Examining student perceptions of custom GPT Chatbots to support introductory physics instruction

Dr. Nicholas Young, University of Georgia

Nicholas Young is an assistant professor of physics with a focus on physics education research at the University of Georgia. He earned his PhD in physics and computational mathematics, science, and engineering at Michigan State University, and he was a postdoctoral scholar at the University of Michigan's Center for Academic Innovation. His current research focuses on assessment in introductory physics courses and graduate education in physics.

Jasmine Freeman, University of Georgia

Jasmine Freeman is a first-year PhD physics student with a focus on quantum technology in condensed matter physics and physics education research at the University of Georgia. She earned her bachelor's degree, in physics and astronomy, at Valdosta State University.

Dr. Sarangan Balasubramaniam, The University of Georgia

Sarangan Balasubramaniam is a Lecturer in the Department of Statistics at the University of Georgia, where he has taught since 2022. He holds a Ph.D. in Computational Mathematics from the University of North Carolina Greensboro, with a dissertation focused on spatial prediction for axially symmetric processes on the sphere. His research interests span spatial statistics, machine learning, statistical learning, and biostatistics. At UGA, he teaches courses ranging from introductory statistics to advanced graduate-level data science, and M.S. in Applied Data Science program. He is also an affiliated faculty member of the UGA Statistical Consulting Center.

Dr. Nandana Weliweriya, University of Georgia

Nandana Weliweriya is a Senior Lecturer in the Department of Physics and Astronomy at the University of Georgia, Athens. His teaching and research focus on physics problem-solving, active learning pedagogies, and undergraduate laboratory instruction.

Examining student perceptions of custom GPT chatbots to support introductory physics instruction

Abstract
Introductory physics courses, often required for STEM majors, typically require a minimum grade of C for students to progress. Those who fall short must retake the course. While individualized feedback is crucial for mastering core concepts, student demand often exceeds what instructors can reasonably offer. Existing computer-graded assignments can help but often only provide a one-size-fits-all approach to student feedback. As such, there is a need to provide students with both additional and immediate personalized feedback and guidance to ensure they meet their learning and educational goals. With generative artificial intelligence tools becoming increasingly widespread and the barriers to entry becoming decreasingly prohibitive, chatbots can be leveraged to provide personalized help and promote student success.

Here, we used ChatGPT's custom GPT feature to develop chatbots for two introductory physics courses at a large, research-intensive university in the Southeastern United States. Each chatbot was solely trained on course materials including the syllabus, handouts, homework solutions, and practice exams. The customized chatbot then serves as an intelligent virtual tutor, providing 24/7 support by answering queries, guiding students through problem-solving, and reinforcing theoretical concepts.

The chatbots were offered as optional resources in three sections of a calculus-based, studio-style introductory physics for engineers course and two sections of an algebra-based, lecture-based introductory physics course. Students could use them for any part of the course except for in-class exams and were told that their instructor could not access the chatbot's history or logs.

To test the chatbot's efficacy, we administered an end-of-semester survey on how often students used the chatbot, the types of queries they asked, how often the chatbot was able to resolve their requests, and general feedback. This paper will present the results of that survey and suggestions for future improvements to better support student learning in important pre-requisite courses.

Introduction
Introductory science, technology, engineering, and mathematics (STEM) courses are prerequisites for many majors [1], and performance in these courses is strongly associated with degree completion [2]. As such, students often desire to do well in these courses so that they can progress in their majors.

Yet, this desire contrasts with the reality that many introductory STEM courses are designed to be or perceived as highly demanding, with low grades being common. Such courses have come to be known as "weed out" courses, given their outsized role in causing students to switch out of STEM majors and even leave the institution [3], [4], [5].

Due to high demand at large institutions, these courses are typically offered in large-enrollment formats. Struggling students in these courses may then have difficulty receiving help given the large student-to-instructor ratio [6], [7]. These difficulties can be exacerbated by practical constraints such as instructors being available only during working hours while students may often need assistance with course assignments outside of working hours, limited opportunities for individualized feedback, and that instructors may not be able to respond to questions promptly during times of peak demands such as near assignment deadlines and exams or before or after class periods [8], [9]. In addition, instructor support services such as office hours, help rooms, and tutoring centers can also suffer from these same constraints and in many large institutions, are already in place.

Generative artificial intelligence (GenAI) powered chatbots offer a potential solution. Chatbots can provide immediate, around-the-clock support for students by answering questions and offering explanations [10], [11], something typically available only during office hours or direct interactions with instructors and TAs. While still relatively uncommon, a recent survey of faculty suggests that around 15% have integrated a chatbot in their courses to support student learning [12], including engineering courses [13], [14], [15], [16], suggesting such an approach has promise.

Yet, despite the promise, minimal empirical work has been conducted on how chatbots are used in education [17]. In this paper, we seek to add to the literature by exploring student perceptions of custom chatbots developed with existing commercial tools offered in two introductory physics courses. We chose physics as a context given that many STEM majors, including engineers, are required to take at least one semester of physics and many non-STEM students choose to take physics courses to fulfill science general education requirements. In addition, introductory physics has a reputation for being an especially difficult course [18].

Specifically, we seek to answer the following research questions:

1) *How frequently do students make use of instructor-provided chatbots?*
2) *What benefits do students believe course chatbots provide them?*
3) *What tasks do students use course chatbots for?*
4) *What reasons prevent students from engaging with course chatbots?*

Brief background on using chatbots for education and physics

In this background, we focus on prior studies of chatbots and GenAI tools for education and learning in general and in physics. We do not discuss how GenAI tools and chatbots work here and direct readers to Polverini and Gregorcic for an overview in the context of education [19].

Prior work has documented multiple benefits of educational chatbots for students. Chatbots can improve learners' motivation and engagement through personalized support [11], [20] as well as knowledge retention and reasoning [21]. Students with low self-esteem or anxiety may also feel more comfortable interacting with a chatbot compared to a human instructor [22], especially outside of one-on-one interactions.

Prior experimental studies have also shown performance benefits of students using chatbots. One study in physics found that students who engaged with ChatGPT while problem solving performed better than those who did not [23]. Likewise, another study found that students who interacted with a chatbot outperformed students who interacted with the course instructor [24]. However, other work has found that the opposite, in that human-human collaborations outperformed human-GenAI collaborations though the authors noted that both groups showed improved physics problem solving [25].

Of course, the ability of chatbots to support course-specific learning depends on their capacity to solve discipline-specific problems. If the chatbot provided incorrect or inaccurate information, student learning could be hindered rather than helped [10]. Recent studies suggest that current commercial GenAI tools are capable of doing so in physics. For example, GenAI outperforms post-instruction students on a variety of conceptual inventories [26], can solve traditional introductory-level physics problems [27], [28], and even difficult physics-olympiad type problems [29], though performance generally decreases on problems requiring advanced calculations and content [27] and problems where the solver needs to make reasonable assumptions [28].

Further, even though the chatbots can correctly solve discipline-specific problems, students must trust their outputs in order to receive the benefits of using them. Recent work suggests that student trust of GenAI tools is variable, with some students showing great trust and others showing minimal [30] and that such trust is variable across the semester [31].

Even with perfect content knowledge, there can still be a few concerns that would prevent widespread usage of chatbots in the classroom. First, chatbots might solve problems using formulas or methods not yet covered in the course or included in the course at all or the output may be too verbose for students to understand the key parts of the solution [27]. While students can ask for clarification through follow up questions, students might become fatigued if they need to ask multiple follow-up prompts to reach an answer.

Second, students might not be interested in using them [32]. In addition, students have different background knowledge and different learning needs and chatbots are not trained to meet those differing needs. Therefore, even if a student is interested in using a chatbot in their course, they may not use it if they do not feel the chatbot was designed for them.

Methodology

Chatbot creation and course context

Using ChatGPT's Create a GPT tool with GPT 5 (available on Plus and higher plans at the time of paper submission, currently starting at $20/month), authors Weliweriya and Young created a custom chatbot for each of their introductory physics courses. They uploaded course materials such as their syllabus, previous homework solutions, a sample of lecture slides, suggested practice problems, practice exams and review guides, and the open-source textbook as the source of knowledge for the chatbot and disabled web search so that the GPT would not search the internet for answers. By combining ChatGPT's general ability to address questions with customized training on course-specific materials, this tool acts as a personalized learning assistant tailored to the unique needs of introductory physics students. The instructors also provided conversation starters to the GPT, which would be displayed to students as examples of questions they could ask.

During the fall 2025 semester at the University of Georgia, a large, research-intensive southeastern university, one of the chatbots was launched in two sections of a first-semester, lecture-based algebra-based introductory physics course (herein referred to as Physics 1A) taught by Young with a total enrollment of 420 students and the other was launched in three sections of a first-semester, active-learning, calculus-based introductory physics for engineers course (herein referred to as Physics 1C) taught by Weliweriya with a total enrollment of 216 students.

Students in Physics 1A are typically upperclassmen and enrolled in a non-physical science major, with many planning to apply to a professional degree program after graduation such as medical, veterinary, or dental school. Additionally, as upperclassmen, Physics 1A students have developed college-level study strategies and are typically more grade-conscious given their pre-professional goals, which may make them more likely to proactively seek additional resources.

Students in Physics 1C are typically freshman intended-engineering majors who must earn at least a “C” in the course before they can enroll in an engineering major. In addition, Physics 1C freshmen are navigating their first college courses and may initially underestimate the course difficulty compared to their high school physics experience.

Students were introduced to the chatbots during the first week of the semester and told they were an optional resource students could use to support them during the course including on homework assignments. In both courses, students had regular online pre-class assignments graded for completion, near weekly online homework assignments graded for accuracy, three in-person midterm exams graded for accuracy, and a comprehensive final exam. As both instructors had previously taught the other course, class slides, examples, and pre-class activities were similar, but exams and lab activities were different.

Survey development and deployment

We used the Technology Acceptance Model 2 (TAM2) as a framework to develop a survey to explore the chatbot’s perceived usefulness and perceived ease of use [33]. The TAM2 builds on the original Technology Acceptance Model by conceptualizing a user’s perceived usefulness of a technology in terms of its subjective norm, image, job relevance, output quality, and result demonstrability. The perceived usefulness and ease of use then leads to an intent to use the technology, which ultimately leads to usage behavior. In the context of this study, we then use the TAM2 model to motivate that student experiences with the course chatbots are important to understand because those experiences inform students’ decisions to engage with the course chatbots. All survey items were developed by the authors.

Students were first asked whether they had used the course chatbot and based on their response, shown one of two sets of questions. For students who had used the course chatbot at any point during the semester, they were first asked about the frequency of their usage and general perceptions of the chatbot. Afterwards, they were asked to indicate specific tasks they used the chatbot for in the course and their perceptions of the chatbot for those tasks. Prior work has found that AI-powered chatbots can provide feedback on student work, offer step-by-step solutions to homework problems, provide explanations of various topics, and help students prepare for exams [10], and we used these as the specific task choices. We also included understanding course policies as that was one of the main tasks instructors hoped students would use the chatbot for.

Students who did not indicate using the chatbot at any point in the semester were asked about their reasons for not doing so.

Both groups then responded to a series of questions about their use of generative artificial intelligence tools in general and demographic information. The final question for both groups asked whether their responses could be used for research purposes.

The survey was then deployed in the two sections of Physics 1A taught by Young and three sections of Physics 1C taught by Weliweriya during the last full week of classes. Students were given two weeks to complete the survey, which included a holiday break, and the survey was due on the last day of classes. Students who completed the survey were given a small amount of course credit in Physics 1A and a small amount of extra credit for completing the survey in Physics 1C.

At the time the survey was given in both courses, which occurred after the withdrawal deadline, there were 383 students still enrolled in Physics 1A and 199 students still enrolled in Physics 1C. 352 and 189 students completed the survey in Physics 1A and 1C respectively, and 278 and 153 students consented for their data to be used for research purposes. Therefore, our sample size for this paper is 278 Physics 1A students and 153 Physics 1C students.

Data analysis
Data from all students who did not consent to participate in this study were removed from the data set before beginning analysis.

For each question, results were analyzed separately for Physics 1A and 1C. All blank responses were only removed for that individual question, meaning that individual questions could have different numbers of responses.

For this paper, only raw numbers are reported and we do not perform any inferential statistics.

Results
Overall, we found that 55.4% of students in Physics 1A and 54.9% of students in Physics 1C reported using the chatbots at any point during the semester.

During the semester, the Physics 1A chatbot logged 1,544 chats and the Physics 1C chatbot logged 294 chats according to OpenAI's tracking.

Regarding usage, we found that the most common response among students in both courses was to use the chatbot a few times a week (Figure 1).

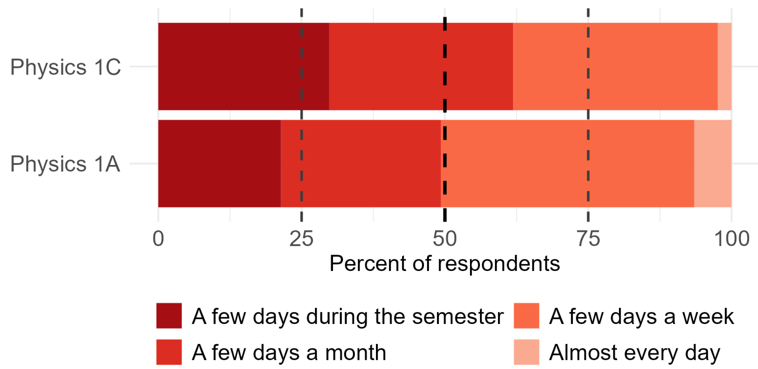

Figure 1: Frequency of course chatbot usage by students. The most common usage pattern was a few days a week.

Regarding usage of time, we found differing results in the two courses. In Physics 1A, we found that most students reported using the chatbot the same amount near the end of the course compared to the beginning of the course and around a third of students using it less near the end of the semester compared to the start. In contrast, we found that nearly 75% of students in Physics 1C reported using the chatbot more frequently towards the end of the course compared to the beginning (Figure 2).

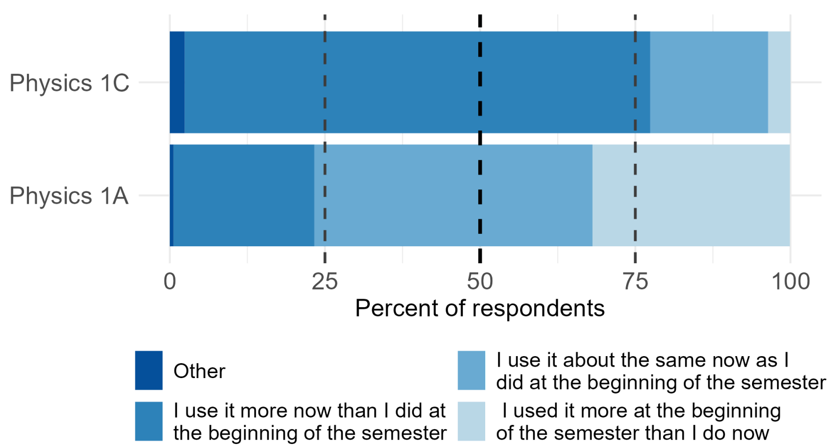

Figure 2: Course chatbot usage over time by students. In Physics 1C, students tended to use the course chatbot more as the semester progressed while in Physics 1A, students tended to use the course chatbot the same amount or less across the semester.

Next, we examined the general attitudes of students regarding the chatbots (Figure 3). In both courses, students tended to agree that the chatbots improved their learning, helped them complete their coursework more efficiently, and enhanced their understanding of the course materials. Students generally agreed that they were quickly able to get the chatbot to do what they wanted and that they would recommend future students use the chatbot.

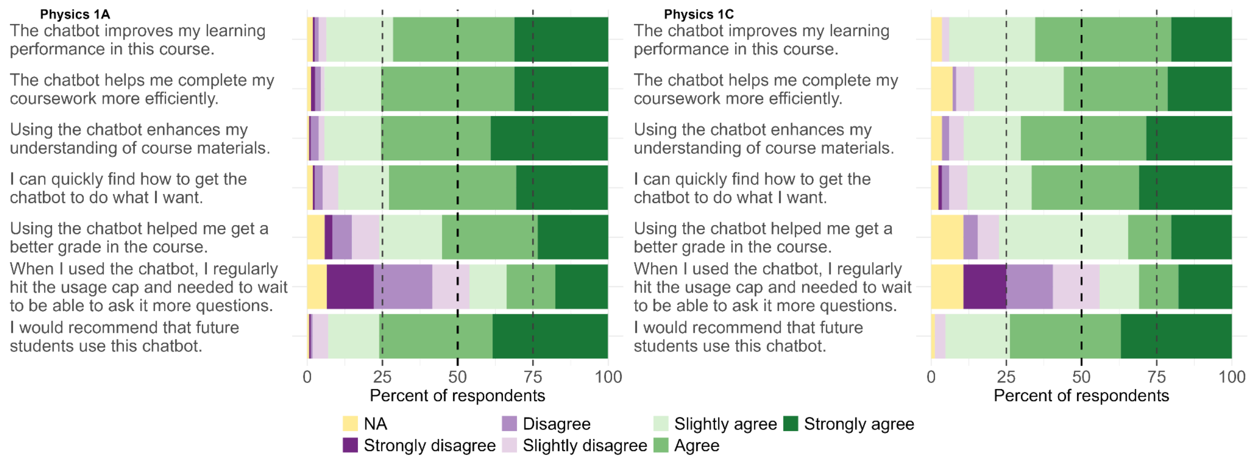

Figure 3: Distribution of responses to items about general usage of the chatbot. In general, students in both courses tend to agree that the chatbot benefitted them.

While the majority of students in both courses agreed that the chatbot helped them get a better grade in the course, nearly a quarter of students in each course disagreed or did not respond to the question.

Finally, we found that most students did not regularly hit usage caps on the chatbot. However, around 40% of students did agree at least somewhat that they hit the usage cap on the chatbot and needed to wait before they were able to ask it additional questions.

Regarding specific tasks students used the chatbot for (Figure 4), we find similar usage between both courses. In both Physics 1A and 1C, the most common reported usage was to help the student study for an exam followed by to clarify course concepts, with over half of the students in each course using the chatbot reporting using the chatbot for this purpose. In the algebra-based course, we also found a large number of students using the chatbot to help them with their homework problems while this was among the least reported usage in the calculus-based course.

In both courses, we also find that students rarely used the chatbots to help them understand course policies, with fewer than 25% of the chatbot-using students in each section reporting using the chatbot for this purpose.

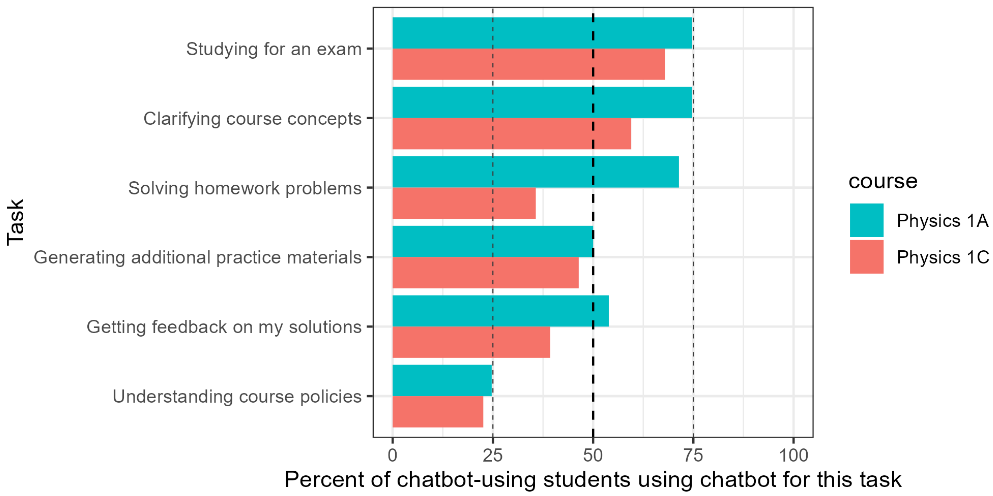

Figure 4: Tasks students used the course chatbot for. Studying for an exam and clarifying course concepts were the most frequent usages in both courses.

Among the students who did not report using the chatbot, we found that 34.7% of Physics 1A students and 5.8% of Physics 1C students did not know there was a course chatbot.

For the 65.3% of students in Physics 1A and 94.2% of students in Physics 1C who did know there was a course chatbot but did not use it, we examined potential reasons why this might be (Figure 5).

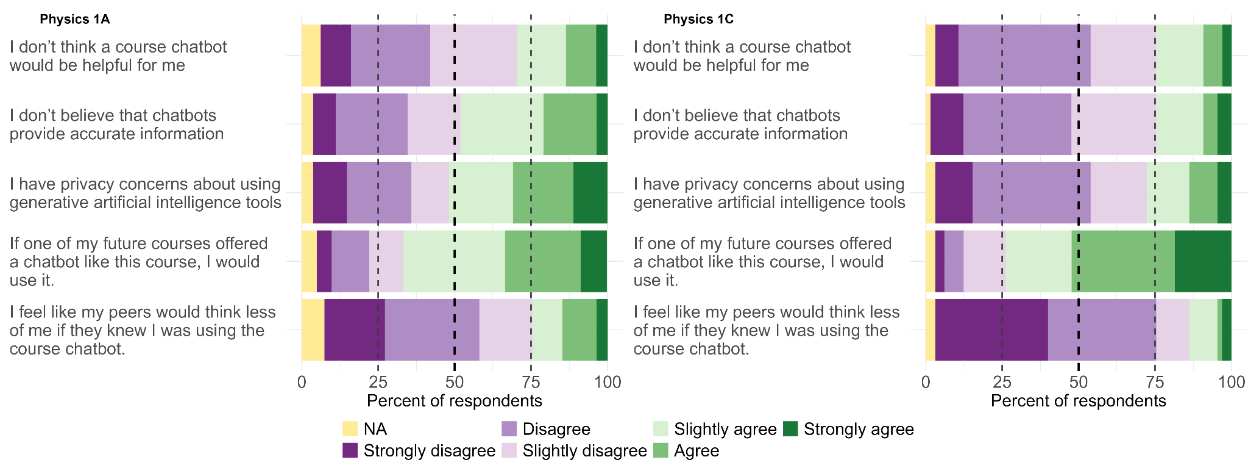

Figure 5: Distribution of responses to items about why students might not have used the course chatbot. In general, there did not seem to be a clear reason why these students did not use the chatbot.

Among the non-chatbot users, students in both courses reported similar responses. Most students disagreed that the chatbot would not have been useful for them or that chatbots do not provide accurate information. Most students agreed that if a future course offered a chatbot, they would use it. In addition, most students disagreed that their peers might think less of them if they found out they were using the course chat.

A notable difference between the two courses arose around the topic of privacy. In the general education physics course (Physics 1A), around half of the students not using the chatbot reported having a privacy concern around using generative artificial intelligence tools. In contrast, only around a quarter of the non-chatbot using students in the engineering physics course (Physics 1C) reported the same concern.

Discussion & Future Work

Here, we briefly provide answers to our research questions based on analysis done to date. We then suggest directions for more in-depth analysis of our survey using additional data students provided as well as future studies.

How frequently do students make use of instructor-provided chatbots?

We find that students in both courses reported using the chatbots frequently, with a few times a week and a few times a month as the most common results. Given that students had weekly homework and near monthly exams, this is perhaps not surprising.

Examining the two courses, we do find differences in how students reported using the chatbots over time. In Physics 1A, most students reported using the chatbot the same amount across the semester while in Physics 1C, most students reported using the chatbot more towards the end of the course.

Future work could examine why this difference in usage over time is present. For example, the difference could be due to differing difficulty in the course or workload expectations. Alternatively, the difference in usage could be the result of instructor messaging around the chatbot. While approximately equal fractions of students used the chatbot in both courses, nearly a third of students in Physics 1A did not know there was a course chatbot while nearly all students in Physics 1C knew there was a course chatbot, suggesting that the chatbot may have been promoted more to the Physics 1C students compared to the Physics 1A students. Finally, the difference could also be due to differing student populations as prior work suggests that students with higher academic demands may be more likely to use ChatGPT [34]. Students in Physics 1A were a mix of class levels while Physics 1C is almost exclusively freshman students.

Future work could also examine actual usage rates on a per-student basis. Examining usage trends would provide further insight into how usage changes across the semester as well as if there are times in the course when usage increases, such as near homework deadlines or before exams. Usage-level data are not available to the investigators because OpenAI's Custom GPT platform does not allow creators to access individual interaction logs or to associate conversations with specific users. Only aggregate metrics (e.g., total number of conversations) are provided. As such, collecting this data would likely require using a different platform. For example, other instructors used NotebookLM to build a physics course chatbot [35], suggesting that other options are available but analytics may still be limited. Other alternatives include institutionally hosted platforms where users authenticate through university credentials and data access is governed by institutional policies.

Larger scale projects could also investigate GenAI usage across courses rather than just a single course. While we do have self-reported data from students about the frequency of their GenAI usage in other courses, we did not collect any data about what courses they are using GenAI in or what types of tasks they are using them for. By looking at GenAI usage across courses, a more comprehensive picture of how students are using these tools could emerge.

At the same time, we also find a large fraction, approximately 45%, of the students in both courses did not use the chatbots. While our survey captured reasons from users who tried and stopped, we did not systematically investigate why nearly half of students never engaged with the tool despite its availability. One potential barrier, raised during our analysis, is student uncertainty about permission and surveillance. Even when instructors create and promote AI tools for learning support, students may fear these tools are surveillance mechanisms or 'traps' to identify academic dishonesty. This trust gap between instructor intentions and student perceptions may be a significant adoption barrier that deserves future investigation, particularly given research showing students' evolving trust in GenAI powered tools over time [31].

What benefits do students believe course chatbots provide them?

Students tended to agree that the course chatbots improved learning and resulted in better understanding of course materials. Most students also agreed to some degree that using the chatbot helped them earn a higher grade in the course, though the fraction of students disagreeing with this statement was the highest among any of the potential benefits.

As the survey was anonymous and we do not have student-level usage data, we are unable to examine whether these benefits actually occurred in practice. Therefore, future work with course chatbots could examine the relationship between usage and learning outcomes. For example, we could compare learning gains on conceptual inventories and final grades among students who used the chatbot with those who did not. Prior work around GenAI has shown mixed results in this area [36], [37], [38]

What tasks do students use course chatbots for?

We find that students in both courses most commonly reported using the chatbots to help them study for an exam and clarify course concepts.

Despite similar types of homework problems, we found that the students in Physics 1A used the chatbot to help them with homework problems more frequently than students in Physics 1C. One potential reason for this may be the existing resources in the course. Physics 1A did not have any graduate teaching assistants while Physics 1C had two teaching assistants assigned to each section along with multiple undergraduate learning assistants. In addition, Physics 1C had occasional “in-class office hours” where no new content was presented and the entire period was devoted to addressing student questions. In this case, students in Physics 1C had substantially more opportunities than students in Physics 1A to walk through homework problems with an instructor instead of needing to rely on a chatbot.

Of the options provided to students, we found that using the chatbot to understand course policies was the least reported usage option. Given that one of the motivations of the two instructors for creating the course chatbots was to streamline common class procedure questions, we found this surprising.

Future work will examine in greater detail how students used the chatbot for each of these tasks. As part of the survey, students answered additional questions about each task they reported using the chatbot for. This paper focuses on high level results and thus, did not include an in-depth examination of task-specific usage. In addition, students were asked to include a sample prompt for each task they used the chatbot. We plan to conduct thematic analysis of these prompts to describe how students are using the chatbot in greater detail.

Future work could include collecting all prompts students enter into the course chatbot for analysis. To our knowledge, such capability is not available in OpenAI's custom GPT tool and would require a different platform in order to conduct such analysis. Collecting the prompts could also help test one of the key assumptions made by the instructors. Given the popularity of ChatGPT, the instructors assumed that students were already familiar with GenAI tools and writing prompts. As such, students did not require any training or scaffolding to effectively use the chatbots. By analyzing all prompts submitted to the course chatbot instead of a select, and likely best, prompts submitted by students, we could examine whether this was actually the case or if students do need scaffolding in this area.

In addition, further analysis of student prompts could allow for additional insights into student AI literacy. While our survey asked about whether students were able to quickly get what they wanted from the chatbot, we did not include any follow up questions about the number of follow up questions students needed to arrive at their result. By examining prompts, we could examine whether students are constructing effective prompts the first time, effectively "debugging" the AI to redirect it towards their desired output, or something else.

What reasons prevent students from engaging with course chatbots?

Unfortunately, our data was unable to provide an answer to this question. Of the potential reasons we offered students who did not use the course chatbot, the students disagreed with most of them. For example, over half of the students disagreed that the course chatbot would not be helpful for them, that chatbots do not provide accurate information, and that their peers would think less of them for using the course chatbot. In addition, more than two-thirds of students who did not use the chatbot agreed that they would use the chatbot in a future course if given the opportunity.

Given the seemingly large disconnect between these students' opinions and their actions, future work should examine why these students did not use the chatbot. Non-chatbot-using students seem to believe the chatbot would be helpful for them and would use a chatbot like the ones offered in Physics 1A and 1C if offered in one of their other courses. Yet, these students did not use the Physics chatbots. A recent study has found a similar result in that students who did not choose to use ChatGPT to support their problem solving still viewed it as a tool to support their learning [23], suggesting that future work should explore this in more detail.

One potential explanation could be privacy concerns as around 50% of students in Physics 1A agreed that they had some privacy concerns around using GenAI tools. However, such a concern was much less prominent in Physics 1C despite similar usage rates suggesting that this is likely not the primary reason for not using the chatbots.

Future work should explore why some students did not use the course chatbot. As this was an observational study, the survey could not determine the underlying reasons for non-use. Conducting interviews with students who did not engage with the chatbot may provide richer insight into these factors.

Additional analysis

In addition to the data presented here, we also collected demographic and background data on students. These include the use of GenAI tools in other courses, gender, race, expected grade, GPA, and previous physics course taking. Future work will examine differences in usage and tasks based on these variables. For example, we might expect students who have not taken a previous physics course to use the course chatbot more than students who have taken a physics course before given that the latter group of students would have more physics knowledge and likely need less support in the course than the former group.

We also collected data about GenAI services that students subscribe to such as ChatGPT, Claude, and Gemini. Such data could provide further insight into usage rates, especially around students hitting the usage cap on the custom GPTs. While our survey results show students value chatbot support, an important equity concern emerged: full access to custom GPTs requires paid ChatGPT Plus subscriptions. In our study, this may have limited adoption rates and could disproportionately affect students from lower-income backgrounds.

Limitations

There are a few limitations to this study. First, the data were collected via a survey only at the end of the semester and only contains self-reported data. Therefore, the results reflect what students think they did rather than what they actually did. In addition, the data may exhibit recall and recency biases where responses are more reflective of what students have used the chatbots for recently as opposed to during the entire course.

Second, because the data were only collected at a single time, we are unable to investigate any changes over time. Further, due to the semester schedule, the window to complete the survey included part of a holiday break in order to give students adequate time. As such, students may have had their recall affected because they were not in “school mode”.

Third, the platform we chose to use for the chatbots contains limited options for analytics. As such, we are unable to actually look at usage trends or see the prompts students entered into the chatbots. Therefore, our results are unable to reflect in depth usage patterns. To our knowledge, this is a limitation placed on custom GPT creators from OpenAI and thus, the chatbots would likely need to be hosted on an alternative platform to obtain usage data.

Fourth, because the survey was anonymous and we do not have any individual-level usage metrics from the chatbot, we are unable to connect chatbot usage with course outcomes. That is, we are unable to examine any relationships between chatbot usage and student outcomes such as grades, test scores or learning gains.

Finally, there is likely a self-selection effect among respondents. Because the survey was framed around the chatbots, students who did not use the chatbots may have been less likely to take the survey, even though students were offered a small amount of course credit for completing it. As such, the fraction of students using the course chatbots may be an overestimation.

Preliminary implications and considerations for instruction and instructors

From our preliminary analysis, we are able to make a few preliminary implications for instructors.

First, messaging from instructors appears to be important for getting students to use the course chatbot. Around a third of students who did not use the chatbot in Physics 1A did not report knowing there was a chatbot while very few students reported the same in Physics 1C, suggesting that the instructor likely needs to advertise the chatbot throughout the course to make sure students are aware of the tool. This should include in-class announcements and demonstrations along with reminders and links available through the course learning management system.

The messaging should also actively build student trust through transparency. This requires: (1) explicit syllabus language clarifying when chatbot use is permitted and encouraged, (2) clear communication about data privacy; what is logged, who can access it, and how it will be used, (3) reassurance that the chatbot is a learning support tool, not a surveillance or assessment mechanism, and (4) clear boundaries distinguishing between appropriate chatbot use (e.g., concept clarification, problem-solving strategies) and academic dishonesty. Without this proactive transparency, students may perceive instructor-created chatbots as surveillance tools, limiting adoption regardless of pedagogical value.

Second, instructors should ensure that all students have access to the chatbot and can support their queries. Around 40% of students in both courses agreed to some degree that they often hit the chatbot usage cap and needed to wait to continue to use the chatbot. For chatbots to better support instruction and realize the goal of 24/7 assistance, students should not need to worry about hitting any usage caps. For our custom GPT chatbots hosted on commercial platforms, students need ChatGPT Plus subscriptions ($20/month as of January 2026) to access custom GPTs without restrictions, creating a financial barrier that likely could not be solved by an instructor alone. This requires institutional licenses or subsidies to ensure all students have equitable access. Failure to do so could allow more-resourced students to experience differential benefits compared to their less-resourced peers, potentially exacerbating existing performance gaps in introductory STEM courses (e.g. [39], [40], [41]).

Finally, instructors should make sure to acknowledge that the chatbots and GenAI tools in general are not always accurate and that students may need to check their outputs. Anecdotally, when solving homework problems, some students would use the chatbot first and if the chatbot gave the incorrect answer, then ask the instructor for help about why they were not correct. As such, instructors should be prepared to assist with AI "debugging" and explaining why the course chatbot may not have given a correct answer.

In addition, we provide two considerations for instructors based on our experience as instructors of these courses.

First, concerns about the chatbots using information not yet covered in the course did occur. Because the chatbots were trained on materials from the entire semester, the chatbots would, for example, occasionally use formulas from unit 3 when solving problems in unit 2. We did not see many occurrences of the chatbot suggesting formulas that were not covered in the course at all or that would be beyond what an introductory course would cover (i.e. suggesting a student construct a Lagrangian to solve a problem).

Second, the chatbot does not have a temporal sense and thus, when asking about time-specific situations, students need to include that information in the prompt. For example, when asking the chatbot to prepare for an exam, the chatbot would suggest topics from across the course rather than from the current unit. Even with the syllabus included as part of the training material, the chatbot would occasionally suggest material from later in the course. Therefore, students need to be explicit about what information and topics should be included in the review to ensure that the correct material is studied.

Initial conclusions

We find students seem to be open to using course chatbots to support their learning, with around half of students in two introductory physics courses using them when offered. We find that these students generally believe using the chatbot improves course outcomes such as grades and understanding of course concepts.

However, there is still a significant fraction of students who did not use the course chatbots. Without understanding the reasons for why this is, chatbots may remain another optional course support tool rather than a key resource for students.

Acknowledgments

We would like to thank the University of Georgia's ChatGPT for Education Trial program, which provided the authors with free access to ChatGPT including the custom GPT feature.

We would also like to thank the members of the Georgia Physics and Astronomy Education Research (G-PhAsER) group for their feedback on survey items.

References

[1] C. L. Meaders *et al.*, "Undergraduate Student Concerns in Introductory STEM Courses: What They Are, How They Change, and What Influences Them," *J. STEM Educ. Res.*, vol. 3, no. 2, pp. 195–216, Jul. 2020, doi: 10.1007/s41979-020-00031-1.

[2] N. Hatfield, N. Brown, and C. M. Topaz, "Do introductory courses disproportionately drive minoritized students out of STEM pathways?," *PNAS Nexus*, vol. 1, no. 4, p. pgac167, Sep. 2022, doi: 10.1093/pnasnexus/pgac167.

[3] J. Mervis, "Weed-Out Courses Hamper Diversity," *Science*, vol. 334, no. 6061, pp. 1333–1333, Dec. 2011, doi: 10.1126/science.334.6061.1333.

[4] T. J. Weston, E. Seymour, A. K. Koch, and B. M. Drake, "Weed-Out Classes and Their Consequences," in *Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education*, E. Seymour and A.-B. Hunter, Eds., Cham: Springer International Publishing, 2019, pp. 197–243. doi: 10.1007/978-3-030-25304-2_7.

[5] M. E. Thompson, "Grade Expectations: The Role of First-Year Grades in Predicting the Pursuit of STEM Majors for First- and Continuing-Generation Students," *J. High. Educ.*, vol. 92, no. 6, pp. 961–985, Sep. 2021, doi: 10.1080/00221546.2021.1907169.

[6] R. J. Swap and J. A. Walter, "An approach to engaging students in a large-enrollment, introductory STEM college course," *J. Scholarsh. Teach. Learn.*, vol. 15, no. 5, pp. 1–21, Nov. 2015, doi: 10.14434/josotl.v16i5.18910.

[7] G. Gibbs and A. Jenkins, Eds., *Teaching Large Classes in Higher Education*. London: Routledge, 2016.

[8] X. Liu, M. Pankiewicz, T. Gupta, Z. Huang, and R. Baker, "A Step Towards Adaptive Online Learning: Exploring the Role of GPT as Virtual Teaching Assistants in Online Education," 2025, pp. 149–166. doi: 10.1201/9781003567257-7.

[9] B. L. Whisenhunt, C. Cathey, M. E. Visio, D. L. Hudson, C. F. Shoptaugh, and A. D. Rost, "Strategies to address challenges with large classes: Can we exceed student expectations for large class experiences?," *Scholarsh. Teach. Learn. Psychol.*, vol. 5, no. 2, pp. 121–127, 2019, doi: 10.1037/stl0000135.

[10] L. Labadze, M. Grigolia, and L. Machaidze, "Role of AI chatbots in education: systematic literature review," *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, p. 56, Oct. 2023, doi: 10.1186/s41239-023-00426-1.

[11] C. W. Okonkwo and A. Ade-Ibijola, "Chatbots applications in education: A systematic review," *Comput. Educ. Artif. Intell.*, vol. 2, p. 100033, Jan. 2021, doi: 10.1016/j.caeai.2021.100033.

[12] C. Matherly, "The Emerging Multimedia Classroom: How AI Is Changing Course Design and Delivery," The Chronicle of Higher Education, 2025. [Online]. Available: https://www.chronicle.com/page/the-emerging-multimedia-classroom

[13] S. Abdulla, Y. M. A. Hamidi, and M. Khraisheh, "Creating and Implementing a Custom Chatbot in Engineering Education," presented at the 2023 ASEE Annual Conference & Exposition, Jun. 2023. Accessed: Jan. 19, 2026. [Online]. Available: https://peer.asee.org/creating-and-implementing-a-custom-chatbot-in-engineering-education

[14] R. L. Smith, R. T. Castles, and E. F. Sondergard, "The Engineering Ethics Coach: Integrating AI Chatbots to Strengthen Ethical Decision-Making in Engineering Economy," presented at the 2025 ASEE Annual Conference & Exposition, Jun. 2025. Accessed: Jan. 19, 2026. [Online]. Available:

https://peer.asee.org/the-engineering-ethics-coach-integrating-ai-chatbots-to-strengthen-ethical-decision-making-in-engineering-economy
[15] A. Kavianpour, "Integration of Chatbot with your course," presented at the 2025 ASEE Annual Conference & Exposition, Jun. 2025. Accessed: Jan. 19, 2026. [Online]. Available: https://peer.asee.org/integration-of-chatbot-with-your-course
[16] E. E. Danahy, M. K. Vora, Y. M. Xu, and W. Church, "Full Paper: Supporting Students' Educational Robotics Experiences through Generative AI Chatbots," presented at the 15th Annual First-Year Engineering Experience Conference (FYEE), Jul. 2024. Accessed: Jan. 19, 2026. [Online]. Available: https://peer.asee.org/full-paper-supporting-students-educational-robotics-experiences-through-generative-ai-chatbots
[17] C. McGrath, A. Farazouli, and T. Cerratto-Pargman, "Generative AI chatbots in higher education: a review of an emerging research area," *High. Educ.*, vol. 89, no. 6, pp. 1533–1549, 2025, doi: 10.1007/s10734-024-01288-w.
[18] F. Ornek, W. R. Robinson, and M. P. Haugan, "What Makes Physics Difficult?," *Int. J. Environ. Sci. Educ.*, vol. 3, no. 1, pp. 30–34, Jan. 2008.
[19] G. Polverini and B. Gregorcic, "How understanding large language models can inform the use of ChatGPT in physics education," *Eur. J. Phys.*, vol. 45, no. 2, p. 025701, Jan. 2024, doi: 10.1088/1361-6404/ad1420.
[20] F. R. Baskara, "Chatbots and Flipped Learning: Enhancing Student Engagement and Learning Outcomes through Personalised Support and Collaboration," *IJORER Int. J. Recent Educ. Res.*, vol. 4, no. 2, pp. 223–238, Mar. 2023, doi: 10.46245/ijorer.v4i2.331.
[21] X. Deng and Z. Yu, "A Meta-Analysis and Systematic Review of the Effect of Chatbot Technology Use in Sustainable Education," *Sustainability*, vol. 15, no. 4, p. 2940, Jan. 2023, doi: 10.3390/su15042940.
[22] N. Annamalai, M. E. Eltahir, S. H. Zyoud, D. Soundrarajan, B. Zakarneh, and N. R. Al Salhi, "Exploring English language learning via Chabot: A case study from a self determination theory perspective," *Comput. Educ. Artif. Intell.*, vol. 5, p. 100148, Jan. 2023, doi: 10.1016/j.caeai.2023.100148.
[23] Q.-A. Mirza and N. S. Rebello, "Help or Hype? Students' Engagement and Perception of Using AI to Solve Physics Problems," Sep. 30, 2025, *arXiv*: arXiv:2509.25642. doi: 10.48550/arXiv.2509.25642.
[24] H. B. Essel, D. Vlachopoulos, A. Tachie-Menson, E. E. Johnson, and P. K. Baah, "The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education," *Int. J. Educ. Technol. High. Educ.*, vol. 19, no. 1, p. 57, Nov. 2022, doi: 10.1186/s41239-022-00362-6.
[25] D. Tong, B. Jin, Y. Tao, H. Ren, A. Y. M. Atiquil Islam, and L. Bao, "Exploring the role of human-AI collaboration in solving scientific problems," *Phys. Rev. Phys. Educ. Res.*, vol. 21, no. 1, p. 010149, May 2025, doi: 10.1103/PhysRevPhysEducRes.21.010149.
[26] G. Kortemeyer, M. Babayeva, G. Polverini, R. Widenhorn, and B. Gregorcic, "Multilingual performance of a multimodal artificial intelligence system on multisubject physics concept inventories," *Phys. Rev. Phys. Educ. Res.*, vol. 21, no. 2, p. 020101, Jul. 2025, doi: 10.1103/98hg-rkrf.

[27] W. Yeadon and T. Hardy, “The impact of AI in physics education: a comprehensive review from GCSE to university levels,” *Phys. Educ.*, vol. 59, no. 2, p. 025010, Feb. 2024, doi: 10.1088/1361-6552/ad1fa2.

[28] R. Horchani, “ChatGPT’s problem-solving abilities in context-rich and traditional physics problems,” *Phys. Educ.*, vol. 60, no. 2, p. 025019, Feb. 2025, doi: 10.1088/1361-6552/adb473.

[29] P. Tschisgale, H. Maus, F. Kieser, B. Kroehs, S. Petersen, and P. Wulff, “Evaluating GPT- and reasoning-based large language models on Physics Olympiad problems: Surpassing human performance and implications for educational assessment,” *Phys. Rev. Phys. Educ. Res.*, vol. 21, no. 2, p. 020115, Aug. 2025, doi: 10.1103/6fmx-bsnl.

[30] M. Amoozadeh *et al.*, “Trust in Generative AI among Students: An exploratory study,” in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, in SIGCSE 2024. New York, NY, USA: Association for Computing Machinery, Mar. 2024, pp. 67–73. doi: 10.1145/3626252.3630842.

[31] D. Song, “How Learners’ Trust Changes in Generative AI over a Semester of Undergraduate Courses,” *Int. J. Artif. Intell. Educ.*, vol. 35, no. 3, pp. 1445–1464, Sep. 2025, doi: 10.1007/s40593-024-00446-6.

[32] J. Q. Pérez, T. Daradoumis, and J. M. M. Puig, “Rediscovering the use of chatbots in education: A systematic literature review,” *Comput. Appl. Eng. Educ.*, vol. 28, no. 6, pp. 1549–1565, 2020, doi: 10.1002/cae.22326.

[33] V. Venkatesh and F. D. Davis, “A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies,” *Manag. Sci.*, vol. 46, no. 2, pp. 186–204, Feb. 2000, doi: 10.1287/mnsc.46.2.186.11926.

[34] B. Agyare, J. Asare, A. Kraishan, I. Nkrumah, and K. Adjekum, “A Cross-National Assessment of Artificial Intelligence (AI) Chatbot User Perceptions in Collegiate Physics Education.,” *Comput. Educ. Artif. Intell.*, vol. 8, p. 100365, Jun. 2025, doi: 10.1016/j.caeai.2025.100365.

[35] E. Tufino, “NotebookLM: An LLM with RAG for active learning and collaborative tutoring,” Apr. 13, 2025, *arXiv*: arXiv:2504.09720. doi: 10.48550/arXiv.2504.09720.

[36] Y. Zhu, Q. Liu, and L. Zhao, “Exploring the impact of generative artificial intelligence on students’ learning outcomes: a meta-analysis,” *Educ. Inf. Technol.*, vol. 30, no. 11, pp. 16211–16239, Jul. 2025, doi: 10.1007/s10639-025-13420-z.

[37] X. Han, H. Peng, and M. Liu, “The impact of GenAI on learning outcomes: A systematic review and meta-analysis of experimental studies,” *Educ. Res. Rev.*, vol. 48, p. 100714, Aug. 2025, doi: 10.1016/j.edurev.2025.100714.

[38] N. Abdallah, R. Katmah, K. Khalaf, and H. F. Jelinek, “Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration,” *Soc. Sci. Humanit. Open*, vol. 12, p. 101866, Jan. 2025, doi: 10.1016/j.ssaho.2025.101866.

[39] R. Matz *et al.*, “Patterns of Gendered Performance Differences in Large Introductory Courses at Five Research Universities:,” *AERA Open*, Nov. 2017, doi: 10.1177/2332858417743754.

[40] S. D. Castle *et al.*, “Systemic advantage has a meaningful relationship with grade outcomes in students’ early STEM courses at six research universities,” *Int. J. STEM Educ.*, vol. 11, no. 1, p. 14, Feb. 2024, doi: 10.1186/s40594-024-00474-7.

[41] K. M. Whitcomb, S. Cwik, and C. Singh, "Not All Disadvantages Are Equal: Racial/Ethnic Minority Students Have Largest Disadvantage Among Demographic Groups in Both STEM and Non-STEM GPA," *AERA Open*, vol. 7, Jan. 2021, doi: 10.1177/23328584211059823.