



*Research article*

## **Potentiality of generative AI tools in higher education: Evaluating ChatGPT's viability as a teaching assistant for introductory programming courses**

**Zishan Ahmed, Shakib Sadat Shanto and Akinul Islam Jony\***

Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh

\* **Correspondence:** Email: [akinul@aiub.edu](mailto:akinul@aiub.edu).

Academic Editor: Alberto Borboni

**Abstract:** With the advent of large language models like ChatGPT, there is interest in leveraging these tools as teaching assistants in higher education. However, important questions remain regarding the effectiveness and appropriateness of AI systems in educational settings. This study evaluated ChatGPT's potential as a teaching assistant for an introductory programming course. We conducted an experimental study where ChatGPT was prompted in response to common student questions and misconceptions from a first-year programming course. This study was conducted over a period of 2 weeks with 20 undergraduate students and 5 faculty members from the department of computer science. ChatGPT's responses were evaluated along several dimensions—accuracy, completeness, pedagogical soundness, and the ability to resolve student confusion by five course faculties through a survey. Additionally, another survey was administered to students in the course to assess their perception of ChatGPT's usefulness after interacting with the tool. The findings suggested that while ChatGPT demonstrated strengths in explaining introductory programming concepts accurately and completely, it showed weaknesses in resolving complex student confusion, adapting responses to individual needs, and providing tailored debugging assistance. This study highlighted key areas needing improvement and provided a basis to develop responsible integration strategies that harness AI to enrich rather than replace human instruction in technical courses. The results, based on the limited sample size and study duration, indicated that ChatGPT has potential as a supplemental teaching aid for core concepts, but also highlighted areas where human instruction may be particularly valuable, such as providing advanced support. Further research with larger

---

samples and longer study periods is needed to assess the generalizability of these findings.

**Keywords:** ChatGPT, teaching assistant, higher education, educational technology, generative AI

---

## 1. Introduction

With recent breakthroughs in artificial intelligence (AI) and natural language processing, a new class of large language models has emerged, exemplified by systems like GPT-3 and ChatGPT<sup>1</sup>. Meanwhile, the generative AI tools like GPT-3 and ChatGPT based on GPT-3.5 already demonstrate quality in human-like dialogues and text generation that give hope for further increase in their use for educational purposes. Some propose that conversational agents like ChatGPT could take on teaching assistant (TA) roles to provide supplemental instructional support alongside human teachers [1].

However, the questions pertaining to the appropriateness, effectiveness, and limitations of deploying AI systems in such educational settings give rise to quite crucial ones that would need answering. More so, questions pertinent to such deployment are in technical subjects like computer programming, which demands from learners detailed, problem-oriented, logical thinking, and actual coding skills. The very nature of programming education—requiring the student to develop computational thinking, understand abstract ideas, and apply them in problem-solving—puts this in a unique place within the challenges presented to an AI-based teaching assistant. Hence, the effectivity with which AI systems provide targeted feedback, adjust with individual learning requirements, and instill practical coding skills is relevant to the success of its integration. Thus, a critical review of the strengths and weaknesses of using AI teaching assistants in the context of computer science education becomes central to informing the appropriate deployment and effective support for student learning.

The “appropriateness” of an AI system refers to its suitability for use in a specific educational setting, considering factors such as its ability to effectively support learning outcomes, align with pedagogical goals, and adhere to ethical and societal norms. This involves appropriateness of the impacts that might be caused by the AI system on the learning experiences of students, impacts on teachers, and also the educational environment as a whole.

The advent of large language models like ChatGPT has prompted extensive public debate on the potential societal implications of increasingly capable AI systems [2]. While acknowledging valid concerns around issues like misinformation, bias, and accountability, it is evident that AI also presents opportunities for human-centric symbiosis and collaborative intelligence if responsibly implemented [3]. Education is one domain primed for beneficial partnerships between human teachers and AI assistants that enhance rather than replace human roles [4]. There has been so much hype around this issue, often to the point of assuming that AI will replace teachers, but some researchers now suggest that real progress in AI-assisted education may be developing AI as a supplementary aid that complements human strengths and compensates for human cognitive limits [5]. However, more empirical evidence is needed to support this claim and to guide the effective integration of AI in educational contexts.

---

<sup>1</sup> ChatGPT is an AI chatbot developed by OpenAI, based on the GPT-3.5 language model. It is important to note that GPT-3 and ChatGPT are not competing systems, but rather ChatGPT is an application built on top of the GPT-3.5 architecture.

---

Programming pedagogy could stand to gain from AI systems that provide scalable, personalized support as students build computational thinking abilities. However, successfully integrating AI in education requires evidence-based insights on the appropriate scope and contributions of AI tools relative to human instruction [6]. Before presuming ChatGPT can serve as a virtual teaching assistant, rigorous investigation is needed to map its capabilities and limitations for adaptive programming instruction compared to human tutors. While large language models can democratize access to knowledge, upholding standards of learning quality requires understanding the unique value of human teachers in mentoring computational competencies [7].

This work aims to inform this understanding for responsible AI adoption in computer science education and to rigorously investigate the viability of ChatGPT as a virtual TA for introductory programming courses in undergraduate computer science curricula. Programming pedagogy relies heavily on constructionist learning, active problem-solving, and iterative coding—areas where current AI systems may lack human teachers’ abilities to scaffold understanding through interactive dialogue, provide adaptive guidance, and properly analyze student code [8]. Hence, while ChatGPT offers some advantages like 24/7 availability and scalability, it remains unclear if the tool can adequately explain complex programming concepts, resolve student misconceptions, and encourage systematic coding skills<sup>2</sup>.

In this study, an experimental study eliciting ChatGPT’s responses to common introductory programming questions and misconceptions was conducted. The AI’s explanations were evaluated by programming faculty along dimensions of accuracy, completeness, pedagogical soundness, and resolving confusion. Additionally, student perceptions of ChatGPT’s utility are surveyed after interacting with the tool. The contributions of this study are multifold. First, novel insights were provided on the effectiveness of ChatGPT specifically for TA roles in programming pedagogy based on faculty and student assessments. Second, gaps were identified between human and AI capabilities for adaptive programming instruction key to the guidelines for appropriate integration.

Overall, this work offers data-driven perspectives on both the promises and limitations of large language models for enhancing computer science education. While they may not fully replicate human teaching, with prudent design AI could still augment programming instruction and student learning. The findings will inform the best practices for incorporating generative AI in higher education in an ethical, equitable manner.

The paper is structured as follows: the literature review in section 2 discusses the growing interest in AI tools in education and highlights the need for assessing their pedagogical effectiveness. The methodology in section 3 details the participant recruitment, the questions posed to ChatGPT, and the survey questionnaires administered to both faculty and students. The results in section 4 provide the means from the faculty and student surveys, presenting where ChatGPT, as an AI TA, stands as a strength or weakness. This section looks into the analysis of the survey results on areas where ChatGPT did well and areas it failed, along with the results from the open-ended question. In section 5, the implications of using generative AI in education are inferred, and recommendations for further research based on the results obtained are drawn. Its structure is systematic and exhaustive, evaluating the potential that ChatGPT holds as a digital teaching assistant and, therefore, contains

---

<sup>2</sup> In this context, “systematic coding skills” refer to the ability to approach programming problems in a structured and logical manner, breaking them down into smaller sub-problems, designing algorithms, and writing clean, efficient, and well-documented code. These skills also encompass the ability to test, debug, and maintain code effectively.

---

valuable insights in the integration of such AI tools at universities and colleges. Finally, the study is concluded in section 6.

## 2. Literature review

Generative artificial intelligence (AI) tools like large language models have sparked growing interest in their potential applications in education [9]. Recent advances in natural language processing have enabled systems like ChatGPT to demonstrate human-like conversational abilities and language generation capabilities [10]. When AI technologies are created, sold, and used, it is important to weigh the advantages and disadvantages [11]. This has prompted questions on whether AI agents could take on teaching and tutoring roles to support or augment human instructors.

Several studies have explored the use of conversational agents and dialogue systems in education [12]. For instance, an AI teaching assistant focused on primary education was proposed that could answer student's questions, generate problems, and work examples, with an adaptive algorithm [13]. Their evaluations found the algorithm's performance proved to be positive. The ways to implement voice driven AI into classrooms were also investigated [14].

While promising, researchers have also highlighted the risks and limitations of relying on AI agents for teaching. For example, conversational agents tend to lack the empathy, social awareness, and theory of mind exhibited by human teachers and hence may struggle to establish rapport or effectively motivate student [15]. There are evident challenges regarding the use of AI in educational settings, such as the lack of adaptability, limited ability to provide personalized feedback, and potential biases in AI systems [16]. These challenges are particularly relevant for conversational agents and AI tutors developed before the advent of large language models, which may have more limited natural language processing capabilities compared to current models like ChatGPT. The extent of AI constraints in education depends on the form of AI framework and the technical basis. For instance, for small to medium conversational agents based on pre-large language models, it might be hard for them to understand the context, produce logical answers, or give feedback that is adaptive and has sense [17]. The current large language models like GPT-3.5 and ChatGPT have demonstrated improved performance in these areas [18]. However, even these advanced models may have limitations in terms of reasoning, common sense understanding, and alignment with educational goals, which require further investigation [19]. Moreover, AI systems trained on limited data can propagate biases and misinformation when used for instruction [20].

Recent works have proposed frameworks and guidelines for incorporating generative AI models into educational settings while addressing concerns around academic integrity. The PAIGE framework [21] outlines an AI system to detect potential cheating on assignments and promote academic honesty by distinguishing human versus AI-generated text. The authors suggest pairing generative models like ChatGPT with plagiarism detectors to uphold integrity standards. Similarly, Enriching the Learning Process with Generative AI [22] puts forth strategies to evaluate ChatGPT's ability to stimulate critical thinking through reflective writing exercises, while having instructors verify originality.

There have been hardly any studies that discuss the role of a ChatGPT-like model in assisting teachers. Therefore, their effectiveness as teaching assistants should be tested further, specially now when they are still new. At first glance, it seems that ChatGPT cannot handle intricate coding exercises involving reasoning; yet it can solve simpler ones related to programming languages [23].

---

This raises uncertainties about its suitability for assisting in technical courses like programming.

As a whole, current literature suggests that it is important to conduct thorough investigations that evaluate how good or bad academic-teaching tools (for various subjects) are in today's age; such work should also assess their efficiency. In this study, authentic data is provided about whether ChatGPT can be used at colleges/universities for teaching these subjects using information provided by teachers and students. The findings are expected to inform guidelines on appropriate integration of generative AI in education.

### 3. Methodology

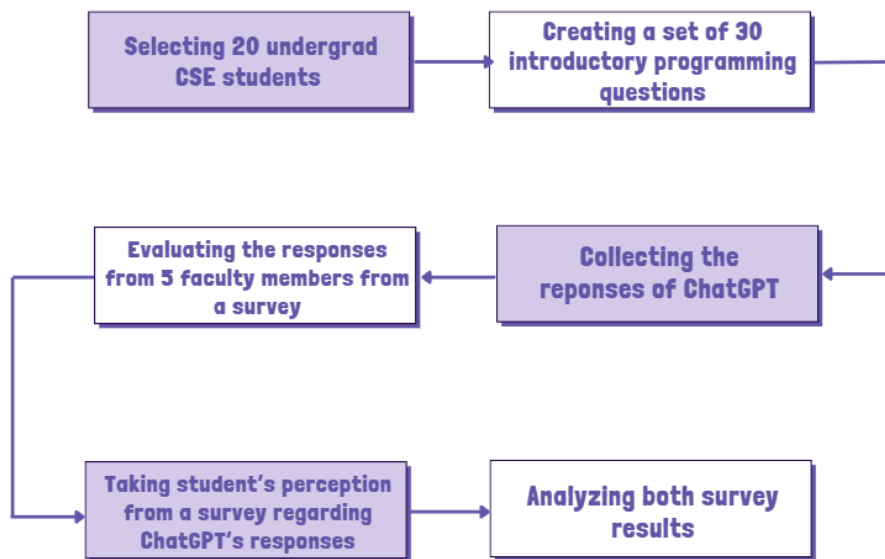
The participants were 20 undergraduate students from the department of computer science (ages 18–22 years old) enrolled in an introductory Python programming course. All participants were recruited through announcements made by the course instructor during class. Interested students were informed that the study involved evaluating an AI teaching assistant for the programming course.

The student participants were given access to ChatGPT for a period of 2 weeks during their programming course. They were instructed to ask at least 15 questions from the pre-compiled set of questions as they arose naturally during coursework. ChatGPT's responses were collected. Five faculty members who commonly teach introductory programming then evaluated the ChatGPT responses.

The five faculty members who evaluated ChatGPT's responses were chosen based on their experience in teaching introductory programming courses. All faculty members had a minimum of three years of teaching experience in computer science, with an average of 5 years. Three of the faculty members held doctoral degrees in computer science or a related field, while the other two held master's degrees. Additionally, all faculty members had completed pedagogical training or workshops focused on effective teaching strategies in computer science education.

Using a 5-point Likert scale, the faculty members assessed accuracy, completeness, pedagogical soundness, and the ability to resolve confusion. After the 2-week period, the students completed the perception survey regarding ChatGPT's responses. They rated dimensions like understandability, accuracy, completeness, and resolving confusion on a 5-point scale. Finally, all of the results were analyzed to determine ChatGPT's ability as a teaching assistant.

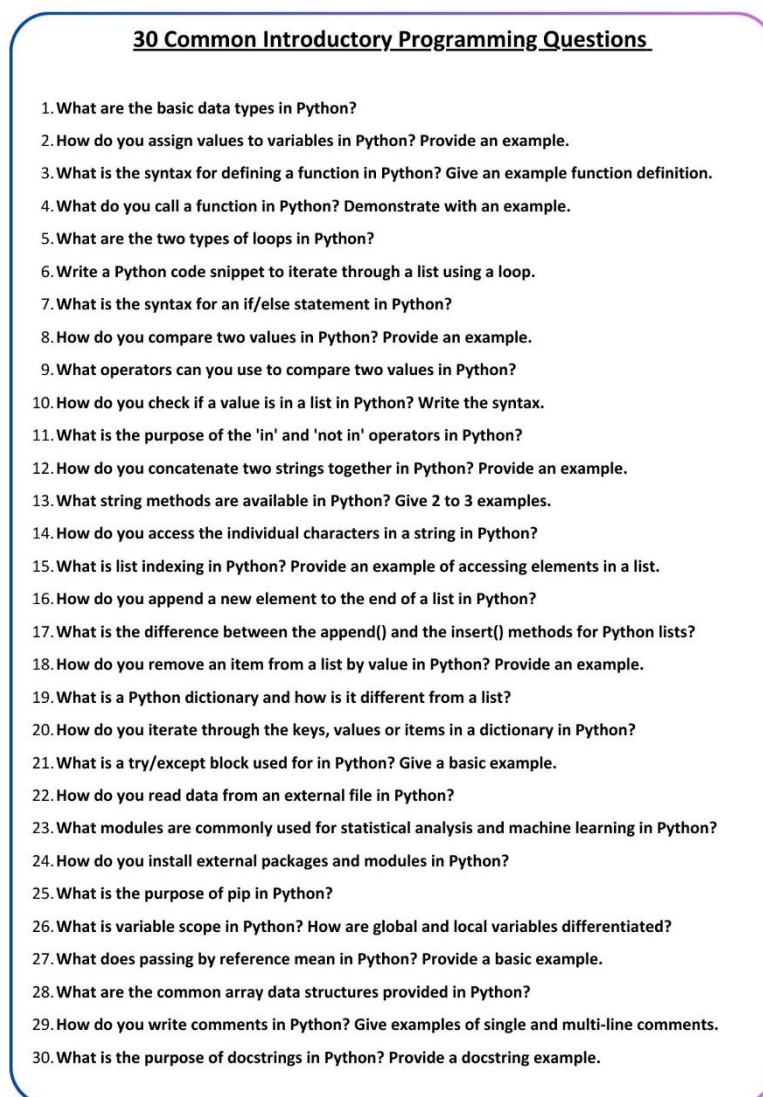
Figure 1 shows the overall methodology design of the study conducted. The diagram that follows outlines the method employed in the course of the study.



**Figure 1.** Methodology diagram.

### 3.1. Introductory programming questions

The questions cover a broad range of fundamental Python programming concepts and skills that would typically be encountered in an introductory coding course or tutorial. Topics include Python data types like strings, lists, and dictionaries as well as foundational programming techniques like variables, functions, conditional logic, loops, and arrays. The questions prompt students to demonstrate Python syntax, write basic code snippets that illustrate language features, use built-in Python functions and methods, work with collections like lists and dictionaries, handle strings and arrays, read/write files, import modules, install packages, understand scope rules, and utilize programming best practices like comments and docstrings. The set of questions aims to assess the comprehension of Python from multiple angles through short-answer, fill-in-the-blank, and coding exercises. There is a balance of theoretical conceptual questions and practical programming questions that require writing code. Overall, the set of 30 questions covers a diverse collection of Python topics and skills that provide insight into a student's mastery of introductory level Python programming. The full question set is provided in Figure 2.



**Figure 2.** Introductory Python programming question set.

### 3.2. Survey questionnaire

A 10-item Likert scale questionnaire was developed to assess faculty perceptions of ChatGPT's performance as a teaching assistant for introductory programming concepts. The faculty survey included statements evaluating various dimensions such as accuracy, completeness, pedagogical strategies, language, resolving confusion, providing examples, handling edge cases, encouraging problem solving, appropriate tone, and overall helpfulness. Additionally, an open-ended question allowed faculty to provide any other relevant feedback on ChatGPT's responses. This questionnaire enabled comprehensive quantitative and qualitative data to be collected from faculty on how ChatGPT's teaching assistant abilities compared to human teaching assistants for introductory programming content. Figure 3 shows the faculty survey questionnaire.

### ChatGPT Teaching Assistant Faculty Perception Survey

Instructions: Please rate how much you agree with the following statements regarding the responses provided by ChatGPT to introductory programming questions on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree).

QUESTIONS:	RATING SCALE:				
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. The responses demonstrated accurate knowledge of programming concepts.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. The explanations provided were complete and thorough.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. The responses used effective teaching methods and strategies.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. The language and terminology used was appropriate for introductory students.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. The responses effectively resolved potential student confusion about concepts.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. The responses provided useful examples and sample code when applicable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. The responses handled exceptions and edge cases appropriately.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. The responses encouraged good programming practices and problem-solving skills.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. The tone and style of the responses was appropriate for a teaching assistant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Overall, the responses would provide helpful pedagogical support for introductory programming students.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Figure 3.** Survey questionnaire on faculty perceptions of ChatGPT's performance as a teaching assistant.

A parallel 10-item Likert scale questionnaire was administered to the student participants to gauge their perceptions of ChatGPT's usefulness as a programming teaching assistant. The student survey prompted participants to rate their level of agreement with statements regarding understandability, clarity, helpfulness of examples, providing code snippets, identifying knowledge gaps, assisting with debugging, encouraging systematic thinking, individualized responses, overall utility, and recommendation to peers. An open-ended question also allowed students to elaborate on their experience interacting with ChatGPT. This questionnaire provided crucial insight into student perspectives on the value of ChatGPT as a supplemental teaching aid for introductory programming pedagogy targeted to their comprehension levels and needs. Figure 4 displays the questionnaire used for surveying the students.



### ChatGPT Teaching Assistant Student Perception Survey

Instructions: Please rate how much you agree with the following statements regarding the responses provided by ChatGPT to introductory programming questions on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree).

QUESTIONS:	RATING SCALE:				
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. The explanations provided by ChatGPT were easy to understand.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. ChatGPT's responses clearly answered my questions.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. The examples given by ChatGPT were helpful for my learning.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. ChatGPT provided useful code snippets and samples when needed.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. ChatGPT helped me identify gaps in my programming knowledge.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. ChatGPT assisted me in debugging errors and exceptions.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. ChatGPT encouraged me to think through problems systematically.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. I felt ChatGPT tailored its responses to my individual questions.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Overall, ChatGPT was a useful teaching assistant for my programming questions.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. I would recommend ChatGPT as a programming teaching assistant to other students.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Figure 4.** Survey questionnaire on student perceptions of ChatGPT's performance as a teaching assistant.

### 3.3. Survey measurement scale

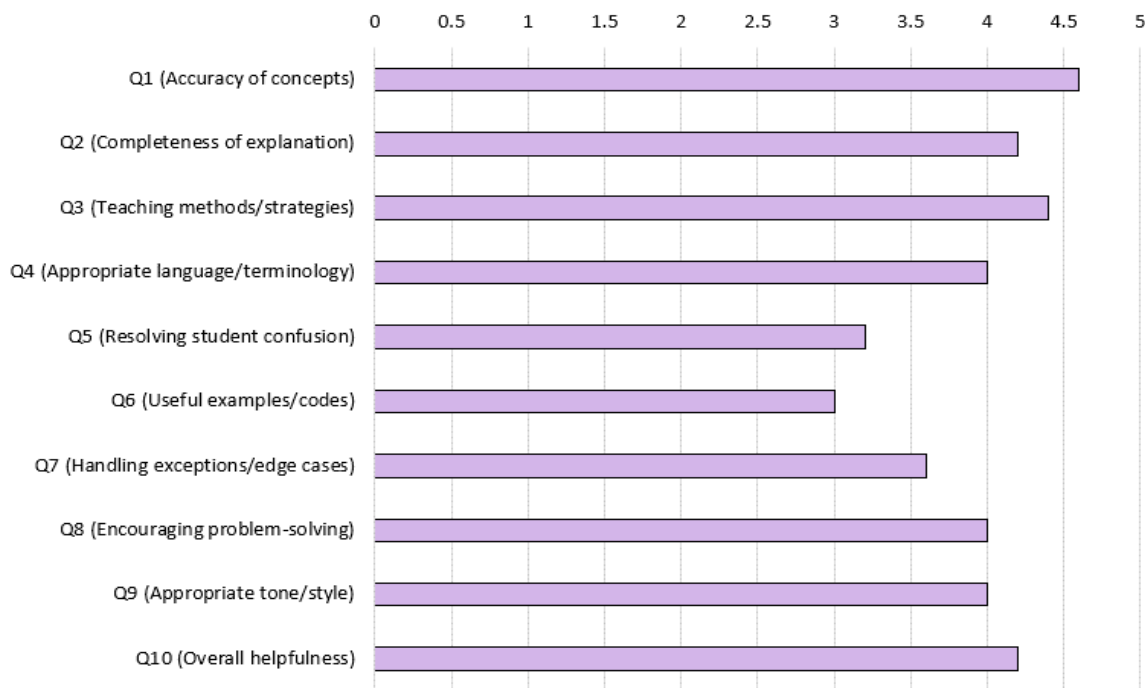
A Likert scale is a psychometric measurement technique commonly used in questionnaires and surveys to gauge attitudes, perceptions, and opinions [24]. In a Likert scale, respondents are presented with a statement and asked to indicate their level of agreement or disagreement on a symmetric agree-disagree scale. A conventional 5-point or 7-point symmetric scale is used, ranging from “strongly disagree” on one end to “strongly agree” at the other end, with a neutral midpoint. Respondents select the point on the scale that aligns best with their view. Each response is associated with a numerical score, allowing quantitative statistical analysis. Likert scales can efficiently capture gradations of opinion, perspective, and subjectivity from participants in a standardized way [25]. They produce interval data that provide insights into the respondent’s feelings toward the target item or experience being evaluated. Likert scale data assumes equidistance between scale points, allowing means and summations to be meaningfully computed. The Likert scale questions used in the faculty and student questionnaires for this study reflect a standard use of this measurement technique. The Likert scale used in this study is provided in Table 1.

**Table 1.** Likert scale.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

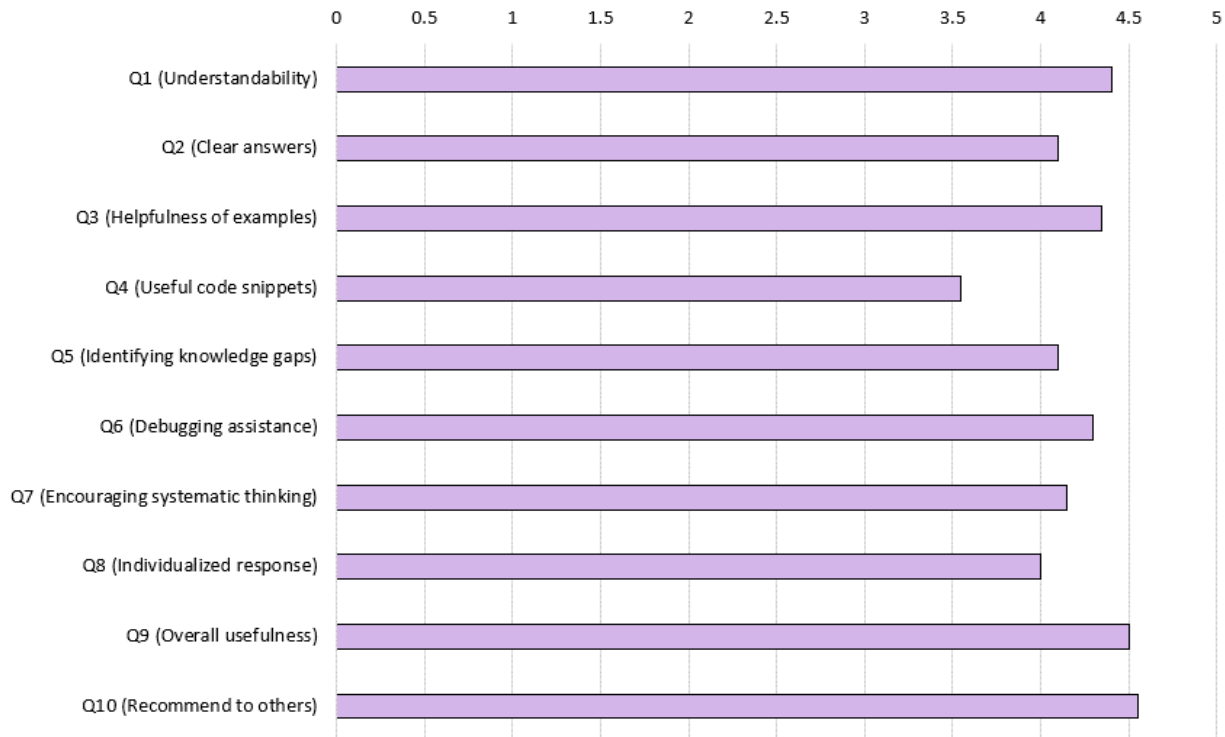
## 4. Results

### 4.1. Faculty survey results

**Figure 5.** Average ratings of the survey from the faculty members.

The 10-item faculty questionnaire generated the following average ratings of ChatGPT's responses across key dimensions of teaching assistant performance. For accuracy of programming concepts (Q1), completeness of explanations (Q2), use of effective teaching strategies (Q3), and encouraging problem-solving (Q8), ChatGPT received mean ratings between 4–4.6 out of 5, indicating general faculty agreement that the AI assistant demonstrated strengths in these areas. However, lower average scores were obtained for resolving student confusion (Q5, mean 3.2), providing useful examples and code (Q6, mean 3), and handling edge cases (Q7, mean 3.6). The lowest ratings came on the aspects of working through complex problems and limitations. The faculty gave an average score of 4.2 for ChatGPT's overall helpfulness as an introductory programming teaching assistant (Q10).

## 4.2. Student survey results



**Figure 6.** Average ratings of the survey from the students.

The 10-item student questionnaire produced the following average ratings regarding ChatGPT's performance across key dimensions of teaching assistant support. Students gave high mean scores for understandability (Q1, 4.4), providing clear answers (Q2, 4.1), usefulness of examples (Q3, 4.35), debugging assistance (Q6, 4.3), encouraging systematic thinking (Q7, 4.15), and overall usefulness (Q9, 4.5). Slightly lower but still agreeable ratings were given for ChatGPT's ability to identify knowledge gaps (Q5, 4.1), individualize responses (Q8, 4), and provide code snippets on request (Q4, 3.55). The highest average rating came on the willingness to recommend ChatGPT as a programming teaching assistant (Q10, 4.55). Generally, students perceived ChatGPT to be an effective teaching aid, with strengths in understandability, usefulness of examples, debugging support, and overall utility.

## 4.3. Open-ended question responses

In addition to the Likert scale questions, both faculty and student surveys included an open-ended question allowing participants to provide further feedback on ChatGPT's performance as a teaching assistant. The responses to these open-ended questions were analyzed using thematic analysis [26] to identify common themes and insights.

Faculty responses to the open-ended question highlighted ChatGPT's strengths in providing clear explanations for basic concepts and its potential to support student learning. However, they also emphasized the need for human instruction to address complex problems and provide individualized support. Some faculty members expressed concerns about the limitations of ChatGPT in handling edge cases and adapting to student needs.

Student responses to the open-ended question generally praised ChatGPT's ability to provide

---

helpful explanations, examples, and debugging assistance. Many students appreciated the accessibility and convenience of having an AI-based teaching assistant available 24/7. However, some students also noted that ChatGPT's responses could be inconsistent or lack the depth of human instructors in certain situations.

## 5. Discussion

The faculty and student surveys provide valuable insights into the potential strengths and weaknesses of ChatGPT as an AI teaching assistant for introductory programming concepts. Overall, both faculty and students gave reasonably high ratings for ChatGPT's performance across most dimensions evaluated in the questionnaires. The open-ended-question responses from faculty and students further support these findings, providing qualitative evidence for ChatGPT's strengths and limitations as a teaching assistant.

It is important to note that this is an exploratory study with a limited sample size of 20 students and 5 faculty members. While this sample size is appropriate for an initial investigation, the findings should be interpreted with caution, and further research with larger sample sizes is needed to establish the generalizability of the results. The conclusions drawn from this study are preliminary and serve as a foundation for future research with more robust sample sizes and study designs. The exploratory nature of this study should be considered when evaluating the potential of ChatGPT as a teaching assistant in introductory programming courses.

The study did not involve a direct comparison between human teaching assistants and ChatGPT. The emphasis that human instruction still plays an important role is seen from identified weaknesses in ChatGPT, and it does not put it head-to-head with human performance. Future work should attempt to identify the respective strengths and weaknesses of human instructors and AI-based systems within the context of introductory programming courses. Furthermore, the identified strengths and weaknesses of ChatGPT are mapped against specific survey questions, and hence they are not clear in this study. Subsequent research should be clear in establishing more evident relationships between the dimensions of interest and the relevant items on the survey to draw more firm conclusions.

The faculty responses suggested that ChatGPT evidenced good accuracy in the explanations of programming concepts, completeness in the breadth and depth of explanations, and an encouraging problem-solving approach. Ratings by faculty for these dimensions with respect to ChatGPT fell between 4.0 and 4.6 points on a 5-point scale, implying general agreement that ChatGPT showed human-like abilities in those aspects of the introductory programming pedagogy. The faculty was also impressed with the style of ChatGPT in the use of proper jargons and tone in the explanations given to the new students regarding the programming concepts.

Faculty, however, identified clear weaknesses in some of ChatGPT's responses. The lowest average faculty ratings received by ChatGPT included its ability to resolve student confusion, provide useful code snippets and examples, handle edge cases, and exceptions. The mean scores on these dimensions fell between 3.0 and 3.6, which does speak to some issues with the ability of ChatGPT to personalize explanations toward specific student difficulties and limitations. This is in line with the fact that large language models like ChatGPT have known weaknesses in properly responding to novel situations and questions too far from the purview of its training data. The importance of human instruction in programming education extends beyond the current limitations of

---

AI systems. Human instructors possess unique qualities such as adaptability, empathy, and the ability to provide individualized support [27]. These qualities are crucial for effectively guiding students through complex problem-solving processes and addressing their unique learning needs [28].

The student perceptions provide a complementary perspective. The students found ChatGPT highly understandable, with clear and complete explanations. They gave high marks for ChatGPT's usefulness of examples, debugging assistance, ability to encourage systematic thinking, and overall utility as a programming teaching aid. However, like the faculty, students identified providing individualized responses and useful code snippets on request as relative weaknesses, though still moderately positive in absolute terms.

It is important to note that while students rated ChatGPT's debugging assistance (Q6) relatively high (mean 4.3), this does not necessarily validate ChatGPT's ability to effectively debug complex code. The survey question assesses students' perceptions of ChatGPT's helpfulness in debugging, but does not directly measure its actual performance in resolving complex coding issues. Further research is needed to objectively evaluate ChatGPT's debugging capabilities compared to human teaching assistants.

The high ratings students gave in sections such as understandability, clarity, and overall usefulness point out that ChatGPT's strength is in making the presented programming topics coherent and clear for them. The students were also very receptive to examples and analogies ChatGPT used in explaining coding ideas, resonating with the tool's ability to tap into knowledge synthesized from huge training datasets. Student responses were also positive with respect to the way it helped them to learn systematic problem-solving in programming and, therefore, provided valuable assistance in debugging. This further emphasizes how useful ChatGPT is as a patient, tireless teaching assistant, demystifying complex concepts and giving step-by-step coaching in problem-solving. The students, however, were able to point out a relative weakness: the bot's inability to provide individualized responses and code snippets on request, thereby reflecting an inherent limitation of current AI systems. On the whole, the student perspective at least validates the strengths of ChatGPT in explaining intro programming basics comprehensively while pointing out that there will still be a need for human TAs to fix things, debug code, and provide tailored instruction. According to a student perspective, the paper reflects on promises and pitfalls of AI in education, emphasizing the guideline requirement concerning intelligent integration and human-AI collaboration.

The study revealed that while faculty members rated ChatGPT lower, students gave it higher marks, for example, in its usefulness. The students may view ChatGPT as an additional teaching tool, valuing its examples and ability to improve their learning process, which could account for this discrepancy in perception. Faculty, on the other hand, may have higher standards for the caliber and educational usefulness of ChatGPT's examples because they are seasoned educators.

Taken together, the faculty and student perceptions suggest that ChatGPT exhibits strengths in explaining introductory programming concepts, such as accuracy, completeness, and comprehension. However, further research is needed to validate these findings and to directly compare ChatGPT's performance to that of human teaching assistants. However, both groups surfaced limitations in ChatGPT's capacity to resolve complex confusion, provide examples and code, handle novel cases, and individualize responses. This aligns with the inherent strengths and weaknesses of large language models—a broad competency within the training distribution but a lack of true adaptability and reasoning.

---

The interpretation of the average rating for the performance of ChatGPT is relatively cautious. Such an average rating, ranging from 3.0 to 3.6, can be regarded as relatively low against other dimensions in this study. Without comparative data for humans performing similar tasks, human TAs or lecturers, a definite conclusion about the significance of these ratings cannot be made. Besides these factors, there are other instructor-related human factors, most likely time pressure and task complexity, that will influence their performance. Future research must aim to conduct comparative AI teaching assistant and human instructor studies which will provide a more holistic view over the relative advantages and disadvantages of these two.

The appropriateness of ChatGPT as an AI teaching assistant in introductory programming courses depends on various factors. While this study highlights ChatGPT's strengths in explaining core concepts and providing general assistance, its limitations in handling complex problems and providing individualized support raise concerns about its overall suitability. The effectiveness of ChatGPT in supporting learning outcomes and aligning with pedagogical goals requires further investigation. Moreover, the ethical implications of using AI systems in educational settings, such as potential biases and the impact on teacher-student interactions, must be carefully considered. Ensuring the appropriate use of ChatGPT and similar AI tools in programming education will require ongoing research, the development of best practices, and collaboration between educators, researchers, and AI developers.

Overall, ChatGPT seems to be of reasonable help in tackling core programming concepts but might fall short in the replacement of human teaching assistants for debugging complex coding challenges and resolving individual confusion by providing individualized coding examples and solutions. Using ChatGPT as an accompaniment to the TA would help to explain programming constructs and how they are put together, with qualified human tutors providing advanced debugging and project assistance. Indeed, faculty and students alike noted the promise of ChatGPT as a new AI teaching tool while surfacing places where human instruction seemed especially valuable, such as when walking the student through complex and open-ended learning tasks.

## 6. Conclusions

This study delves into what appears to be one of the most beneficial uses of ChatGPT, the large language model, in an artificial intelligence teaching assistant role in a first-semester course for programming students. These relevant findings were obtained from prompts and surveys through an experimental study, showing the strengths and weaknesses of ChatGPT in this educational context. Furthermore, the appropriateness of ChatGPT and similar AI tools in programming education should be carefully evaluated, considering factors such as their alignment with learning outcomes, pedagogical goals, and ethical considerations.

The results of the present study suggest that ChatGPT has promising potential for the coherent explanation of basic programming concepts. Most faculty and students rated the chatbots high in terms of accuracy, completeness, pedagogical strategies employed, and general help in learning the information. However, the deficiencies would be apparent in the case of the resolution of complex confusion and adapting to individual requirements. Deficiencies include offering help with debugging and tackling new cases.

Therefore, the findings suggest that ChatGPT has merit as a supplemental teaching aid, but also highlight the need for human instruction to address advanced concepts and provide individualized

---

support. The results align with the broad competency but lack of adaptability inherent in large pretrained language models like ChatGPT.

While promising, ChatGPT and similar technologies may be most effective when used as complementary tools focused on core competencies like explaining concepts, rather than as direct replacements for human teaching assistants. With carefully planned integration and a clear understanding of its strengths and limitations, AI has the potential to enhance programming education when used in conjunction with human instruction.

Future research should directly compare the performance and perception of human teaching assistants and AI-based systems like ChatGPT in introductory programming courses. Such studies will provide more concrete evidence regarding the relative strengths and weaknesses of each approach and inform the development of effective strategies for integrating AI tools into programming education.

Several promising ways of working in this area can be used as a starting point for research. Initially, it would be important to increase the number of participants in the experiment on introductory programming which would involve students from different programming courses. Moreover, it is necessary to test different models beyond ChatGPT to establish AI teaching assistants' effectiveness. Another area that requires further exploration is focusing on the development of instructions for integrating ChatGPT efficiently into the study process.

### **Author contributions**

Zishan Ahmed: Conceptualization, Methodology creation, Investigation, Resources, Data Curation, Survey questionnaire preparation, Writing - Original Draft, Writing –Abstract, Writing – Methodology, Writing – Discussion, Writing- Conclusion; Shakib Sadat Shanto: Result Analysis, Data Curation, Survey conduction, Writing – Introduction, Writing – Literature review, Writing - Results; Akinul Islam Jony: Conceptualization, Resources, Survey conduction , Writing - Review & Editing, Supervision. All authors have read and approved the final version of the manuscript for publication.

### **Use of AI tools declaration**

The authors declare that they have not used artificial intelligence (AI) tools in the writing of this article. However, the artificial intelligence (AI) tool, ChatGPT, was used to generate responses for the purpose of the study.

### **Acknowledgments**

The authors would like to thank the reviewers for their constructive feedback.

### **Conflict of interest**

The authors declare that they have no conflicts of interest.

### **Ethics declaration**

The authors declare that the ethics committee approval was waived for the study.

---

## References

1. Abedi, M., Alshybani, I., Shahadat, M.R. and Murillo, M., Beyond Traditional Teaching: The Potential of Large Language Models and Chatbots in Graduate Engineering Education. *Qeios*, 2023. <https://doi.org/10.32388/md04b0>
2. Lund, B.D., Wang, T., Mannuru, N.R., Nie, B., Shimray, S. and Wang, Z., ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, 2023, 74(5): 570–581. <https://doi.org/10.1002/asi.24750>
3. Horvatić, D. and Lipic, T., Human-Centric AI: The Symbiosis of Human and Artificial Intelligence. *Entropy*, 2021, 23(3): 332. <https://doi.org/10.3390/e23030332>
4. Yang, S.J., Ogata, H., Matsui, T. and Chen, N.S., Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence*, 2021, 2: 100008. <https://doi.org/10.1016/j.caeai.2021.100008>
5. Rastogi, C., Zhang, Y., Wei, D., Varshney, K.R., Dhurandhar, A. and Tomsett, R., Deciding Fast and Slow: The Role of Cognitive Biases in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 2022, 6: 1–22. <https://doi.org/10.1145/3512930>
6. Zhang, K. and Aslan, A.B., AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2021, 2: 100025. <https://doi.org/10.1016/j.caeai.2021.100025>
7. Rane, N.L., Choudhary, S.P., Tawde, A. and Rane, J., ChatGPT is not capable of serving as an author: ethical concerns and challenges of large language models in education. *International Research Journal of Modernization in Engineering Technology and Science*, 2023, 5(10): 851–874. <https://doi.org/10.56726/irjmets45212>
8. Romero, M., Lepage, A. and Lille, B., Computational thinking development through creative programming in higher education. *International Journal of Educational Technology in Higher Education*, 2017, 14: 1–15. <https://doi.org/10.1186/s41239-017-0080-z>
9. Alasadi, E.A. and Baiz, C.R., Generative AI in Education and Research: Opportunities, Concerns, and Solutions. *Journal of Chemical Education*, 2023, 100(8): 2965–2971. <https://doi.org/10.1021/acs.jchemed.3c00323>
10. Orr ù, G., Piarulli, A., Conversano, C. and Gemignani, A., Human-like problem-solving abilities in large language models using ChatGPT. *Frontiers in artificial intelligence*, 2023, 6: 1199350. <https://doi.org/10.3389/frai.2023.1199350>
11. Berendt, B., Littlejohn, A. and Blakemore, M., AI in education: learner choice and fundamental rights. *Learning, Media and Technology*, 2020, 45(3): 312–324. <https://doi.org/10.1080/17439884.2020.1786399>
12. Xu, W. and Ouyang, F., A systematic review of AI role in the educational system based on a proposed conceptual framework. *Education and Information Technologies*, 2022, 27(3): 4195–4223. <https://doi.org/10.1007/s10639-021-10774-y>
13. van Dijk, L.J., *AI as the assistant of the teacher: an adaptive math application for primary schools*. MS thesis, University of Twente, 2021. Available from: <https://essay.utwente.nl/88893/>.
14. Borthwick, K., Bradley, L. and Thou šny, S., *CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017*. Research-publishing.net, 2017.



15. Kim, J., Merrill, K., Xu, K. and Sellnow, D.D., My Teacher Is a Machine: Understanding Students' Perceptions of AI Teaching Assistants in Online Education. *International Journal of Human-Computer Interaction*, 2020, 36(20): 1902–1911. <https://doi.org/10.1080/10447318.2020.1801227>
16. Zhai, X., Chu, X., Chai, C.S., Jong, M.S.Y., Istenic, A., Spector, M., et al., A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, 2021, 2021: 1–18. <https://doi.org/10.1155/2021/8812542>
17. Aggarwal, D., Exploring the Scope of Artificial Intelligence (AI) for Lifelong Education through Personalised & Adaptive Learning. *Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN)*, 2024, 4(01): 21–26. <https://doi.org/10.55529/jaimlnn.41.21.26>
18. Osmanoglu, B., Forms of Alliances between Humans and Technology: The Role of Human Agency to Design and Setting up Artificial Intelligence-based Learning Tools. *Training, Education, and Learning Sciences*, 2023, 109. <https://doi.org/10.54941/ahfe1003154>
19. Anzor, F., Zulkifli, N.A., Jannah, D.S.M. and Krisnaresanti, A., Adaptive Learning Based on Artificial Intelligence to Overcome Student Academic Inequalities. *Journal of Social Science Utilizing Technology*, 2023, 1(4): 202–213. <https://doi.org/10.55849/jssut.v1i4.663>
20. Pedro, F., Subosa, M., Rivas, A. and Valverde, P., *Artificial intelligence in education : challenges and opportunities for sustainable development*. MINISTERIO DE EDUCACIÓN, 2019. Available from: <https://repositorio.minedu.gob.pe/handle/20.500.12799/6533>.
21. Shanto, S.S., Ahmed, Z. and Jony, A.I., PAIGE: A generative AI-based framework for promoting assignment integrity in higher education. *STEM education*, 2023, 3(4): 288–305. <https://doi.org/10.3934/steme.2023018>
22. Shanto, S.S., Ahmed, Z. and Jony, A.I., Enriching Learning Process with Generative AI: A Proposed Framework to Cultivate Critical Thinking in Higher Education using Chat GPT. *Tuijin Jishu/Journal of Propulsion Technology*, 2024, 45(1): 3019–3029.
23. Yilmaz, R. and Yilmaz, F.G.K., Augmented intelligence in programming learning: Examining student views on the use of ChatGPT for programming learning. *Computers in Human Behavior: Artificial Humans*, 2023, 1(2): 100005. <https://doi.org/10.1016/j.chbah.2023.100005>
24. Taherdoost, H., *What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale / Attitude Scale / Likert Scale*. papers.ssrn.com, 2019. Available from: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3588604](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3588604).
25. Ho, G.W.K., Examining Perceptions and Attitudes. *Western Journal of Nursing Research*, 2016, 39(5): 674–689. <https://doi.org/10.1177/0193945916661302>
26. Braun, V. and Clarke, V., Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 2006, 3(2): 77–101. <https://doi.org/10.1191/1478088706qp063oa>
27. Deng, Q., Zheng, B. and Chen, J., The Relationship Between Personality Traits, Resilience, School Support, and Creative Teaching in Higher School Physical Education Teachers. *Frontiers in Psychology*, 2020, 11: 568906. <https://doi.org/10.3389/fpsyg.2020.568906>
28. da S. Fernandes, P.R., Jardim, J. and de Sousa Lopes, M.C., The Soft Skills of Special Education Teachers: Evidence from the Literature. *Education Sciences*, 2021, 11(3): 125. <https://doi.org/10.3390/educsci11030125>

---

### Author's biography

**Zishan Ahmed** an enthusiastic undergraduate pursuing a Bachelor of Science in Computer Science and Engineering. In data science, natural language processing (NLP), and machine learning, he sees the greatest potential for innovation and influence. He is well-versed in several programming languages, including Python, Java, and C++, and is always keen to acquire new tools and technologies. His education has included data structures and algorithms, database management, artificial intelligence, and computer vision.

**Shakib Sadat Shanto** an undergraduate currently pursuing a Bachelor of Science in Computer Science and Engineering at American International University-Bangladesh. He is extremely passionate about artificial intelligence and the data science domain. He wants to do further research on Educational Technology, Natural Language Processing, and Cyber Security.

**Akinul Islam Jony** currently works as an Associate Professor of Computer Science at American International University-Bangladesh (AIUB). His current research interests include AI, machine learning, e-Learning, and educational technology.



AIMS Press

©2024 The Author(s), licensee by AIMS Press. This is an open access article under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).