

Generative AI as a lab partner: A case study

Sebastian Kilde-Westberg[✉],* Andreas Johansson[✉], and Jonas Enger[✉]

Department of Physics, *University of Gothenburg*, SE 412 96 Gothenburg, Sweden



(Received 18 December 2024; accepted 8 August 2025; published 27 August 2025)

[This paper is part of the Focused Collection in Artificial Intelligence Tools in Physics Teaching and Physics Education Research.] Generative AI tools, including the popular ChatGPT, have had a significant impact on discourses about future work and educational practices. Previous research in science education has highlighted the potential of generative AI in various education-related areas, including generating valuable discussion material, solving physics problems, and acting as a tutor. However, little research has been conducted on the role of generative AI tools in laboratory work, an essential part of science education, and physics education specifically. Here, we show various ways in which high school students use ChatGPT during a physics laboratory session and discuss the relevance of using generative AI tools to investigate acoustic levitation and the speed of sound in air. Additionally, employing variation theory as a theoretical lens in the analysis, we identify how generative AI can be used to further develop students' problem-solving skills in the physics laboratory. However, although our study shows that generative AI tools may handle some relevant questions and problems during laboratory work, the teacher still plays a crucial role in identifying students' needs and capabilities of understanding the potential and limitations of generative AI. Finally, this study serves as an important point of discussion regarding the ways in which students need support and training to efficiently utilize generative AI to further their learning of physics.

DOI: [10.1103/ggy1-3kjk](https://doi.org/10.1103/ggy1-3kjk)

I. INTRODUCTION

The introduction of generative artificial intelligence (GenAI) tools, spearheaded by the public release of the chatbot ChatGPT [1], has led to a broad debate on their impact on education and future workforce needs [2–5]. In physics education research (PER), recent studies involving GenAI tools have examined their ability to generate valuable (i.e., reliable or interesting, but sometimes incorrect) answers to physics problems [6–9]. Although a need has been identified for physics instructors to incorporate artificial intelligence (AI) in the classroom, there is still little focus on how GenAI tools can be used in the physics laboratory, a crucial part of students' education when learning physics.

In this study, we investigate the potential future role of GenAI tools, such as ChatGPT, in the context of a high school physics laboratory. We do this in part by analyzing students' awareness during the lab by drawing on the variation theory of teaching and learning as a theoretical framework. Specifically, our work provides valuable

insights into how students are utilizing the freely available version of ChatGPT, GPT-3.5 at the time of data collection, with varying prior knowledge of the phenomenon they are studying, as well as with little to no prior training on how to use GenAI tools in school. Our aim is to contribute to the discussion on the ways in which GenAI tools may impact future laboratory work and implications for teachers, both in planning and executing such lessons. We do this by answering the following research questions (RQs):

RQ1: What do students perceive as useful interactions with GenAI tools like ChatGPT during a lab session?

RQ2: How do high school students actively engage with ChatGPT to solve problems during a lab investigating the speed of sound using acoustic levitation?

RQ3: How can interacting with ChatGPT during the lab aid students in problem solving and conceptual understanding?

The structure of the paper is as follows: In Sec. II, we provide a broader context for the study by conducting a literature review to identify prior research that can help inform the use of GenAI in the physics laboratory. Section III describes the theoretical framework of variation theory, which is a theory of teaching and learning that was used in the current study. By introducing a theory of teaching and learning, it provides a clear lens through which we can connect our results to practical implications for teaching and learning in the physics laboratory. In the

*Contact author: sebastian.lofgren@physics.gu.se

Published by the American Physical Society under the terms of the *Creative Commons Attribution 4.0 International license*. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by *Bibsam*.

current study, we draw upon our theoretical framework specifically in analyzing the results as a way to answer RQ3. Our lab design, as well as the study design, is presented in Sec. IV, which also includes a discussion of the study's limitations and how the data were analyzed. Our findings are presented in Sec. V. In Sec. VI, the RQs are discussed in relation to our findings. There, we also discuss implications for teaching and learning in the science laboratory. Finally, Sec. VII provides a summary of the paper, including an outlook that relates the findings of the current paper to future research needs in the field.

II. BACKGROUND

PER involving modern GenAI tools (postpublic release of ChatGPT) has had a broad focus, including comparing different models' performance on tasks [10,11], performance on physics problems [8,9,12–15], ways in which physics teachers can use these tools [7,16–20], and the potential for interactions between students and GenAI [6,21–24]. A common theme that can be observed in PER is that the use of GenAI tools, whether for students, teachers, or researchers, must be carefully considered to achieve a beneficial outcome. Regarding the potential use of GenAI tools for students, a primary benefit has been identified to be connected to increased motivation and engagement [25]. However, a recent systematic review on the influence of ChatGPT on student engagement has found that the increased engagement can, in some instances, be linked to cheating and reduced critical thinking, resulting from overreliance on answers provided by chatbots during problem-solving activities [26]. These findings reinforce the notion that the use of such tools in educational contexts requires careful consideration.

In science education, research about incorporating GenAI tools in laboratory work has primarily been done in the context of chemistry education, including comparing different models [27], aiding students developing methods [28], producing code to assist in data analysis [29], and in writing lab reports [30]. Two studies, both conducted at the university level and utilizing ChatGPT with the freely available version, GPT-3.5, involved students who had little to no prior experience with GenAI tools [28,29]. In using ChatGPT to help students produce code to analyze lab results, students found the GenAI tool to be useful and reported an increased level of understanding of the programming language they used [29]. The other study had students use ChatGPT to help them develop methods for lab proposals in analytical and inorganic chemistry [28]. Results indicated that it was challenging to obtain useful information from ChatGPT about specific procedures, as it often provided some correct information alongside vague or incorrect information. Additionally, the authors found that students with more theoretical and experimental knowledge were able to utilize ChatGPT more effectively than novice students.

A link between students' prior knowledge and their ability to evaluate the quality of information provided by GenAI tools has also been identified in PER, outside laboratory contexts. In a study investigating students' ability to evaluate the scientific quality of answers to three physics problems, 102 first- and second-year undergraduate students were presented with 4 different answers to each problem, 3 incorrect and generated using ChatGPT, and 1 correct generated by the researchers [6]. Their findings identified that the students' ability to gauge the quality of the answer was closely linked to their prior knowledge of the underlying physics.

Finally, in addition to prior knowledge influencing students' ability to identify incorrect information provided to them by GenAI tools, their implicit trust in chatbots can also influence how they gauge answers they get when interacting with the tool. This was investigated in a study in which students used ChatGPT as a tutor [21]. They found that students' preconceptions about the capabilities of GenAI tools influenced their trust in answers provided by ChatGPT. Additionally, they also identified that many of the participating students blindly trusted answers provided by ChatGPT, pointing toward suggesting that students had received little prior teaching about GenAI tools or how they are typically developed.

III. VARIATION THEORY

The current study is exploratory in nature, and the findings presented in this paper were identified using an inductive approach, similar to applying grounded theory [31] as a research approach or conducting a thematic analysis [32]. In this section, we introduce the variation theory of teaching and learning as a theoretical framework to be used in discussing our findings, allowing for further analysis related to students' awareness, as well as discussing general principles on teaching and learning identified by the theory. As such, variation theory provides a lens through which we can discuss the implications of our findings for teaching and learning in the science laboratory.

Variation theory is a theory of teaching and learning asserting that for learning to take place, the learner must notice what is to be learned, and that noticing requires variation [33–36]. It argues that it is easier to notice what differs between two contexts than what is similar, and learning is to be understood as “learning to see” [35]. Here, we can identify that learning is closely connected to a learner, meaning that in variation theory, one must acknowledge that the process of learning about a specific concept or phenomenon within a particular subject cannot be viewed as an isolated event, void of individual differences, including previous experiences of what is to be learned. Rather, a teacher must account for the students when designing teaching sequences, because students' prior knowledge and experiences of the world affect how they discern the current content.

A. Object of learning, structure of awareness, and critical aspects

What is to be learned is referred to as an object of learning. It is related to a specific group of students and the surrounding educational context, including curriculum-imposed goals [35,36]. An object of learning is constituted by a direct and an indirect object of learning. The content (e.g., standing waves) to be learned refers to the direct, whereas a particular skill or capability to be developed (e.g., predicting how particles can be trapped using acoustic levitation) refers to the indirect object of learning [35]. Together, they make up the whole object of learning, e.g., using the concept of standing waves to explain and predict how to trap particles using acoustic levitation, which is then identified as what is to be learned.

For a particular object of learning, one must also consider how a particular group of students may initially perceive or discern it. How a student perceives an object of learning is related to their current structure of awareness, an idea stemming from the phenomenographic tradition [37]. In the phenomenographic tradition, it assumes that a learner's structure of awareness is their experiences of the world. The structure of awareness continuously changes, as humans can only be consciously aware of a limited set of the world simultaneously. When considering an object of learning, we are thus never capable of being actively aware of all parts, or aspects, of it at the same time. As a student, what is to be learned can, from this point of view, be understood as becoming aware of parts of the object of learning in more powerful ways. These parts, that the student has yet to discern, are known as critical aspects [37,38], and the structure of awareness dictates how students become aware of these aspects [39]. Finally, the object of learning can be divided into an intended (the teacher's aim prior to the current lesson), an enacted (what aspects are made possible to discern during the lesson), and a lived object of learning (what the students actually learned).

B. Putting it into practice

When planning teaching and learning situations, variation theory emphasizes that focus should be on how variation can be used in handling the content that is taught, not variation in terms of varying teaching methods, such as types of classroom activities [36]. By varying the handling of the content, one can highlight the critical aspects of the object of learning in ways that make them more clearly visible to the students. However, if the teacher is not cognizant of their students' current possible structure of awareness, introduced variations may shift their focus toward unproductive, misleading aspects of the phenomenon under study. Specifically, during a lesson, students are ideally better at discerning critical aspects of the phenomenon under study toward the end of the lesson. This means that the students have become more proficient in identifying what the important (critical) aspects are. The students'

capacity to better connect the critical aspects to the object of learning means that they have developed a more refined awareness of the phenomenon [39]. If students become aware of the intended object of learning during the lesson, it also means that the enacted object of learning is in line with the intended. Such an alignment relies on the students' discerning the critical aspects, but also that the students' structure of awareness is directed toward the object of learning.

Studies employing variation theory commonly focus on analyzing qualitative data, and the same holds for studies using this theory in PER [40–42]. The three studies mentioned here to exemplify the use of variation theory in PER employ similar setups in that they follow a study design where a lesson has been designed. However, Refs. [40,41] both have one set intervention design, whereas [42] iterates the design throughout the study, following a design-based research approach to study students' understanding and improve the lesson design. Results from all three studies provide fruitful insights into what learning is made possible with each specific lesson design, and how teachers' intervention can aid in further directing students' attention toward critical aspects. On a more general note, they also highlight how variation theory can serve as a powerful tool in designing and planning good teaching and learning sequences in physics.

In the context of the current study, variation theory allows us to identify and discuss the implications of the findings presented in this paper for designing fruitful teaching and learning sequences in the age of AI in education. This is done in part by identifying instances during the lab where students' awareness shifts, indicating that their conception of the object of learning, or of some critical aspect(s), has changed.

IV. METHOD

The current study is designed to identify different ways students might naturally find it useful to make use of GenAI tools readily available to them in a laboratory situation. To that end, we opted to let students have access to one of the most common ways they might have come in contact with AI, the current free version of ChatGPT. Further, a lab was designed with the research questions in mind. As such, we intentionally focused on constructing a lab session so that students had little to no direct knowledge about the phenomenon under study and would struggle to complete the lab without consulting ChatGPT. In this section, we provide an introduction to the lab that was designed, as well as the context of the study in terms of participants, ChatGPT, data collection, and the analysis process. Finally, limitations connected to the chosen approach are also discussed.

A. Experimental setup and lab design

In acoustic levitation, particles typically in the millimeter range are trapped midair due to the creation of an acoustic

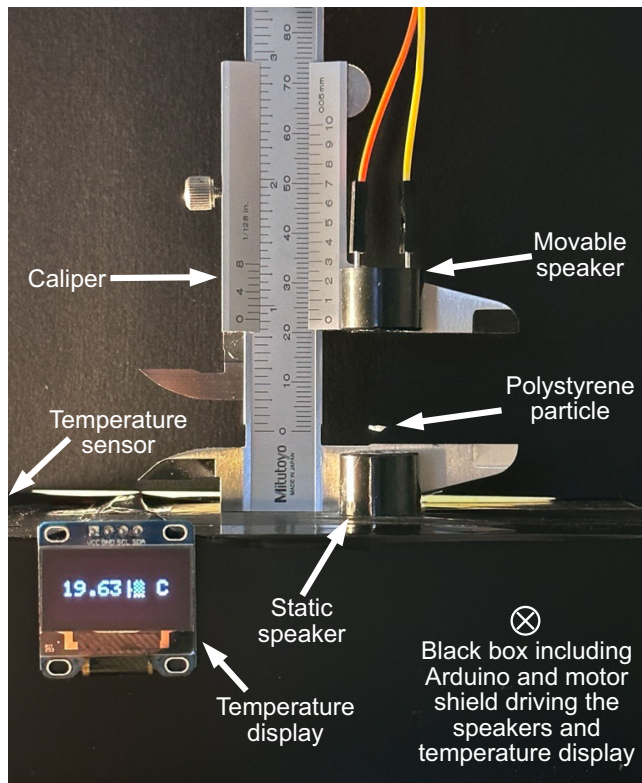


FIG. 1. The experimental setup, also known as LeviLab. The external low-voltage dc power supply powering LeviLab is not included in the figure.

field of standing waves. Millimeter-sized particles consisting of materials denser than air are attracted to the sound intensity nodes [43]. The experimental setup (LeviLab, [44]) comprises two ultrasonic speakers mounted vertically and opposite to one another on a caliper (Fig. 1). As the speakers, which are connected in phase, are moved axially, the standing waves between them shift. If a millimeter-sized object is placed resting on the bottom speaker grill, and the top speaker is slowly moved closer, the particle will be attracted to the closest node, which can either be above or below the speaker grill. This results in that the particle will, at certain distances when a node above the speaker grill is sufficiently close, seem to jump up and begin to hover midair. The configuration allows for the variable and precise measurement of the distance between the loudspeakers. In addition, LeviLab outputs a constant frequency of 40 kHz and is equipped with a thermometer and a digital display, which allows a measurement of the ambient temperature while performing experiments using the setup.

Additionally, the students were encouraged to also bring their physics textbook, formula book, and any calculator. The lab manual guided the students in obtaining data to analyze, from which they could calculate the wavelength and the speed of sound. The method for data collection was precisely described. However, the instructions related to the analysis had been intentionally designed to be somewhat

vague, asking students to perform a linear regression and to calculate the speed of sound using a formula that includes wavelength and frequency without explicating the exact form of the equation. Additionally, students were also tasked with comparing the speed of sound with another model in which the speed of sound depends on temperature. The intentional vagueness ensured that all or most groups would need to consult ChatGPT to identify the appropriate equation, as it was beyond their expected prior knowledge, regardless of educational level. Finally, to increase the likelihood of consulting ChatGPT, discussion questions were added, asking the students to construct a figure that describes how sound waves can be used to counteract gravity, identify potential sources of errors, explain the relevance of measuring temperature, and discuss applications of acoustic levitation. Students were considered to have successfully finished the lab if they finished the analysis, even if not all the discussion questions were answered.

B. Participants

The current study was conducted with Swedish second- and third-year high school students taking one of the three physics courses offered in the Swedish high school curriculum [45]. Having students from different levels of physics courses was made possible since the phenomenon of acoustic levitation is not included in the Swedish high school or compulsory school physics curriculum. However, in compulsory school, students learn about the physics of sound, specifically related to how it originates, propagates, can be reflected, as well as the particle model of matter including, e.g., pressure and density [46]. In the high school curriculum, however, sound is not covered in the course Physics 1, but again brought up in Physics 2 as part of learning about mechanical waves, and in Physics 3 in more advanced study of mechanical waves [45]. Thus, we expected the participants to have different, but at least some, knowledge about sound regardless of what physics course they are currently taking.

Participating students taking Physics 1 or 2 were sought at one high school, and students taking Physics 3 were sought at a university that offered the nonmandatory Physics 3 course to third-year high school students. The sampling of students was done by convenience sampling, where all willing students got to participate in the study. In total, seven groups participated: four with second-year students taking Physics 1, one with third-year students taking Physics 2, and two with third-year students taking Physics 3. All groups consisted of three students, except group 7 (Table I).

Furthermore, since each lab group participated in the study parallel to their normal lesson, there was some variance in the time each group had for the lab and interviews. On average, the pre- and postinterviews took 5 min each and between 5 and 10 min was required for transport to the classroom and informing the participating students about the study. In total, each group was given

TABLE I. A summary of each lab group, including pseudonyms for each of the participants ($n = 19$), educational year, what physics course they are taking, and their effective lab time rounded up to the nearest minute. *Pseudonyms that preserve anonymity regarding name and gender by choosing names from the list of the most common baby names in Sweden in 2005 [47], half boys and half girls. Names were organized and chosen in alphabetical order, with one exception in E, where the first boy's name, Ebbe, was skipped due to the similarity to Ebba.

Lab group	Students*	Educational year	Physics course	Lab time (minutes)
1	Adele, Adam, Agnes	2	1	43
2	Benjamin, Beatrice, Bill	2	1	44
3	Caroline, Carl, Cassandra	2	1	56
4	Daniel, Daniela, Dante	2	1	55
5	Ebba, Eddie, Edith	3	2	46
6	Fabian, Fanny, Felix	3	3	47
7	Gabriella	3	3	56

between 65 and 70 min to participate in the study. The effective lab time, which was either until they declared they were finished or the researcher present in the room had to stop them, for each group is presented in Table I.

Ethical considerations for this study have been made following local rules and guidelines set by the Swedish Research Council [48]. All participating students were old enough to understand the purpose of the study and the data collection method. Participants all gave written consent after being informed of the purpose of the study, what data were being collected, how it was to be handled to preserve their anonymity, and that participating was voluntary.

C. ChatGPT

Since the behavior of ChatGPT is changing over time, even for the same model [49], here, we provide context regarding what model was used during the study, as well as when it was used. Data collection took place between March 5 and March 28, 2024. During each lab session, students had access to freely accessible version of ChatGPT, GPT-3.5. Further, between each lab session, the entire chat history of the account was downloaded and then cleared from the account, meaning each lab group faced a seemingly empty account with no access to previous chats or any initial prompting done by the researchers. The ChatGPT account used by the students was created by the researchers.

We opted for the free version of ChatGPT after carefully considering its ability to assist with the analysis during the lab and the types of answers it could provide for the optional discussion questions toward the end of the lab. In our testing of GPT-3.5 in February 2024, we identified that the current version of ChatGPT could be a sufficient help in that it was able to provide answers, albeit of varying quality depending on how the prompts were constructed, to important questions related to the analysis students were tasked with conducting during the lab. Another argument for using the free version of ChatGPT during the study was that it would more closely perform on the level students who had prior experience with ChatGPT or similar GenAI

tools thought it would. Thus, their use of ChatGPT during the study would be similar to how they would use it in a more naturalistic setting.

D. Data collection

To get relevant information about how the students worked during the lab, each group was filmed during the entire lab session, including the interview before and after the lab. Additionally, the entire chat history with ChatGPT was collected, as well as any written notes the groups created during the lab session.

During the interview before the lab session, students were asked whether they had used ChatGPT or similar services before for school-related tasks, with a follow-up on use in physics specifically. If they had such experiences, then follow-up questions asked for examples and what version of ChatGPT or other tool they had used. Then, to get a sense of the students' prior knowledge and how they approach solving problems during labs, we also asked them what they think acoustic levitation is and how they would handle a situation where they run into a problem, practical or theoretical, during a normal lab session.

During the lab, the students were told that the researcher present in the room during the lab session would not interfere or answer any questions unless something unexpected happened with the equipment. In addition to the material needed to conduct the experiment during the lab session and what material students decided to bring to the session, a computer was also supplied by the researcher with ChatGPT. A typical view of their workspace can be seen in Fig. 2. The researcher would also intervene if students' handling of the equipment could have led to breakage or posed any danger to the students. Interference from the researcher was needed in groups 1, 2, 5, and 6 for about 1 min within the first 5 min of the lab session. Groups 1, 2, and 5 had trouble understanding how to use the caliper to vary the distance between the speakers, and it was deemed necessary for the researcher to explain this, as they were at risk of breaking the glue holding the caliper and

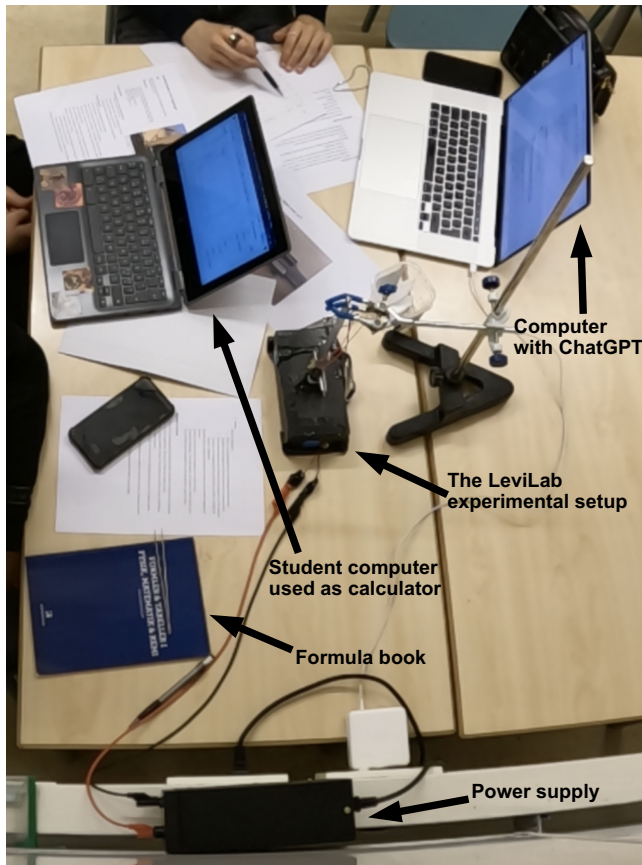


FIG. 2. What the students' laboratory workspace typically looked like, and what materials they chose to use.

speakers in place. Group 6 experienced glitches with the cables between the setup and the power supply, and the researcher had to step in to provide new cables and perform a quick check to see if any of the electronics had been damaged. Finally, since group 7 consisted of only one student, the researcher informed the student that they might ask clarifying questions about what they were doing during the lab session if it was deemed difficult to interpret what the student was working on only from the recorded video. The researcher asked questions regarding what the student did after 16, 40, and 50 min, respectively.

After the lab, another group interview was conducted in which the students were asked to first think aloud regarding how they thought about using or not using ChatGPT during the lab. That question was then followed up by asking if they discovered any problems in the answers they got from ChatGPT during the lab. Then, the students were asked to reason about whether they thought tools such as ChatGPT could be used to get help with explanations in physics, as well as whether they could use such tools to get help with various problems during a lab session. Finally, the groups were asked how they think AI tools could be used during lab sessions in physics.

E. Analysis and limitations

The research questions for this study are of an exploratory nature, with the aim of getting information about students' views of the potential role of GenAI tools in the physics laboratory, as well as furthering our understanding of the potential use cases of such tools in educational lab sessions. By collecting qualitative data and identifying that the three parts of the study, the pre- and postinterviews and the lab session, provide a narrative account of students' experiences as a group, the analysis focused on extracting these by drawing upon an inductive narrative approach for the analysis.

Following the typical structure of working with qualitative data, including conducting a narrative analysis [50], the analysis includes multiple steps. First, to get familiar with the data, all of the video recordings were observed first as a whole, followed by a second time where the data were carefully transcribed. The transcription process of the video recordings was aided by first having a local version of Whisper [51] transcribe the data. Then, the first author certified that the transcription provided by Whisper was correct, adding or changing wording as needed, anonymizing the transcript, and formatting the transcription so it was possible to identify the different speakers. After the transcription was complete, the first step in becoming familiar with the data also involved identifying what the groups did and when during the lab session. This was done by constructing Gantt charts (Fig. 3).

The second step of the analytical process involved inductively identifying relevant themes and passages that were in line with the general research aim of investigating the use of GenAI tools. Here, it was relevant to make use of the Gantt charts, transcripts, and each group's chat history with ChatGPT to conclude first what they had done when, how they discussed interactions, or lack thereof, of ChatGPT, as well as connect why they answered the way they did during the postinterview in light of their reasoning during the preinterview and interactions with ChatGPT during the lab. Following the second step, parts 3 and 4 of the analysis involved identifying the overall narrative tone and relevant themes and data to include in the findings sections. Here, tentative findings were discussed both among all authors of the papers as well as presented and discussed with colleagues external to the current study. Then, the main author proposed an initial structure of the findings and the main conclusions to be presented, which were again discussed among all authors to certify that the identified conclusions are supported by the data. Practically, this was achieved by proposing themes and supporting transcripts, which were then discussed in terms of their reliability and potential problematic subjective interpretations. After reaching a consensus among the authors on the main emerging themes identified in the data, the final analytical step involves reporting and

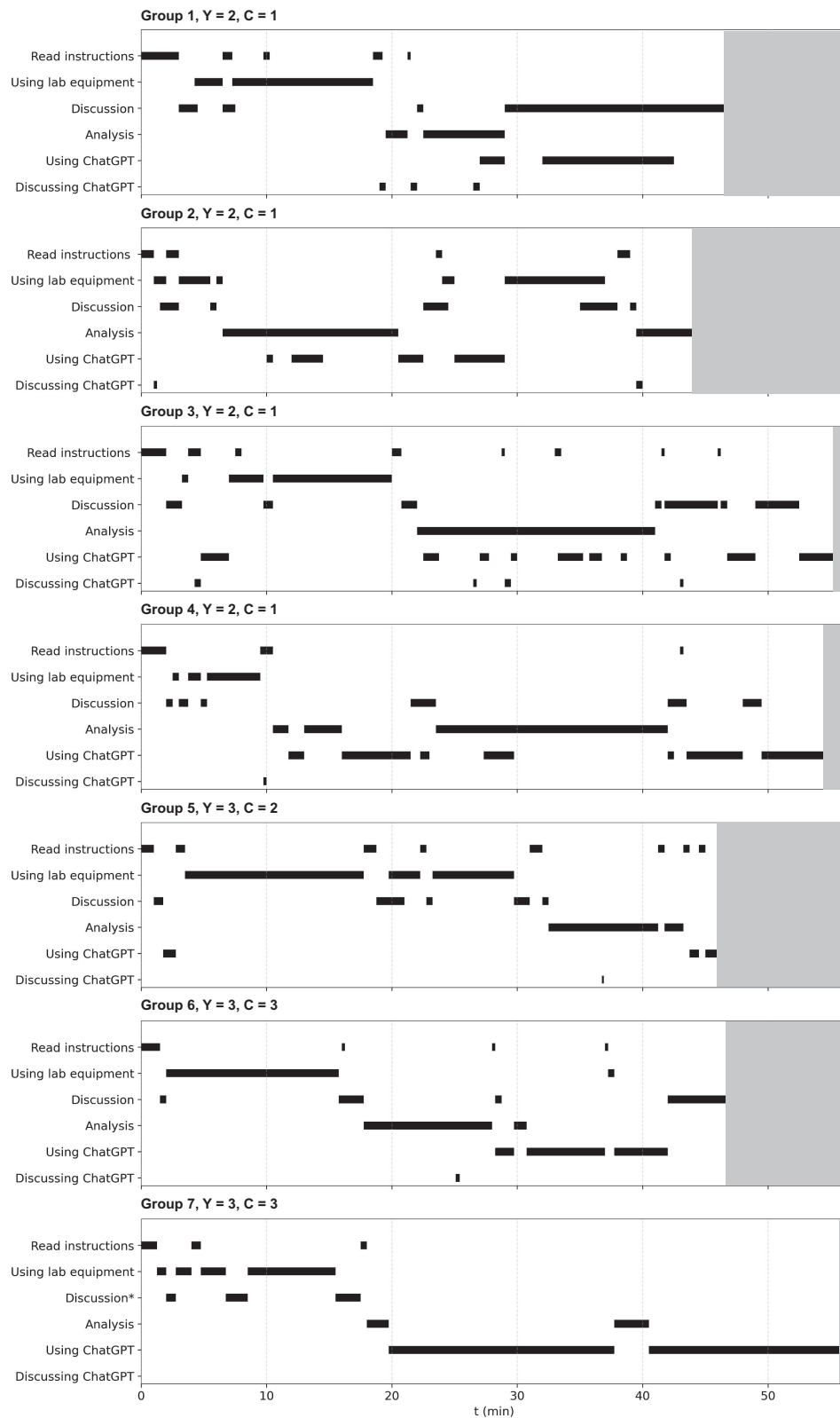


FIG. 3. Visualizing the workflow for each group, including information about their year (Y) and course (C), during the lab using Gantt charts and the categories defined in Table II. Since group 7 had only one student, the criteria for Discussion* was extended here also to include times when the student's actions did not fit in the other categories. An example is when the student conducted more general note taking during the lab. Gray bars are added to indicate each group's end time, as specified in Table I. The “*” indicates a special case regarding how the criteria were applied in the Gantt chart for group 7.

discussing the findings, as presented in Sec. V through Sec. VII of this paper.

In discussing the findings, the theoretical framework of variation theory was used in further analysis, related to the students' awareness. Specifically, we relied on the structure of awareness as a way of understanding, from the students' point of view, the enacted object of learning. The setup of the study, where no teacher was present to enact the intended object of learning, made it relevant to conceptualize the enacted object of learning. In our analysis, we treated the enacted object of learning as something dynamic, akin to the dynamical nature of the structure of awareness. As such, we could identify differences in what the enacted object of learning was between the groups, which aided our understanding of why the groups had seemingly different foci during the lab. Additionally, variation theory was used as a lens in the discussion for identifying implications for teaching and learning.

Collecting and reporting qualitative data, as is done in the current study, is an important step on the path of creating a collective understanding of the phenomenon under study [52]. As with all qualitative research, however, it is important to recognize how the researcher influences the analysis, which brings to light the issue of validity [50]. Another issue that relates to the overall validity of the current study is the lack of representability and thus generalizability. It is, therefore, important to recognize that the current study only included students enrolled in the Swedish school system. As such, any conclusions drawn from this study have to be viewed in the light of the current overall cultural context of that system. For example, it is important to recognize that soon after the public release of ChatGPT, the Swedish National Agency for Education urged schools to be cautious in letting students use such tools and to limit home assignments in general, as it would be nearly impossible to verify whether the student or some GenAI tool had completed the submitted work. The level of knowledge the involved researchers had about the current local educational landscape, as well as the societal debate, was therefore deemed relevant to be cognizant of when conducting the current study.

Finally, it is important to again stress what conclusions and insights can be gained from the current study. It is not to provide findings to produce an objective truth or provide absolute certainty but rather to identify and argue for justifiable conclusions that produce likely interpretations that have some utility for the intended reader [50].

V. FINDINGS

Here, we present findings from the group interviews before and after the lab and what the students did during the lab, providing relevant data for answering the research questions. The current study is a case study involving a small number of participating students ($n = 19$), and as

such, the findings presented here should not be seen as directly generalizable. Further, in this section, we include prompts by the groups but give a more descriptive take on what ChatGPT answers. We opted for this approach since all groups, except group 6, interacted with ChatGPT in Swedish, and the act of translating the answers could have a profound impact on how the answers are interpreted by a reader. The entire chat history for each group is provided as Supplemental Material [53]. As such, the interested reader may translate the chat history to get a more complete context of each group's chat history.

A. Group interviews

In analyzing the interviews before and after the lab, we primarily sought to answer RQ1 and provide a potential additional perspective on some parts of the lab, namely the students' self-reflection of what they had done and why. This additional information could be helpful to discuss RQ2 and RQ3.

Regarding which interactions with ChatGPT students perceive as useful in the context of a lab session, participants were primarily negative about the general usefulness of ChatGPT in physics and mathematics before the lab. All groups did mention that AI chatbots like ChatGPT were not good at math but could help provide comprehensible explanations of "words" (concepts) they had forgotten or felt the book or teacher had explained poorly. However, groups 4, 6, and 7 all stressed the importance of being cautious regarding the facts provided by ChatGPT and consistently trying to double-check with other reputable sources. A few students (Benjamin and Beatrice in group 2, Caroline in group 3, and Daniel in group 4) had no prior experience using ChatGPT in physics prior to the current study. Of those students, Beatrice claimed she had not used AI chatbots for any school work previously, whereas Benjamin, Caroline, and Daniel mentioned using them for help in other subjects that were not science or math.

After the lab, all groups except group 5 were still skeptical of its ability to perform calculations or do any meaningful analysis. Apart from group 5, all students mentioned that it might not be a good idea to rely on ChatGPT for calculations. Groups 4 and 6 said ChatGPT was helpful during the analysis since it could provide relevant equations and that they could trust them by either studying the equations and see if they looked correct or by performing calculations and verifying if the answer was in line with their hypothesis. However, group 6 specifically mentioned that this requires that you know something about the underlying physics. Further, group 3 said they got an equation from ChatGPT to use during the analysis, but that they could not know if it was correct since ChatGPT did not provide any source. Within group 5, the reasoning was that ChatGPT was useful when one has to do "simple analysis that only takes time to write down and calculate.

For that, ChatGPT is useful and good to use ... it gives a good estimate based on previous research and is therefore a good base answer.” (Ebba, group 5).

The students who had no previous experience with using AI chatbots in physics, Benjamin, Beatrice, Caroline, and Daniel, stated after the lab that they would use such tools in the future when they are looking to get short and simple explanations of things they struggle to understand. Furthermore, most groups stated that their answers from ChatGPT during the lab seemed reliable and felt no need to fact-check theoretical answers. Groups 4, 6, and 7 had a similar stance on trustworthiness before and after the lab regarding theoretical information. However, all mentioned that they thought it was reliable in providing physics formulas because they did not see any reason for fake information about those in the training data. Fabian in group 6 explicated this as “We got two different formulas for the speed of sound. The first gave like 2000 m per second and the other 344 comma something, so it felt more reasonable. Often there is some truth in all it says because it always comes from some context, but it might be a context we don’t know about. For example the first formulae I still think is right because it makes no sense to keep fake physics in the training [data], but it might be that it gave wrong units or something.”

When asked about how they thought AI tools like ChatGPT could be utilized in physics labs in the future, all groups, except group 5, said it could be a good substitute for the teacher in answering conceptual questions. Group 5 thought it could also be good to help with calculations or get answers to some practical questions if they did not have enough time to finish the lab during class properly. Gabriella (group 7) explicitly mentioned the problem of using AI tools to cheat but stated that “ChatGPT can help you formulate explanations you are not fully done with. It is less obvious cheating in labs because it can help you with explanations or clarifications, but never do the work for you.” Additionally, groups 4, 6, and 7 were all of the mindset that AI chatbots could be a valuable tool during labs in the future because they can help students “get answers to small stuff like helping you remember some word or technique and help the teacher by outsourcing simple or ‘stupid’ questions. But it’s also important that

students get better knowledge about how AI tools work and how to ask good questions.” (Fanny, group 6). In relation to “stupid” questions, no group clearly defined what they meant by similar wording, but most students mentioned that it could be useful to ask ChatGPT about the meaning of nonphysics words that appeared on lab instructions that they had forgotten or to ask about concepts they should know about but had forgotten since they had not heard the word for some time.

B. Interactions with ChatGPT during the lab

In this paper, we define one interaction with ChatGPT as consisting of one written prompt by the user and the reply from ChatGPT. During the lab, the number of interactions the groups had with ChatGPT varied from 2 to 17. Figure 3 illustrates the manner in which each laboratory group engaged with the assigned tasks, categorized according to the following parameters: Read instructions, Using lab equipment, discussion, analysis, using ChatGPT, and discussing ChatGPT. The meaning of the categories is elucidated in Table II. For each lab group, Fig. 3 indicates how they worked during the lab using the categories: read instructions, using lab equipment, discussion, analysis, using ChatGPT, and discussing ChatGPT (see Table II).

Two groups, 2 and 5, tried to get some guidance from ChatGPT regarding data collection and equipment handling. Group 5 tried to get practical guidance on how to do something with the lab equipment using the prompt “How do you adjust sound frequency on a Levilab.” This was asked after having read the instructions once and interpreting the instructions of varying the distance between the speakers to be synonymous with adjusting the sound frequency. The response was seemingly irrelevant and mentioned a product called Levilab. However, it differed from the equipment they had access to, and the students reread the instructions without interacting with ChatGPT until the end of the lab session. Group 2, who struggled to identify what data to collect during the lab and, as such, had to collect data twice, did consult ChatGPT on “how DO YOU PERFORM LABORATION of acoustic wavelength and sound speed” about 20 min into the session. For them,

TABLE II. Categories and corresponding criteria used to identify the overall workflow for each group during the lab.

Category	Criteria
Read instructions	If students seem to be actively engaged in reading the lab manual
Using lab equipment	If students seem to be actively engaged in reading the lab manual
Discussion	If students talk among each other about the lab such as what to do, how to interpret results, sorting out questions, or talk about something that does not fit in the other categories
Analysis	If students are performing calculations or are otherwise involved in tasks related to data analysis, such as generating relevant graphs, as well as discussing or looking up formulas to use
Using ChatGPT	If students write to ChatGPT, read responses, or directly discuss responses or the interaction as a whole
Discussing ChatGPT	If students discuss something where ChatGPT is mentioned (or if any abbreviations such as “Chat,” “the AI,” or similar wording are used), without interacting with it

even though the reply from ChatGPT did not describe how to conduct measurements with the available equipment, the answer led them to understand that the concepts of sound waves, speed, and frequency were relevant to the experiment.

Groups 1, 3, 4, 6, and 7 had no interactions with ChatGPT connected to the data collection. During the data collection process, Carl, in group 3, who was not involved in handling the equipment or taking notes, was tasked with asking ChatGPT to guide the group on what acoustic levitation does. He did this first by asking “What is acoustic levitation?” with a follow-up to get more relevant information that was “Explain to a high school student.” Overall, the group was content with the second answer. There, ChatGPT used a metaphorical invisible basket and explained that it was created by special patterns of sound waves that made several pressure points that could be used to hold an object in place. ChatGPT additionally made an analogy to trying to levitate balls in the air by blowing on them from below. By accepting this explanation, the experimental setup the students worked with does not make much sense as it uses two opposing speakers to make a particle levitate, and the fact that the particle only levitates for some distances between the speakers [54]. The answer ChatGPT gave to the first question Carl posed only mentioned that acoustic levitation uses sound waves to lift and manipulate objects and that sound waves can create an acoustic field that generates pressure points in a volume of air or other medium. Thus, it was similarly vague regarding what happens but omitted the use of analogies.

Group 6 also attempted to get an introduction from ChatGPT on acoustic levitation. However, the group was more familiar with mechanical waves, including relevant concepts, such as standing waves, and utilized prior knowledge in their interactions with ChatGPT. Additionally, they chose to interact with ChatGPT using English, as they claimed it had a lot more English than Swedish in its training set. From the prompt “could you give us a introduction to acoustic levitation?” they got a reply that involved relevant concepts such as standing waves, nodes and antinodes, and acoustic radiation pressure. However, the claim that objects can be levitated in both nodes and antinodes did not agree with the group’s hypothesized explanation of why the particle did not levitate at all speaker separation distances. As such, they continued to probe if ChatGPT gave what they could gauge as a legitimate explanation of this claim by asking “Where in the standing wave can you balance an object and why?” The new response from ChatGPT made them more confident in their hypothesis that the particle could only levitate in the displacement nodes since it once again mentioned levitation to be possible in both nodes and antinodes but also stated that levitation in the antinodes would be unstable but would still be attracted there due to a “stronger radiation pressure.” Group 6 discussed that this explanation was confusing and possibly based on flawed reasoning, although they were not sure because they felt they lacked the proper knowledge about the physics involved.

In general, groups 1, 2, 3, 5, and 7 were identified as struggling during the lab, meaning that they had trouble understanding what data to collect, got stuck during the analysis, or both. For groups 1, 2, 3, and 5, interactions with ChatGPT focused on asking simple questions and they did little to no follow-up. Group 7 struggled mainly due to a lack of knowledge on how to use the online calculator Desmos. They opted for using Desmos since they did not bring a calculator to the lab session, and thus, much of the interaction with ChatGPT focused on trying to get help with understanding how to use Desmos to perform the necessary analysis.

Groups 4 and 6 both did well during the lab and showed a deeper understanding of how to interact with ChatGPT during the lab. Questions they posed also included some poorly formulated ones such as “which SI unit is wave-length written with?” However, a meaningful difference was the amount of follow-up and additional probing included in the interactions. Group 6 notably chose to double-check the answers given by ChatGPT by opening up additional chats and asking similar questions again but without having the context of the entire chat history. Both groups also tried to gain a deep understanding of how acoustic levitation works by posing different questions to ChatGPT, discussing the answers among themselves, and posing new questions to see if their understanding aligned with the new answers provided by ChatGPT.

C. On interacting with ChatGPT and students’ awareness

Drawing upon our theoretical framework, we can identify instances where the students’ interactions with ChatGPT during the lab impacted their structure of awareness, which in turn has an effect on what the object of learning is (and thus what aspects are relevant to consider). As an initial step, it must be acknowledged what students perceived to be the primary focus at the beginning of the lab session. An aim for the lab session was provided to students by including a short sentence in the lab manual (see the Appendix) stating that they should explore acoustic levitation and determine the speed of sound in air. During the data collection phase of the lab, both groups 3 and 6 verbally reiterated the written aim. This led to two different interactions with the goal of better understanding acoustic levitation, where group 6 discussed among themselves and reasoned about the phenomenon of acoustic levitation based on their prior physics knowledge. Group 3 instead approached the aim of exploring acoustic levitation by having Carl interact with ChatGPT, as described above. For the other groups, no discussions touching on the aim of exploring acoustic levitation were present until either during the data analysis phase of the lab or after having finished the analysis. Among these, group 4 stood out by having conceptual discussions on why the particle could levitate during the data analysis stage. Overall, most

students perceived the aim of the lab session to be related to determining the speed of sound, and thus, the enacted object of learning was identified to be more closely connected to the capability of determining the speed of sound using the available equipment. In light of the identified enacted object of learning, it was relevant for the groups to spend most of their time on data collection and analysis, as illustrated by Fig. 3.

From the previous subsection presenting how groups interacted with ChatGPT, it became evident that the interactions were akin to having quick access to the teacher, a teaching assistant, or web searching capabilities. In these interactions, we can identify passages where students' awareness of what could be important to discern during the lab was seemingly affected. Here, we present three examples, from groups 2, 4, and 5, highlighting how the interactions affected the students' structure of awareness during the episodes.

The first episode is of group 2 and an initial interaction with ChatGPT during the data analysis phase. They asked ChatGPT "How to convert kelvin to celsius?" with the follow-up "how many degree celsius is 273 kelvin?" In the first response, ChatGPT provided the formula $T_C = T_K - 273.15$ and also included an example where $T_K = 300$ K. In the second response, ChatGPT said that with $T_K = 273$ K, $T_C = -0.15^\circ\text{C}$. After the replies, the group went on to have a discussion without ChatGPT about whether they needed the temperature data or not for the experiment. Here, the group still had no equation or conception of a model that included the temperature dependency for the speed of sound, making the interaction episode rather irrelevant. The students' focus was locked on the aforementioned enacted object of learning, and they had yet to succeed with the first part of the analysis, which was to calculate the wavelength. However, with no apparent understanding of how the speed of sound, or how the concept of wavelengths were connected to the experimental setup, group 2 would have needed clearer guidance toward aspects relevant to the object of learning. Instead, here the group's interaction with ChatGPT strengthened their focus on temperature, and they kept on haphazardly trying various calculations to finish the lab.

In episode 2, group 5 also struggled with the analysis and getting an estimate of the speed of sound using the collected data. Toward the end of the lab session, they turned to ChatGPT and asked, "What is speed of sound in 22 degrees celsius?" The reply told them that the approximate speed of sound at 22°C is 344 m/s, and the group was then happy to not continue with their own calculations but simply agreed that the answer given by ChatGPT was a good estimation of what they would have ended up with. During the group's short discussion following this interaction, they said that the reply was reasonable since ChatGPT provides "a good average of previous research results" (Eddie, group 5), and the answer could thus be

understood as ChatGPT had done the analysis for them. Here, we could identify a similar enacted object of learning, albeit more closely connected to simply being able to provide an acceptable value for the speed of sound in air. Similar to group 2, group 5 showed no prior understanding of the speed of sound or the connection to the concept of wavelength. Drawing upon the identified enacted object of learning, together with the fact that group 5 saw tools like ChatGPT as being able to provide a synthesis of all available information, their way of interacting with ChatGPT was perfectly reasonable. Again, here, we see yet another episode where the students' structure of awareness, both regarding the object of learning and GenAI tools, is not affected by spontaneous interactions with ChatGPT.

Episode 3 consists of an interaction group 4 had with ChatGPT toward the end of the lab session. There, they discussed how particles could be levitated using the experimental setup together with ChatGPT using the prompts "how can sound waves counteract gravity?," "what do you mean by the gravitational force is balanced?," "simplify how sound waves counteract the gravitational force?," and "is it particles that have bounced on each other and in that way creates a force upward that takes out the gravitational force?" In between the prompts, group 4 tried to identify if the answers they got agreed with their prior knowledge and results from the experiment. For group 4, we identified a clear contrast to the previous two episodes, namely that the enacted object of learning is more closely connected to the full stated aim in the lab manual. This is evident in part due to the fact that the group thought it important not only to calculate the speed of sound but also to understand its connections to the phenomenon of acoustic levitation. Another contrast is seen in the way group 4 interacted with ChatGPT. Here, the group's discussions prior to and between the first two interactions indicate that they initially discerned a connection between the concept of wavelength and the levitating styrofoam particles. However, the students perceive that ChatGPT provides confounding information and introduces concepts and ideas they are currently unfamiliar with. Additionally, their lack of adequate knowledge about the underlying physics makes it hard for them to gauge the correctness of the answers ChatGPT provides. Thus, the interaction led to a shift of awareness regarding critical aspects of acoustic levitation with the equipment; from having to do with sound waves from the two speakers sometimes "*strengthening*" each other (I.e., on their way to discovering standing waves) and toward one student's original notion that the styrofoam levitates mainly due to "air particles" physically pushing on it from below.

VI. DISCUSSION

The findings in the current study are based on seven lab groups and the work and reflection of, in total, 19 students. As such, they should not be interpreted as directly

generalizable. However, this study provides important insights into the potential use, current and future, of GenAI tools in high school science laboratories. Here, we discuss the findings in relation to the research questions, as well as broader implications of using GenAI tools in laboratory work. The discussion is distilled into four parts: The role of prior knowledge and trustworthiness, utilizing ChatGPT as a lab partner or a source of information, GenAI as a tool to develop problem-solving skills in laboratory settings, and the ongoing problem with GenAI and hallucination.

A. The role of prior knowledge and trustworthiness

Regarding RQ1: What do students perceive as useful interactions with GenAI tools like ChatGPT during a lab session? Our findings show that, in agreement with previous research [6], students require knowledge about the underlying physics in order for them to gauge whether answers provided by ChatGPT are trustworthy. This is exemplified well by group 6, one of two groups of third-year students taking the optional course Physics 3, where they could get meaningful assistance from ChatGPT during the analysis stage of the lab in identifying a formula to calculate the speed of sound that included temperature dependence. If not for the group's prior knowledge of the physics of sound, they might not have been able to conclude that the initial equation provided by ChatGPT was faulty in that the form of the equation was correct, but the gas constant R was given with the wrong units and would have had to be converted prior to using it to calculate the speed of sound.

The importance of prior knowledge about the specific area to successfully identify the validity of answers given by ChatGPT is further strengthened by examining an interaction during the lab by group 4. Although they were one of the groups with second-year students taking Physics 1 and thus having access to very little prior knowledge of the physics of sound, they were identified as performing well during the lab since they were able to identify what data they should collect and perform the analysis without issues. However, they ended up struggling toward the end of the lab session when they were tasked with discussing how acoustic levitation works. Due to their limited to nonexistent knowledge of wave mechanics and the physics of sound, they had few tools available to them to challenge what answers ChatGPT gave them. As a result, despite group 4 demonstrating an understanding of the need to pose not just one question but many to determine whether the answers provided by ChatGPT made sense or not, they were ultimately led to a wrong explanation that it is mainly the lower speaker that supplies a force by pushing on the styrofoam particle and thus counteracting gravity.

Finally, in most other interactions with ChatGPT during the lab session, the groups seemed to take the answers provided by ChatGPT as simply being the truth. This may

be connected to how the students, at times, anthropomorphized ChatGPT during the interviews. Group 3 stated that they might not fully trust the answers if ChatGPT did not provide any sources or citations, and group 5 stated that ChatGPT can be trusted since it has knowledge about previous research results. In general, during the group interviews before and after the lab, when asked about the capabilities of ChatGPT, students were quick to use words such as knowledge and talk less about it as some tool and more as a person with much knowledge. During the interviews, it was mainly groups 6 and 7 that stood by the argumentation that what ChatGPT can answer and not reliably is dictated by what data were part of its training, which indicates that they are more cognizant that GenAI tools do not have knowledge and reasoning skills similar to humans.

B. Utilizing ChatGPT as a lab partner or a source of information

Concerning RQ2 and RQ3, they focus more on how students interacted with ChatGPT during the lab to overcome specific hurdles. Previous studies provide little information about its capabilities to function as a lab partner. What it does show, however, is that depending on the type of questions and problems students face, it might be able to provide reliable answers, as it can be good at solving subject-specific problems [8,9]. Further, it has also been identified that GenAI tools, like ChatGPT, can reproduce language and provide answers relevant to tackling problems in laboratory settings [55]. Our findings indicate similar conclusions. In RQ2, How do high school students actively engage with ChatGPT to solve problems during a lab investigating the speed of sound using acoustic levitation?, Fig. 3 shows that students tend to make use of ChatGPT after having done all or most of the data collection. This indicates that despite having little to no knowledge about the phenomenon under study, the groups had little problem collecting data using the experimental setup.

During the group interviews, students stressed that a major use case of ChatGPT is to get answers to conceptual questions or to clarify problems or instructions, and that it made little or no sense in asking about lab equipment. This sentiment is in agreement with how the groups did make use of ChatGPT during the lab session: during the analysis and to get help with the discussion questions about acoustic levitation. Here, the primary beneficial use case for the groups was related to getting help with identifying what equations to make use of, as well as getting an introduction to acoustic levitation. A noteworthy special case was group 5, who “outsourced” the analysis to ChatGPT after failing to understand what to do with the data by simply asking what the sound speed was at the room's current temperature. During the analysis, it is important to stress that the lab instructions were intentionally vague in that no equations were provided. Instead, the assumption was that

students would utilize a combination of ChatGPT and their formula book to identify how to make use of the data to get the speed of sound using the equation $v = f\lambda$ and compare that to the speed of sound they get when using an equation that included temperature dependence. Additionally, if the groups had time, they were tasked with discussing questions related to acoustic levitation, including the underlying principles of acoustic levitation. As such, it was expected that all groups, regardless of their natural intention to use or not use ChatGPT during the study, were incentivized to at least try to use it during the analysis and discussion part.

C. GenAI as a tool to develop problem-solving skills in laboratory settings

Overall, the findings presented in the current study demonstrate fruitful and problematic interactions with ChatGPT in terms of aiding students in solving problems or helping them understand the phenomenon of acoustic levitation. Groups 4 and 6 were both able to use ChatGPT to aid them during the analysis in that group 4 successfully utilized it to aid them in reasoning about whether the wavelength they got from the data collection was reasonable. Further, both groups showed proficiency in discussing how to identify if the equations given to them from ChatGPT were reliable by drawing upon their prior knowledge of dimensional analysis. Contrary to these, groups 1, 2, 3, and 5 all demonstrated that lacking knowledge about the relevant physics and how GenAI tools like ChatGPT “reasons” or has “knowledge,” having access to such tools in a classroom setting, including laboratories, might end up disadvantageous for the students’ learning process.

In this contrast between groups 4 and 6, and groups 1, 2, 3, and 5, we can see a situation where ChatGPT can serve as a tool to help reduce the importance of prior content knowledge during lab sessions, which teachers can utilize to help their students develop certain capabilities. Namely, groups 4 and 6, despite significant differences regarding relevant knowledge of the underlying physics, were both able to use ChatGPT to assist them in moving forward with the data analysis. What instead separated these groups from 1, 2, 3, and 5 was their assumptions and knowledge about how GenAI tools function and “reasons.” The kind of interactions groups 4 and 6 demonstrated during the analysis provided them with new tools on how to approach problem solving in a laboratory context. Drawing upon variation theory, we can identify an alternative approach to the lab session that could allow the struggling groups to succeed more similarly to groups 4 and 6. If an object of learning is chosen related to utilizing GenAI tools in problem solving during laboratory work, a potential critical aspect may be related to discerning how current GenAI tools’ reasoning differs from that of humans. Here, the argument for the critical aspect is connected to the identified stark difference between groups 4, 6, and 7 and the other groups when it came to assumptions about

how GenAI tools work. Such a focus would also open up for more active scaffolding regarding steering the students’ focus toward critical aspects [36]. This would include rather heavily aiding students with content-specific conceptual understanding, leaving them to instead focus on problem solving and GenAI-related aspects. Finally, this shift in focus toward more general skills has been identified in previous work in PER using variation theory, as exemplified by Ref. [42] on how students develop skills related to developing internal idealized models and use them to develop conceptual understanding in the physics laboratory (referred to as being able to modelize). It is also a focus that has widespread acceptance in the current PER literature as a means toward getting a more expertlike view of laboratory work in physics [56–59].

Continuing, we can connect the suggested change in focus regarding the object of learning to RQ3, How can interacting with ChatGPT during the lab aid students in solving problems and with conceptual understanding? With the aim of having students develop new skills using GenAI when solving problems during laboratory work, a crucial part of their scaffolding during the lab would revolve around probing the students on their conceptual understanding of the phenomenon studied during the lab by actively interacting with the groups during the lab. We base this suggestion on that our findings show instances of ChatGPT providing information to the students about the underlying physics, which is wrong or misleading. An example of such an instance is seen in the interaction Carl in group 3 has with ChatGPT. There, the analogy of keeping a ball in the air by blowing on it from below reinforced the notion that the main contributor to the phenomenon seen during the lab was due to the bottom speaker pushing air particles on the styrofoam particle. If a teacher were present and identified Carl’s acceptance of this analogy, they could intervene and provide an explanation to correct the problems with the analogy. Thus, Carl and the rest of group 3 would be back toward the intended object of learning instead of having their awareness shifted toward aspects related to a more concept-focused object of learning.

D. The ongoing problem with GenAI and hallucinations

Becoming better at addressing the issue of getting problematic and made-up answers from GenAI tools in the classroom relates to the growing need for AI literacy in education, for students and teachers alike [21,26,60]. As is evident from our findings, and in agreement with previous work [61], students with a combination of less knowledge of the subject (here, physics) and not being familiar with GenAI tools or have little to no knowledge about how they function struggle with making use of these new tools. Proposed solutions from the literature involve training students and teachers in prompt engineering [21,60,62]. This proposal is further supported by our findings, where groups 4 and 6 utilized their understanding and assumptions about how

ChatGPT functions to help them distinguish between reliable and unreliable information in the answers provided by the chatbot. From their interviews and observations of their lab sessions, it became evident that their assumptions led them to ascribe a less humanlike awareness in how ChatGPT generated responses.

Developing skills related to prompt engineering can be a fruitful way to give students and teachers a greater understanding of not only how to extract more useful information from GenAI tools but also to better comprehend changes in limitations over time between and within models [21,60]. For educators, an important step in learning how the behavior of GenAI tools varies over time is for researchers to continuously monitor how various models perform, with and without proper prompt engineering and further fine-tuning. To that end, the current study has explored how largely untrained students can and do make use of a free version of ChatGPT, GPT-3.5 at the time of data collection, when solving problems during a physics lab. The findings and discussions presented here can benefit educators and researchers in gaining a deeper understanding of the naïve and partially spontaneous interaction between students and GenAI, particularly in identifying fruitful ways of interaction and problematic instances of overreliance on ChatGPT as a substitute for or absolute source of information.

The use of GPT-3.5, rather than, for example, GPT-4o, has profound implications in terms of the type of misinformation presented by ChatGPT, as well as how it responds to questions in this study. Additionally, it is reasonable to assume that the way ChatGPT behaves has an effect on how participants in this study rely on the answers and find ChatGPT helpful. However, subsequent releases of GenAI tools still struggle with the problem of hallucinations, albeit less frequently, and it is a problem that may be inherent to the way large language models work [63]. This implies that although the specific problems ChatGPT exhibits in this study may not be present if replicating the study today, we would expect similar conclusions, but perhaps that more groups may have been able to successfully use ChatGPT, particularly in relation to understanding when and why GenAI tools can be useful, depending on what is to be learned.

VII. CONCLUSION AND OUTLOOK

GenAI has made a clear impact on the public and educational discourse about the future of work and education. GenAI and AI more generally are likely to become an even larger part of people's daily lives in the future, both at an individual and professional level. Therefore, it is essential that, as far as possible, both teachers and students are instructed in how to interact with GenAI tools to extract the greatest benefits from them. This has happened with other technologies in the past, including calculators, digital circuits, computers, and the Internet. However, GenAI tools are different in that they respond to the user in

fundamentally new ways, thus requiring new kinds of training related to formulating questions and gauging the reliability of results and answers from such tools.

In the current study, we address the research gap related to the need of exploring how GenAI tools, specifically the free version ChatGPT at the time of the data collection, could be used in the educational physics laboratory. Our findings are in agreement with previous research regarding the importance of educating students about the capabilities and limitations of AI tools, as well as gaining some level of understanding about how they are developed and “think.” Additionally, in our analysis, we have explored a novel way of conceptualizing the enacted object of learning from a student perspective and its connection to the structure of awareness. This allowed us to gain further insight into how GenAI tools can be utilized in laboratory settings to develop students' problem-solving skills.

In science, technology, engineering, and mathematics (STEM) education specifically, there is a need to further explore and develop educational strategies and materials that can help students and teachers better understand how to best utilize and think about the role of GenAI in education, not least in laboratory settings. As these tools continue to be developed, their ability to simulate a trustworthy chain of reasoning will improve, furthering the need to carefully consider the implications such interactions may have for learning concepts in disciplinary-accepted ways. Here, the role and knowledge of the teacher about AI is of utmost importance. Finally, we have identified potential for using GenAI tools like ChatGPT in the physics laboratory as a lab partner, but their usefulness appears to depend on students' knowledge of the phenomenon under study as well as their knowledge about how GenAI tools function. Furthermore, the specific ways in which such a tool is best employed in the physics laboratory require further research and will also be dictated by the future development of these tools.

ACKNOWLEDGMENTS

We would like to thank the students who participated in the study. We also thank our dear friend and colleague Dr. Ricardo Méndez-Fragoso for providing valuable insights in discussing and reviewing the manuscript.

S. K.-W.: conceptualization, data curation (lead), formal analysis, investigation (lead), methodology, project administration, resources (lab design; supporting), validation (lead), visualization (lead), writing—original draft, writing—review and editing (lead). A. J.: investigation (supporting), resources (lab design; lead), validation (supporting), visualization (supporting), writing—original draft (supporting), writing—review and editing (supporting). J. E.: data curation (supporting), investigation (supporting), supervision, validation (supporting), writing—review and editing (supporting).

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain sensitive personal information. The data are available from the authors upon reasonable request.

APPENDIX: LAB MANUAL

1. Speed of sound and wavelength

a. Aim

To explore acoustic levitation using *LeviLab* and determine the speed of sound in air.

b. Equipment

- *LeviLab* (two ultrasonic speakers mounted on a caliper, a function generator, an amplifier, and a thermometer.)
- Lab stand equipment.
- Voltage supply (DC, 20 V).
- Polystyrene particles.
- ChatGPT.

c. Experiment

1. Information, assembly, and operation of *LeviLab*
 - Place *LeviLab* on a flat surface.
 - Stabilize *LeviLab* using the lab stand equipment.
 - Vary the loudspeaker distance using the caliper.
 - The emitted sound frequency is 40 kHz.
2. Calculation of the speed of sound in air
 - Verify that the power supply is on.
 - Place a polystyrene particle on the lower speaker.
 - Gradually lower the upper loudspeaker and identify points where the particle appears to “jump up” and start levitating. Does increasing the distance between the speakers produce the same effect?
 - Place the loudspeakers at a distance from each other so that as the loudspeaker distance approaches zero, the particle has jumped 7–10 times. Here is a closer description of a measurement:

- Carefully lower the top loudspeaker. Stop exactly at the position where the polystyrene particle jumps up and starts to levitate.
- Use the caliper to record the distance between the loudspeakers.
- Record the temperature of the ambient air using the thermometer.

3. Analysis

- Calculate the wavelength.
 - Label your measurements such that 1 represents the shortest speaker spacing. This is called an ordinal number.
 - Conduct a linear regression analysis with loudspeaker spacing as the dependent variable and the ordinal number as the independent variable. The slope of the regression line corresponds to the wavelength.
- Calculate the speed of sound using a formula involving wavelength and frequency.
- Compare your result with the speed of sound as determined by the temperature reading. This is achieved by applying a linear model that expresses the relationship between the speed of sound and the temperature in degrees Celsius.

d. Discuss in the group

1. What are the fundamental principles of acoustic levitation that make it possible for sound waves to balance the gravitational force on an object? Construct a descriptive figure.
2. Discuss possible sources of error and how to minimize them to calculate more reliable speed of sound values.
3. Explain why it is important to measure the ambient temperature when calculating the speed of sound.
4. Discuss potential applications of acoustic levitation in situations outside school.
5. Explore and describe other applications where standing waves are used.

-
- [1] OpenAI, ChatGPT, <https://www.openai.com>.
- [2] European Commission: European Education and Culture Executive Agency, AI report: By the European digital education hub’s squad on artificial intelligence in education (2023), Publications Office of the European Union, 2023, <https://data.europa.eu/doi/10.2797/828281>.
- [3] Y. Shen and X. Zhang, The impact of artificial intelligence on employment: The role of virtual agglomeration, *Hum. Soc. Sci. Commun.* **11**, 122 (2024).

- [4] S. Wells, Ready or not, AI is coming to science education—and students have opinions, *Nature (London)* **628**, 459 (2024).
- [5] World Economic Forum, The future of learning: How AI is revolutionizing education 4.0, <https://www.weforum.org/stories/2024/04/future-learning-ai-revolutionizing-education-4-0/> (2024).
- [6] 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).
- [7] B. Gregorcic and A.-M. Pendrill, ChatGPT and the frustrated Socrates, *Phys. Educ.* **58**, 035021 (2023).
- [8] G. Kortemeyer, Could an artificial-intelligence agent pass an introductory physics course?, *Phys. Rev. Phys. Educ. Res.* **19**, 010132 (2023).
- [9] V. López-Simó and M. F. Rezende, Challenging ChatGPT with different types of physics education questions, *Phys. Teach.* **62**, 290 (2024).
- [10] G. Polverini and B. Gregorcic, Evaluating vision-capable chatbots in interpreting kinematics graphs: A comparative study of free and subscription-based models, *Front. Educ.* **9**, 1452414 (2024).
- [11] R. K. Fussell, M. Flynn, A. Damle, M. F. J. Fox, and N. G. Holmes, Comparing large language models for supervised analysis of students' lab notes, *Phys. Rev. Phys. Educ. Res.* **21**, 010128 (2025).
- [12] W. Yeadon, O.-O. Inyang, A. Mizouri, A. Peach, and C. P. Testrow, The death of the short-form physics essay in the coming AI revolution, *Phys. Educ.* **58**, 035027 (2023).
- [13] W. Yeadon and T. Hardy, The impact of AI in physics education: A comprehensive review from GCSE to university levels, *Phys. Educ.* **59**, 025010 (2024).
- [14] G. Polverini and B. Gregorcic, Performance of ChatGPT on the test of understanding graphs in kinematics, *Phys. Rev. Phys. Educ. Res.* **20**, 010109 (2024).
- [15] K. A. Pimblett and L. J. Morrell, Can ChatGPT pass a physics degree? Making a case for reformation of assessment of undergraduate degrees, *Eur. J. Phys.* **46**, 015702 (2025).
- [16] S. Küchemann, S. Steinert, N. Revenga, M. Schweinberger, Y. Dinc, K. E. Avila, and J. Kuhn, Can ChatGPT support prospective teachers in physics task development? Can ChatGPT support prospective teachers in physics task development?, *Phys. Rev. Phys. Educ. Res.* **19**, 020128 (2023).
- [17] D. Domenichini, A. Bucchiarone, F. Chiarello, G. Schiavo, and G. Fantoni, An AI-driven approach for enhancing engagement and conceptual understanding in physics education, in *Proceedings of the 2024 IEEE Global Engineering Education Conference (EDUCON)* (IEEE, New York, NY, 2024), pp. 1–3.
- [18] B. Gregorcic, G. Polverini, and A. Sarlah, ChatGPT as a tool for honing teachers' Socratic dialogue skills, *Phys. Educ.* **59**, 045005 (2024).
- [19] G. Kortemeyer, J. Nöhl, and D. Onishchuk, Grading assistance for a handwritten thermodynamics exam using artificial intelligence: An exploratory study, *Phys. Rev. Phys. Educ. Res.* **20**, 020144 (2024).
- [20] T. Wan and Z. Chen, Exploring generative AI assisted feedback writing for students' written responses to a physics conceptual question with prompt engineering and few-shot learning, *Phys. Rev. Phys. Educ. Res.* **20**, 010152 (2024).
- [21] L. Ding, T. Li, S. Jiang, and A. Gapud, Students' perceptions of using ChatGPT in a physics class as a virtual tutor, *Int. J. Educ. Technol. Higher Educ.* **20**, 63 (2023).
- [22] Y. Liang, D. Zou, H. Xie, and F. L. Wang, Exploring the potential of using ChatGPT in physics education, *Smart Learn. Environ.* **10**, 52 (2023).
- [23] R. Kahaleh and V. Lopez, Evaluating large language models in high school physics education: Addressing misconceptions and fostering conceptual understanding, *Phys. Educ.* **60**, 025013 (2025).
- [24] J. J. Trout and L. Winterbottom, Artificial intelligence and undergraduate physics education, *Phys. Educ.* **60**, 015024 (2025).
- [25] T. Adiguzel, M. H. Kaya, and F. K. Cansu, Revolutionizing education with AI: Exploring the transformative potential of ChatGPT, *Contemp. Educ. Technol.* **15**, ep429 (2023).
- [26] C. K. Lo, K. F. Hew, and M. S. yung Jong, The influence of ChatGPT on student engagement: A systematic review and future research agenda, *Comput. Educ.* **219**, 105100 (2024).
- [27] K. Hallal, R. Hamdan, and S. Tlais, Exploring the potential of AI-Chatbots in organic chemistry: An assessment of ChatGPT and Bard, *Comput. Educ. Artif. Intell.* **5**, 100170 (2023).
- [28] E. F. Ruff, J. L. Franz, and J. K. West, Using ChatGPT for method development and green chemistry education in upper-level laboratory courses, *J. Chem. Educ.* **101**, 3224 (2024).
- [29] P. M. Hare, Coding with AI in the physical chemistry laboratory, *J. Chem. Educ.* **101**, 3869 (2024).
- [30] T. Humphry and A. L. Fuller, Potential ChatGPT use in undergraduate chemistry laboratories, *J. Chem. Educ.* **100**, 1434 (2023).
- [31] B. G. Glaser and A. L. Strauss, *The Discovery of Grounded Theory* (Routledge, London, 2017).
- [32] V. Braun and V. Clarke, Using thematic analysis in psychology, *Qual. Res. Psychol.* **3**, 77 (2006).
- [33] F. Marton, A. B. Tsui, P. P. Chik, P. Y. Ko, and M. L. Lo, *Classroom Discourse and the Space of Learning* (Routledge, London, 2004).
- [34] M. Holmqvist, L. Gustavsson, and A. Wernberg, Variation theory: An organizing principle to guide design research in education, in *Handbook of Design Research Methods in Education* (Routledge, London, 2014), pp. 129–148.
- [35] F. Marton, *Necessary Conditions of Learning* (Routledge, London, 2015).
- [36] A. Kullberg, Åke Ingerman, and F. Marton, *Planning and Analyzing Teaching*, 2nd ed. (Routledge, London, 2024).
- [37] F. Marton and S. Booth, *Learning and Awareness* (Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, 1997), pp. 1–224.
- [38] M. F. Pang and W. W. Ki, Revisiting the idea of “critical aspects”, *Scand. J. Educ. Research* **60**, 323 (2016).
- [39] M. Hasan, M. S. H. Khan, and A. F. Ahmed, Application of variation theory in STEM education: A comprehensive guideline for STEM teachers, *MethodsX* **12**, 102500 (2024).
- [40] Å. Ingerman, C. Linder, and D. Marshall, The learners' experience of variation: Following students' threads of learning physics in computer simulation sessions, *Instr. Sci.* **37**, 273 (2009).
- [41] E. Euler, B. Gregorcic, and C. Linder, Variation theory as a lens for interpreting and guiding physics students' use of

- digital learning environments, *Eur. J. Phys.* **41**, 045705 (2020).
- [42] S. Kilde Löfgren, J. Weidow, and J. Enger, Rolling balls or trapping ions? How students relate models to real-world phenomena in the physics laboratory, *Sci. Educ.* **107**, 1215 (2023).
- [43] A. Johansson, R. Méndez-Fragoso, and J. Enger, Developing a self-calibrating system for volume measurement of spheroidal particles using two acoustically levitated droplets, *Rev. Sci. Instrum.* **95**, 114901 (2024).
- [44] A. Johansson, S. Kilde-Westberg, and J. Enger, LeviLabs: Learning about sound through acoustic levitation, *Phys. Educ.* **59**, 063005 (2024).
- [45] Swedish National Agency for Education, Physics [subject syllabus] (2011), <https://www.skolverket.se/download/18.189c87ae1623366ff374c3/1521539980000/Physics-swedish-school.pdf>.
- [46] Swedish National Agency for Education, Curriculum for compulsory school, preschool class and school-age education—Lgr22 (2024), <https://www.skolverket.se/publikationer?id=13128>.
- [47] Statistics Sweden, Newborns, the 10 and 100 most common first names normally used, by year. Year 1998–2022 [data set] (2024), https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0001__BE0001D/BE0001T05AR/.
- [48] Swedish Research Council, Good research practice 2024 (2010), <https://www.vr.se/english/analysis/reports/our-reports/2025-07-03-good-research-practice-2024.html>.
- [49] L. Chen, M. Zaharia, and J. Zou, How is ChatGPT's behavior changing over time?, *Harvard Data Sci. Rev.* **6** (2024).
- [50] M. Crossley, in *Analysing Qualitative Data in Psychology*, edited by E. Lyons and A. Coyle (SAGE Publications Ltd., London, 2007), Chap. 9, pp. 131–144.
- [51] Using the app Ragnar, an open-source project available on Github: <https://github.com/mickekring/ragnar.git>.
- [52] M. Cortazzi, Narrative analysis, *Lang. Teach.* **27**, 157 (1994).
- [53] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/ggy1-3kjk> for each group's entire chat history with ChatGPT during the lab sessions.
- [54] Rather, acoustic levitation utilizes the phenomenon of standing waves to trap particles in sound intensity nodes. As a visual example, see Fig. 3 in [43].
- [55] J. L. Araújo and I. Saúde, Can ChatGPT enhance chemistry laboratory teaching? Using prompt engineering to enable AI in generating laboratory activities, *J. Chem. Educ.* **101**, 1858 (2024).
- [56] E. Etkina, Pedagogical content knowledge and preparation of high school physics teachers, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020110 (2010).
- [57] E. Etkina, A. Karelina, M. Ruibal-Villasenor, D. Rosengrant, R. Jordan, and C. E. Hmelo-Silver, Design and reflection help students develop scientific abilities: Learning in introductory physics laboratories, *J. Learn. Sci.* **19**, 54 (2010).
- [58] B. R. Wilcox and H. Lewandowski, Students' views about the nature of experimental physics, *Phys. Rev. Phys. Educ. Res.* **13**, 020110 (2017).
- [59] N. G. Holmes and C. E. Wieman, Introductory physics labs: We can do better, *Phys. Today* **71**, No. 1, 38 (2018).
- [60] G. Polverini and B. Gregorcic, How understanding large language models can inform the use of ChatGPT in physics education, *Eur. J. Phys.* **45**, 025701 (2024).
- [61] Q. Xia, T. K. F. Chiu, C. S. Chai, and K. Xie, The mediating effects of needs satisfaction on the relationships between prior knowledge and self-regulated learning through artificial intelligence chatbot, *Br. J. Educ. Technol.* **54**, 967 (2023).
- [62] M. Wang, M. Wang, X. Xu, L. Yang, D. Cai, and M. Yin, Unleashing ChatGPT's power: A case study on optimizing information retrieval in flipped classrooms via prompt engineering, *IEEE Trans. Learn. Technol.* **17**, 629 (2024).
- [63] Z. Xu, S. Jain, and M. Kankanhalli, Hallucination is inevitable: An innate limitation of large language models, *arXiv:2401.11817*.