating the number of unknown variables with the relevant equations by substituting the known values until v^* is confirmed; (4) summarize the final answer. This thinking path can help to build steps in the chain of thought (Wei et al., 2022b) annotations and formalize the prompt with few-shot exemplars.

As pretrained large language models, like GPT, have excellent performance in many natural language processing tasks, this work aims to explore how to use ChatGPT to fulfill the task that transforms the text content of an exercise into the desired output.

[2] This physics problem is a worked example in the physics textbook (Sang et al., 2014, p. 39).

[3] In this paper, we assume that one physics principle corresponds to one equation.

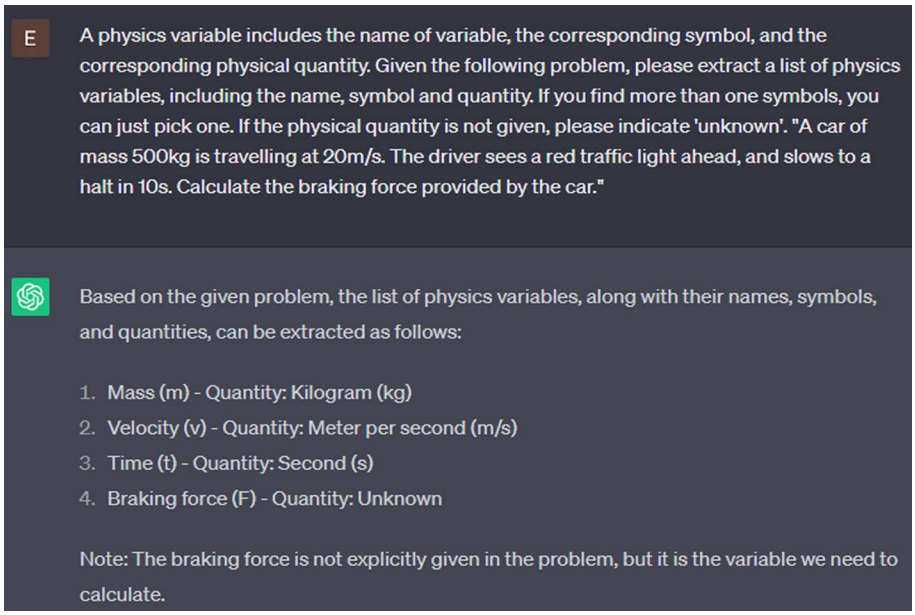

Fig. 3 Give instruction to ChatGPT on extracting variables

Recently, the reasoning ability of large language models has been unlocked by chain-of-thought (CoT) with few-shot (Wei et al., 2022b) or zero-shot (Kojima et al., 2022) prompting and the experiments show that CoT improves the performance on the tasks of arithmetic, commonsense, and symbolic reasoning. This research work aims to investigate the performance of ChatGPT in solving physics calculation problems by adapting zero-shot-CoT (Kojima et al., 2022).

Prompting for physics

Demonstration on extracting variables

The in-context learning ability (Brown et al., 2020) is introduced in GPT-3 which can generate the expected output under a natural language instruction. Given the input text of a physics computation problem, we can extract physics variables by using the following prompt:

> **Prompt**: *"A physics variable includes the name of the variable, the corresponding symbol, and the corresponding physical quantity. Given the following problem, please extract a list of physics variables, including the name, symbol, and quantity. If you find more than one symbol, you can just pick one. If the physical quantity is not given, please indicate 'unknown'. [problem]"*

This prompt contains four components: (1) context: tell the model what information a variable should include; (2) task instruction: extract a list of physics variables from a given problem; (3) constraint: guide the model to do some filter on generating the output; (4) input: [problem] denotes the text of the problem. The reason for adding constraints of symbols in this prompt is that the notation for physics variables can vary depending on the context of the problem, such as with or without subscripts. For example, both F and F_b (with the subscript "b" indicating braking) are suitable as the

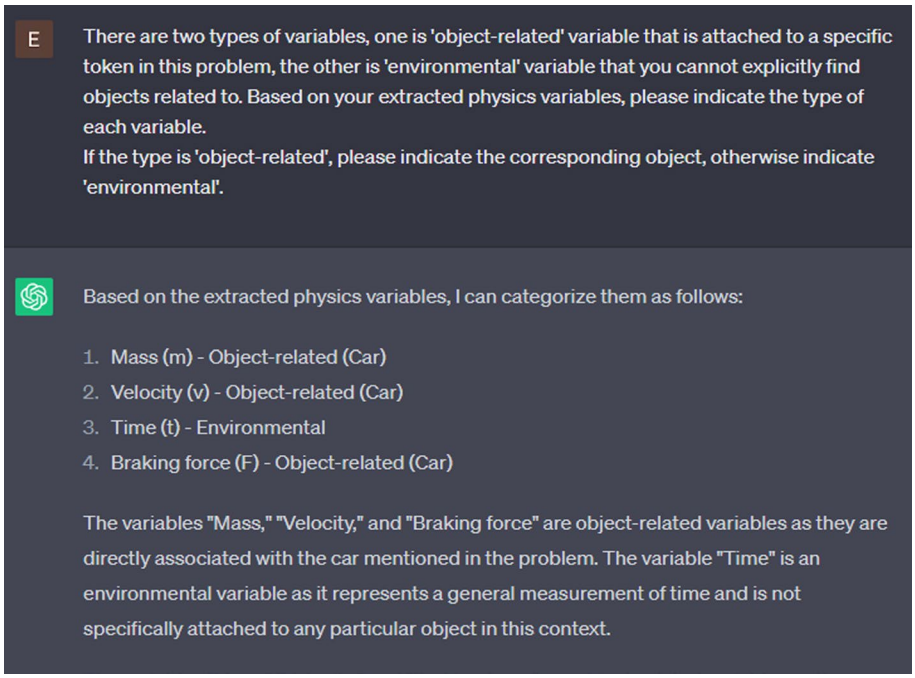

Fig. 4 Given an instruction to ChatGPT on classifying variables

candidate symbol for 'breaking force' in this problem. From the response shown in Fig. 3, ChatGPT was able to indicate four physics variables (mass, velocity, time, and breaking force) from a given problem in terms of successfully specifying the name, the corresponding quantity, and our predefined variable type.

Based on the retrieved variables, we designed the following prompt to classify the variables.

> **Prompt**: *There are two types of variables: one is an 'object-related' variable that is attached to a specific token in this problem, and the other is an 'environmental' variable that you cannot explicitly find objects related to. Based on your extracted physics variables, please indicate the type of each variable. If the type is 'object-related', please indicate the corresponding object, otherwise indicate 'environmental'.*

The context of this prompt is to inform ChatGPT of the information on variable types and the instruction is a binary classification task. The constraint is that an object-related variable needs to be attached to the corresponding object in terms of a token in this problem. From the result in Fig. 4, ChatGPT can accurately classify each variable and indicate its corresponding object as required, suggesting that ChatGPT can understand the problem by recognizing the semantics of variables. To summarize the results in Figs. 3 and 4, we can use a prompt to generate a table summarizing the results, as shown in Fig. 5, in accordance with our proposed model for defining a physics calculation problem. The benefit of utilizing ChatGPT to summarize the variables in a physics problem is that decomposing the problem into a set of variables with physics semantics (e.g., description, physical quantity, and corresponding object) can help students get

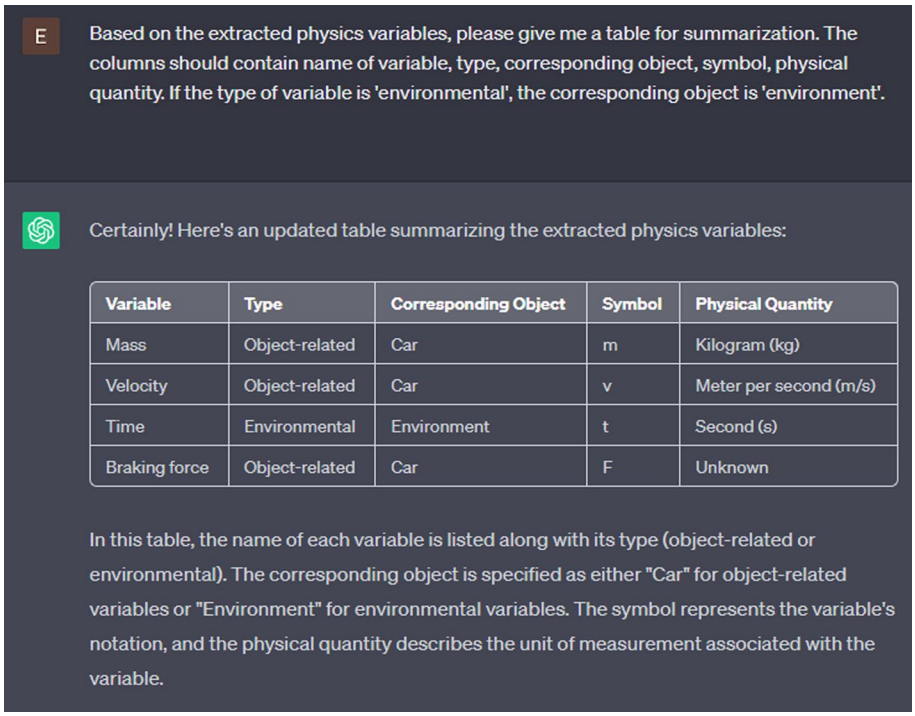

Fig. 5 Given an instruction to ChatGPT on summarizing the extracted variables

Table 2 ChatGPT solved the example problem

Prompt:	Based on the extracted physics variables, please solve this problem. Let's think step by step
ChatGPT:	Step 1: Given information, show the list of extracted variables Step 2: Identify relevant physics principles. To calculate the braking force, we can use Newton's second law of motion Step 3: Calculate the acceleration. Using the formula for acceleration ($a = (v' - v) / t$) Step 4: Calculate the braking force. Using Newton's second law ($F = m * a$), we can calculate the braking force. Step 5: Final Answer. The braking force provided by the car is -1000 N (opposite direction of motion).

Due to the space limitation, part of the response is shown

hints from the system and enable them to try to solve the problem themselves in subsequent steps.

Demonstration on generating solution

With the extracted variables from ChatGPT, we can adopt the zero-shot-CoT prompting technique (Kojima et al., 2022) to solve this problem. From the response shown in Table 2, ChatGPT came up with a solution in five steps. In particular, in step 1, list the known physics variables, including symbol and quantity, and specify the unknown variable that needs to be solved. In step 2, in order to calculate the unknown variable (braking force F), ChatGPT identifies relevant physics principles, e.g., Newton's second law of motion ($F = m \cdot a$). Calculating the unknown braking force is transferred into the calculation acceleration a in this case. In step 3, calculate the acceleration by using the formula $a = (v' - v)/t$ where the substituted values on the right side can refer to the

Table 3 Permute variables to generate a new computation problem

Original problem	New generated problem
"A car of mass 500 kg is travelling at 20 m/s. The driver sees a red traffic light ahead and slows to a halt in 10 s. Calculate the braking force provided by the car." (Sang et al. 2014, p. 39)	A car with a mass of 800 kg experiences a constant braking force of 400 N. If the car comes to a halt from an initial velocity of 15 m/s, calculate the time it takes for the car to stop.
Mass: Symbol: *m*, Quantity: 500 kg	**Mass:** Symbol: *m*, Quantity: 800 kg
Velocity: Symbol: *v*, Quantity: 20 m/s	**Velocity:** Symbol: *v*, Quantity: 15 m/s
Time: Symbol: *t*, Quantity: 10 s	**Braking force:** Symbol: *F*, Quantity: 400
Braking force: Symbol: *F*, Quantity: Unknown (to be solved)	**Time:** Symbol: *t*, Quantity: Unknown (to be solved)
principle: (1) Newton's second law of motion; (2) definition formula of acceleration	

extracted variables in step 1. In step 4, calculate the braking force by using the principle (i.e., Newton's second law of motion) indicated in step 2. In the final step, ChatGPT summarizes the final answer with more semantics, e.g., indicating the direction of the force is in the opposite direction of motion. In summary, there are two key intermediate steps for solving this problem. First, use the formula for acceleration (acceleration = change of velocity/time). Second, apply Newton's second law of motion (force = mass · acceleration).

Demonstration on generating new problems

ChatGPT is not only capable of solving a physics computation problem but also generating a new problem that share the same physics principle with the given input problem. For example, we use a permute-instruction prompt "Permute the physics variables and give me another problem. [input problem]" to accomplish this task. For comparison, the original problem and the newly generated problem are placed together in Table 3. These two problems share the same set of physics variables but with different unknowns for calculation. In addition, these two problems share the same physics principle for the problem solution. Specifically, Newton's second law of motion connects mass and force with acceleration, and acceleration definition formula connects velocity and time with acceleration. This particular function of ChatGPT can help instructors prepare homework materials that allow students to practice different exercises related to the same principles.

Similarly, we can use ChatGPT to generate a problem related to the given physics principles. For instance, we can ask ChatGPT to give a problem that contains the principles of momentum conservation law and Newton motion laws shown in Fig. 6. Intuitively, the more principles involved in a calculation problem, the more difficult the problem becomes. ChatGPT can help instructors automatically generate exercises at different levels and provide personalized learning materials as a practice to students who are not familiar with some particular topics. Besides, according to Gerjets et al. (2004), presenting modular worked examples that break down complex solutions (e.g., multiple physics principles) into smaller meaningful solution elements (e.g., individual principle), can help reduce learners' intrinsic cognitive load. As a result, ChatGPT can help instructors design modular worked examples (Gerjets et al., 2004) so as to improve learners' problem-solving skills.

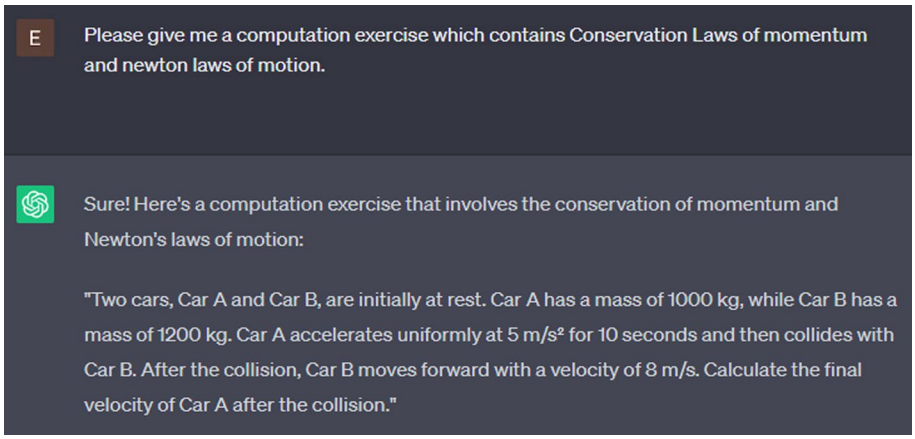

Fig. 6 Propose a problem related to some given principles

Table 4 ChatGPT solve rate

	Solve rate (%)	Error
One equation	80	Computation, direction
Two equations	80	Computation, direction

Performance on solving problems

There are some publicly accessible datasets for evaluating the solving rate of large language models on math problems, but to the best of our knowledge, there is still no benchmark dataset for physics problems. For evaluating the solve rate of ChatGPT, we manually picked up 20 calculation physics problems on dynamics topic and used the zero-shot-CoT (Kojima et al., 2022) prompt "Please solve the following problem. Let's think step by step. [problem]". In addition, according to the number of related equations to solve the problem, the problems were divided into two sets: (1) one physics equation; (2) at least two physics equations.

The result of the solve rate is shown in Table 4. Two domain experts who are high school physics teachers manually examine each solution provided by ChatGPT and find that ChatGPT can accurately point out given variables from the text of the problem and identify relevant physics principles. For example, for the problem "*A trolley with a 5.0cm long card passed through a single light gate. The time recorded by a digital timer was 0.40s. What was the average speed of the trolley in m/s?*" (Sang et al., 2014, p. 4), ChatGPT can indicate two known variables distance and time and one unknown variable average speed, and decide to use the formula 'speed=distance/time' for speed calculation. If the problem is related to only one principal formula within the format '$var3 = var1 \times var2$', ChatGPT can successfully solve it. This suggests that large language models, like ChatGPT, are adept at handling problems requiring simple reasoning and deduction.

Furthermore, ChatGPT is capable of doing unit transformation[4] to calculate the physical quantity as the problem requires. However, regarding the problem related to some

[4] Unit transformation is the process of converting a measurement from one unit to another while maintaining an equivalent quantity.

equations with complex expressions, like the equation of motion $s = v_0 t + \frac{1}{2}at^2$, ChatGPT sometimes makes errors in computation. On the other hand, when computation concerns the signs of vector variables,[5] ChatGPT sometimes makes errors in judging the direction of vectors. LLM is just a language model and it sometimes struggles with performing complex arithmetic operations. Using an external calculator (Gao et al., 2023) developed by the program model can help LLMs lower the chance of making arithmetic mistakes.

Discussion

Integrating ChatGPT content into course activities

Concerning the incorporation of prompts generated by ChatGPT into physics course activities, we propose the following structured approach to ensure relevance, accuracy, and pedagogical efficacy.

To begin with, teachers should clearly define the learning objectives. Before generating any prompts, teachers should have a comprehensive understanding of what they aim for their students to learn, whether that's foundational concepts, problem-solving skills, or a combination of the two. Once the teaching and learning objectives are set, teachers can then leverage ChatGPT to generate pertinent questions by inputting specific topics or concepts related to physics. For example, when addressing kinematics, one might input "Create a question related to the concept of acceleration." However, we strongly suggest that it's vital for teachers to remember that while ChatGPT is an advanced tool, it's not infallible. As such, teachers must rigorously review all generated prompts and refinement to ensure their relevance, accuracy, and appropriateness in terms of difficulty and adjust or reformulate the questions as needed. After curating these prompts, teachers can categorize them based on their complexity and intended use (e.g. 'warm-up questions,' 'discussion starters,' 'homework assignments' or 'formative assessment'). That is, some prompts might serve as introductory warm-up questions for lectures, while others could act as triggers for group discussions or even as components of graded assignments and exams. During lectures, some prompts can serve as a bridge to introduce new topics or recapitulate previous discussions. Moreover, teachers can use more challenging or open-ended prompts in collaborative settings to stimulate enriching group discussions or problem-solving sessions. Another way of making use of the ChatGPT prompts is to incorporate them into homework assignments or formative assessments as a continuous feedback loop with students is paramount, while ChatGPT can work well in providing real-time post-activity feedback.

Additionally, as the realm of AI, including tools like ChatGPT, is in a state of constant evolution, teachers ought to stay abreast of the latest developments, periodically reviewing and updating their approach to benefit from new features or capabilities. Collaborative efforts with peers can further enhance this, as sharing and refining prompts with colleagues can introduce diverse perspectives, enriching the overall pool of questions.

[5] In physics, a vector is a mathematical quantity that has both magnitude and direction.

Pedagogical benefits of using ChatGPT for learning physics

Using ChatGPT for the resolution of physics questions brings to the table an array of pedagogical advantages that are well-suited to modern educational needs. Among the foremost is the provision of immediate feedback. It's widely acknowledged that students flourish in their understanding when feedback is prompt. In the intricate world of physics, understanding where a mistake was made or having confirmations of correct reasoning can significantly impact a student's learning trajectory. ChatGPT stands out in this regard, offering instantaneous answers, be it for clarifications or corrections. This is very important for the learning of physics.

Equally noteworthy is the adaptive learning capability of ChatGPT. Every student is unique, possessing their strengths, weaknesses, and pace of grasping concepts. With the adaptability intrinsic to ChatGPT, the complexity of explanations is tailored in real-time based on the student's queries. Whether a learner requires a simplified overview or a comprehensive breakdown, the tool adjusts, ensuring that the explanations resonate with the individual's needs in learning physics.

In today's digital age, the conventional 9-to-5 study timetable doesn't resonate with all. The 24/7 availability of ChatGPT addresses this modern challenge. Students, whether they're night owls, early risers, or weekend scholars, have the liberty to seek assistance whenever a query strikes—a feature that conventional classrooms or even tutors might not always offer. With such flexibility, students can learn physics whenever and wherever they prefer.

Moving beyond traditional rote learning, ChatGPT champions interactive learning. The dynamic of engaging in a dialogue, as opposed to the passivity of mere reading, enriches comprehension. The student can delve deeper, ask ancillary questions, and even seek repeated clarifications until clarity is achieved. In this way, they can self-regulate their own learning pace and content, which tends to increase students' learning motivation.

A common challenge in education is the variance in students' foundational knowledge. Some might grapple with advanced concepts simply because their basic grounding is shaky. Recognizing such gaps, ChatGPT can elucidate fundamental principles, paving the way for a more solid understanding of intricate topics. This personalized learning experience is conducive to students' effective learning of physics.

While the benefits are myriad, ChatGPT, with its extensive capabilities, is best viewed as a supplementary resource. It complements but doesn't necessarily supplant, traditional teaching avenues like textbooks or human instructors. Its interactive format offers students an alternative means to engage with content, diversifying their learning experience.

Yet, one of the underrated benefits of ChatGPT is its creation of a safe learning environment. In traditional settings, the fear of judgment can deter students from voicing out their doubts, especially if they deem them too rudimentary. With ChatGPT, such hesitations dissipate. The platform becomes a sanctuary where every question is valid, devoid of potential ridicule or judgment.

Nevertheless, it's imperative to approach this tool with a balanced perspective. While ChatGPT offers numerous pedagogical advantages, an exclusive dependency on it might be counterproductive. The nuances of certain advanced topics in physics might

demand the seasoned expertise of human educators. Furthermore, an over-reliance on technology can diminish the value of human interaction and guidance in the educational journey. Hence, the ideal approach positions ChatGPT as an adjunct to traditional educational methodologies, harnessing the best of both worlds for a comprehensive learning experience.

ChatGPT performance in solving physics calculation problems

Based on the depth of understanding and cognitive skills required to answer them, we proposed the following taxonomy of physics questions.

1. Recall Questions. At the foundation of this taxonomy lie Recall Questions. These are straightforward queries that primarily test the student's memory of specific terms, facts, or basic concepts.
2. Comprehension Questions. Ascending from recall, we encounter Comprehension Questions. Moving beyond mere regurgitation of facts, these questions task students with interpreting or elucidating known concepts.
3. Application Questions. As the name suggests, Application Questions revolve around the practical utilization of knowledge. These questions necessitate students to employ known principles in fresh or unfamiliar situations.
4. Analysis Questions. The complexity ratchets up with Analysis Questions. In these, students dissect intricate concepts or problems, reducing them to their core components. This breakdown assists in discerning underlying structures or patterns.
5. Problem Solving Questions. At the pinnacle of cognitive complexity, are Problem Solving Questions. These are multi-faceted and intricate, obliging students to weave together various threads of knowledge, logical reasoning, and critical thinking.

To evaluate how well ChatGPT performs in answering the five types of questions above, the example of projectile motion was taken. For the recall question, 'Where does projectile motion usually take place and what kind of path does it look like?'. From Fig. 1, ChatGPT accurately identifies air and curve as the place and the trajectory of movement regarding this question. For the comprehension question, 'What force influences the projectile motion?'. ChatGPT is able to indicate gravity as the force that actually pulls the flying object downward. For the application question, 'How to solve motion in a curved path?'. ChatGPT gives the answer that breaks down the complicated motion into two orthogonal components, namely horizontal and vertical components. For the analysis question, 'How to analyze the horizontal and vertical motion accordingly?'. According to Fig. 1, ChatGPT points out that the horizontal component remains constant due to no external forces acting horizontally, whereas the vertical velocity changes over time due to gravity. For the problem-solving question, ChatGPT is evaluated based on two aspects, generating a new exercise given a specific topic (e.g., projectile motion), and generating a solution for a particular exercise. As seen from Fig. 2, ChatGPT provides an exercise connecting to a real-life example, i.e., throwing a baseball, and the three subsequent questions are all related to the important concepts behind projectile motion, e.g., time of flight, maximum height, horizontal range. Regarding solving the problem, ChatGPT can be prompted to generate hints related to the answers to the analysis question. In addition, ChatGPT can do some

fine-grained tasks in solving an exercise, like extracting physics variables as shown in Fig. 3, classifying variables as defined in Fig. 4.

Conclusions and future work

This paper has presented preliminary ideas regarding the ways in which ChatGPT can facilitate the learning of physics. This work defines a model for a physics calculation problem which can be further developed as hints to help learners understand the problem. In addition, this paper has provided some demonstration cases on how to use ChatGPT for learning physics problems, including extracting variables from the input problem, generating new problems, and solving the problem step by step. According to the demonstration results, we can find that ChatGPT can understand the problem in terms of recognizing the semantics of variables and indicating the principles behind the problem.

Nevertheless, there are several limitations of the present research. Firstly, its primary objective was to conceptualize the possibility of employing ChatGPT to tackle physics problems and to envision its application within physics classrooms, therefore, the scope of this investigation was limited to exploring this potential, without any empirical implementation in a real classroom setting. This stands as one of the limitations of our study and also paves the way for future research directions. Moreover, our research only theoretically discussed the potential of using ChatGPT in physics classrooms, we have not yet conducted experiments with real learners to validate the efficacy of these recommendations in aiding physics learning. Furthermore, we have not interviewed teachers and students to gather their perspectives on this method of learning. Therefore, our findings lack empirical support in these aspects. Nevertheless, the main objective of this study was to explore the potential pedagogical benefits of using ChatGPT in physics and to evaluate its proficiency in solving physics problems. As such, we did not conduct empirical research or collect data from real learners and teachers, which means we have yet to substantiate its application value and benefits in the educational domain. Future research may consider addressing these gaps in the literature by providing more practical insights.

In future work, we will explore hint generation and create a teacher model that can procedurally generate hints for learners to solve physics problems. With a hint generator, students can be guided to better understand the problem, from solely giving a numeric answer to demonstrating a higher level of explainability. Furthermore, we will integrate the student profiles into LLMs to provide more personalized feedback, learning paths, and scaffoldings (Zou et al., 2021; Xie et al., 2019; Zou & Xie, 2018; Xie et al., 2017). Another important data source for improving the interaction experience with LLMs is the behavioral data from various platforms (e.g., social media platforms, learning management systems, open learning resources, and so on) (Wang et al., 2021; Wong et al., 2020; Xie et al., 2015).

Appendix A: Ten ways to improve physics using ChatGPT

1. Using ChatGPT to acquire some background about a physicist Suggested prompt: Please tell me some stories about physicist [name].

2. Using ChatGPT to recommend study resource Suggested prompt: Please recommend some study resources in terms of textbooks, online courses, and other resources to aid in learning physics.
3. Using ChatGPT to explain complex concepts and terminology Suggested prompt: Please give a clear description of the concept of [concept name].
4. Using ChatGPT to offer real-world applications related to physics Suggested prompt: Physics concepts are applied in various real-world scenarios. Please provide examples of how [concept name] is applied.
5. Using ChatGPT to extract physics variables from problem Suggested prompt: Given the following problem, please extract a list of physics variables, including the name, symbol, and quantity. [problem]
6. Using ChatGPT to summarize variables from problem Suggested prompt: Please give me a table for summarization based on the extracted variables from the following problem. The columns should contain the name of the variable, symbol, and physical quantity. [problem]
7. Using ChatGPT to re-generate new problems with variables permutation Suggested prompt: Given the following problem, please permute the physics variables, and give me another problem. [input problem]
8. Using ChatGPT to indicate physics principles from a given problem Suggested prompt: Give the following problem, please indicate the related physics principles. [problem]
9. Using ChatGPT to generate problems related to given physics principles Suggested prompt: Please give me a computation exercise, which contains [physics principles].
10. Using ChatGPT to solve physics calculation problems Suggested prompt: Please solve the following calculation problem. Let's think step by step. [problem]

Abbreviations

LLM	Large language model
PLM	Pretrained language model
NLP	Natural language processing
GAI	Generative artificial intelligence
MWP	Math word problem
CoT	Chain of thought
CLT	Cognitive load theory

Acknowledgements
Not applicable.

Author contributions
YL: Conceptualization, Methodology, Formal analysis, Data curation, Writing—original draft. DZ: Conceptualization, Methodology, Supervision, Writing—review and editing. HX: Methodology, Data curation, Supervision, Writing—review and editing, Funding acquisition. FLW: Supervision, Resources, Project administration, Writing—review and editing.

Funding
The research has been supported by the Direct Grant (DR23B2) and the Faculty Research Grants (DB23A3 and DB23B2) of Lingnan University, Hong Kong.

Availability of data and materials
Not applicable.

Code availability
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors agree to publish this paper, which has not been published elsewhere.

Competing interests
The authors declare no competing interests.

Received: 22 June 2023 Accepted: 10 October 2023
Published online: 20 October 2023

References

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & Agarwal, S. (2020). Language models are few-shot learners. In *Proceedings of the 34th international conference on neural information processing systems* (pp. 1877–1901).

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., et al. (2023). PaLM: Scaling language modeling with pathways. *Journal of Machine Learning Research, 24*, 1–113.

Drori, I., Zhang, S., Shuttleworth, R., Tang, L., Lu, A., Ke, E., Liu, K., Chen, L., Tran, S., Cheng, N., & Wang, R. (2022). A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level. In *Proceedings of the national academy of science* (Vol. 119, p. e2123433119).

Fedus, W., Zoph, B., & Shazeer, N. (2022). Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *The Journal of Machine Learning Research, 23*(1), 5232–5270.

Frieder, S., Pinchetti, L., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., Chevalier, A., & Berner, J. (2023). Mathematical capabilities of ChatGPT. arxiv:2301.13867

Gao, L., Madaan, A., Zhou, S., Alon, U., Liu, P., Yang, Y., Callan, J., & Neubig, G. (2023). PAL: Program-aided language models. In *Proceedings of the 40th international conference on machine learning* (Vol. 202, pp. 10764–10799).

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32*, 33–58.

Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Taylor, R., & Chartash, D. (2023). How does ChatGPT perform on the United States medical licensing exams? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education, 9*, e45312. https://doi.org/10.2196/45312

He-Yueya, J., Poesia, G., Wang, R. E., Goodman, N. D. (2023). Solving math word problems by combining language models with symbolic solvers. arxiv:2304.09102

Hwang, G.-J., & Chen, N.-S. (2023). Editorial position paper: Exploring the potential of generative artificial intelligence in education: Applications, challenges, and future research directions. *Educational Technology and Society*, *26*(2), i–xviii. https://doi.org/10.30191/ETS.20230426(2).0014

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550.

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. *Advances in Neural Information Processing Systems, 35*, 22199–22213.

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaÿno, C., et al. (2023). Performance of ChatGPT on USMLE: Potential for uppercase AI assisted medical education using large language models. *PLoS Digital Health, 2*(2), e0000198.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*(1/2), 1–8.

Phan, H. P., Ngu, B. H., & Yeung, A. S. (2017). Achieving optimal best: Instructional efficiency and the use of cognitive load theory in mathematical problem solving. *Educational Psychology Review, 29*, 667–692.

Sang, D., Jones, G., Chadha, G., & Woodside, R. (2014). *Cambridge international as and a level physics coursebook with CD-ROM*. Cambridge University Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation, 55*, 37–76.

Tran, S., Krishna, P., Pakuwal, I., Kafle, P., Singh, N., Lynch, J., & Drori, I. (2021). Solving machine learning problems. In *Asian conference on machine learning* (pp. 470–485).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, l., & Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 6000–6010).

Wang, J., Xie, H., Wang, F. L., Lee, L.-K., & Au, O. T. S. (2021). Top-n personalized recommendation with graph neural networks in MOOCs. *Computers and Education: Artificial Intelligence, 2*, 100010.

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., & Zhou, D. (2023). Self-consistency improves chain of thought reasoning in language models. In *International conference on learning representations*.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., & Chi, E.H. (2022a). Emergent abilities of large language models. *Transactions on Machine Learning Research*. Retrieved from https://openreview.net/forum?id=yzkSU5zdwD

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E. H., Le, Q. V., & Zhou, D. (2022b). Chain-of-thought prompting elicits reasoning in large language models. In *Advances in neural information processing systems*.

Wong, T.-L., Xie, H., Zou, D., Wang, F. L., Tang, J. K. T., Kong, A., & Kwan, R. (2020). How to facilitate self-regulated learning? A case study on open educational resources. *Journal of Computers in Education, 7*, 51–77.

Xie, H., Zou, D., Lau, R. Y., Wang, F. L., & Wong, T.-L. (2015). Generating incidental word-learning tasks via topic-based and load-based profiles. *IEEE Multimedia, 23*(1), 60–70.
Xie, H., Zou, D., Wang, F. L., Wong, T.-L., Rao, Y., & Wang, S. H. (2017). Discover learning path for group users: A profile-based approach. *Neurocomputing, 254*, 59–70.
Xie, H., Zou, D., Zhang, R., Wang, M., & Kwan, R. (2019). Personalized word learning for university students: A profile-based method for e-learning systems. *Journal of Computing in Higher Education, 31*, 273–289.
Zhang, D., Wang, L., Zhang, L., Dai, B. T., & Shen, H. T. (2019). The gap of semantic parsing: A survey on automatic math word problem solvers. *IEEE Transactions on pattern analysis and machine intelligence, 42*(9), 2287–2305.
Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., & Du, Y. (2023). A survey of large language models. arxiv:2303.18223
Zou, D., Wang, M., Xie, H., Cheng, G., Wang, F. L., & Lee, L.-K. (2021). A comparative study on linguistic theories for modeling EFL learners: Facilitating personalized vocabulary learning via task recommendations. *Interactive Learning Environments, 29*(2), 270–282.
Zou, D., & Xie, H. (2018). Personalized word-learning based on technique feature analysis and learning analytics. *Journal of Educational Technology & Society, 21*(2), 233–244.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Yicong Liang is a Ph.D. student at the School of Science and Technology, Hong Kong Metropolitan University, Hong Kong. His research interests include AI in education and generative AI applications.

Di Zou is the Head and Associate Professor of the Centre for English and Additional Languages, Lingnan University, Hong Kong. Her research interests include language learning and technology-enhanced learning. She has over 150 publications. She has been listed as one of the World's Top 2% Scientists by Stanford University in 2021 and 2022. She is the Associate Editor of Computers & Education and Computers & Education: X Reality.

Haoran Xie is the Head and Associate Professor of the Department of Computing and Decision Sciences at Lingnan University, Hong Kong. His research interests include artificial intelligence, big data, and educational technology. He has over 350 publications. He is the Editor-in-Chief of Computers & Education: Artificial Intelligence, Computers & Education: X Reality, and Natural Language Processing Journal. He has been listed as one of the World's Top 2% Scientists by Stanford University in 2021 and 2022.

Fu Lee Wang is the Dean of School of Science and Technology and Professor at Hong Kong Metropolitan University, Hong Kong. His research interests include e-learning and information retrieval. Professor Wang has over 300 publications and 40 grants with more than 80 million Hong Kong dollars. He was also the Chair of ACM Hong Kong Chapter and IEEE Hong Kong Section Computer Society.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com