

*The Development and Evaluation of Artificial Intelligence (A.I.) Tutor
for a Java Programming Class*

Alessia Tripaldelli, Embry-Riddle Aeronautical University, United States
Brian Butka, Embry-Riddle Aeronautical University, United States
Casey Elder, Embry-Riddle Aeronautical University, United States

The Paris Conference on Education 2023
Official Conference Proceedings

Abstract

The advent of Artificial Intelligence has brought numerous opportunities for innovation in the field of education. One such tool is ChatGPT, a conversational AI model, which can aid students in their learning process and enhance their Java programming skills. This study aims to evaluate the effectiveness of ChatGPT in helping students with Object Oriented Java programming tasks. The research was conducted by having four students work on the class assignments using ChatGPT as a supplementary tool. The students were helped and guided by the researchers to input the prompts in a way to make the AI produce the desired answer. Then, they were assessed through an exam taken first without the AI, and then with its help. The assessment included the topics where they struggled the most in the class. The students' performance was graded by the researchers and the students were interviewed about their experience with ChatGPT. In today's rapidly advancing technological world, it is imperative that education keeps pace. This study highlights the importance of adapting to technological advancements in education and teaching methods and provides valuable insights into the effectiveness of using AI as a teaching tool. The growing use of AI in language modeling has increased the interest in this field of research, and this study contributes to the body of knowledge by demonstrating the potential of ChatGPT as an AI tutor for programming tasks. The results of the study reveal that ChatGPT has the potential to play a valuable role in reinforcing teaching techniques and improving students' understanding. ChatGPT's features, such as code readability assessment, grading based on a predefined rubric, quality check of comments, and debugging, can provide students with immediate feedback and help them make necessary improvements.

Keywords: Artificial Intelligence, Object Oriented Programming, Machine Learning

iafor

The International Academic Forum
www.iafor.org

Introduction

Chat-GPT is a large language trained by OpenAI designed to respond to natural language input. This AI has the ability to respond to a large variety of questions, including programming and software development technical topics. ChatGPT uses a deep learning technique called a transformer neural network, which consists of multiple layers of self-attention mechanisms and feed-forward neural networks, enabling it to understand and generate coherent and contextually relevant text. This AI has been trained on a big amount of data from the internet, which helps it learn grammar, facts, reasoning abilities, and even some degree of common sense. When a prompt is input into ChatGPT, the response gets generated by sampling words from a probability distribution over the model's vocabulary. Thus, ChatGPT can generate appropriate responses based on patterns and information learned. The two phases in training this AI model are pre-training and fine-tuning. Pre-training involves gaining knowledge by predicting future events in a large dataset that includes portions of the internet, such as grammar, syntax, and semantics. Instead, during the fine-tuning phase, OpenAI created specialized datasets that are used to train the model, which compare and contrast alternative responses and examples of proper behavior. However, ChatGPT does not process information after September 2021 [1].

ChatGPT can be a useful tool in several different areas, such as writing, coding, explaining theoretical concepts, and many more. For instance, the creation of scientific abstracts and journal articles is one area where ChatGPT has shown its potential. In a recent study, researchers discovered that ChatGPT was capable of producing high-quality scientific abstracts that were equivalent to those produced by humans when combined with domain-specific data. Professor John Pavlik conducted a study on the effects of this AI on journalism, stating that this AI has become an important and influential technology for journalism and media, making it increasingly relevant for journalism and media education. Therefore, as this article argues, AI has the capacity to produce relevant, possibly high-quality written content that is relevant to the journalism and media domains [5].

In addition, our previous research titled *Can Artificial Intelligence Pass a Sophomore Level Digital Design Laboratory?* shows that ChatGPT can be a useful peer or a tutor in coding and electrical engineering material. While ChatGPT was unable to physically wire the circuits in the laboratories for the CEC222 class at Embry-Riddle Aeronautical University, it was able to describe the wiring, write the programs needed to run the laboratories, and provide answers to the lab questions. When the labs were scored individually, ChatGPT received a 73% [3]. Therefore, according to the study, ChatGPT performed on par with many other students and may be regarded as a peer in a laboratory environment where peer-to-peer learning is encouraged. ChatGPT has the potential to be a beneficial tool for students because students may feel more at ease asking simple questions to ChatGPT rather than to a person, which may increase their interest in the course content. However, while applying AI in education, it is vital to proceed with caution and consider any potential consequences on student learning and critical thinking abilities.

In this regard, even while ChatGPT can produce coherent and well-written essays, there are concerns about cheating and the loss of critical thinking skills that come with relying too much on AI-generated content. However, many professors carefully analyzed what ChatGPT may provide if used similarly to other programs already in use, including Grammarly. Although it is nothing new, learning quickly is not necessarily a bad thing. The professor John Warner in his essay "How About We Put Learning at the Center?" states that: "While much of the discussion of ChatGPT is about how disruptive it seems to be, my view is that it merely reveals

what was already present, rather than fundamentally changing the picture. For example, students taking a shortcut to a grade in order to avoid the longer journey through learning is not a new phenomenon” [4].

Moreover, it can be a useful tool to aid students with Object Oriented Java programming tasks, representing a virtual Teaching Assistant available all day to help students better understand what was taught in class. Chat-GPT performs well in coding technical assessments and code debugging. In fact, Chat-GPT was used for a LinkedIn Python skill assessment, producing a score of 85%. Thus, in the paper titled “ChatGPT outperforms 85% of the 4m programmers assessed on Python on LinkedIn,” was outlined that: “ChatGPT has proven its ability to understand and write code at a basic level, and beat basic assessment tests that are used to assess a candidate’s programming skills. Given the emphasis of hard skills in the industry today and many companies hiring based on these criteria, it is imperative to assess the way we assess the suitability of a candidate. Assessing only hard skills is no longer the solution if an AI machine like ChatGPT can do better than 85% of 3.9 million people” [5]. This AI is not only able to solve basic tasks and answer questions, but it can interpret and perform advanced tasks, which are acquired by programmers in years of studying. Therefore, Chat-GPT could represent a valid help for students that are struggling in understanding coding tasks providing examples and detailed explanations.

This research investigates whether Chat-GPT can represent a coding tutor for students, providing answers to their questions and a large variety of examples. This study worked with four students of varying abilities at Embry-Riddle Aeronautical University that are taking the Computer Science II Object Oriented Java class (CS 225). The researchers visited the class to ask the students the areas where they struggle the most. The students reported the critical areas as class inheritance between communicating programming classes, UML diagrams, and File I/O. In addition, the students reported a lack of examples and active learning in class, which was based more on theory. Thus, all the four students were helped by the researchers using Chat-GPT as a tutor to perform their homework, which reported a faster and more efficient completion of the task, which resulted in a better grade. In particular, the researchers helped in Chat-GPT’s prompt construction, to make the AI produce the desired answer.

Then, the four students were asked to take a coding exam first without the help of ChatGPT, and then with ChatGPT to test the knowledge they gained being tutored through it. In particular, three of the students were part of the class and were a computer science major, while one student was not currently enrolled in the class. The exam tested the students on two inheritance coding questions, a UML diagram question, one programming question, and two theoretical questions. The first time, the students took the exam in 30 minutes completely unaided. Then they were able to ask ChatGPT questions, and they were given the exam again.

The results showed that the student who was not in the class improved drastically, going from a failing grade to a passing grade (B). The other three students improved by a letter grade. All the students expressed positive feedback and evaluated the time they spent using ChatGPT more valuable than passively sitting in class listening to the professor. In this sense, we might be looking at a new way of tutoring students at any time, which raises the question if professors should be changing the way they teach in the future. The following section will provide more details about the way students were tutored and assessed.

Methodology

This research paper presents a pilot study aimed at investigating how students can use AI as a supplemental tutor in their classes. It is important to note that this study was conducted with a small number of subjects, and therefore, should be considered with caution and as a preliminary study as opposed to a definitive study of the topic. The use of the small sample size in the study is acknowledged and was driven by practical constraints.

To test the quality of ChatGPT's ability to help students perform on assignments, an exam containing 6 questions was given to the students twice. The first attempt was taken without the aid of Chat-GPT, while the second was supported by the help of the AI. Each question was worth five points, for a total of 30 points. The first question included the following:

Attributes and Methods (5 points): Create a Java class called Person with the following attributes:

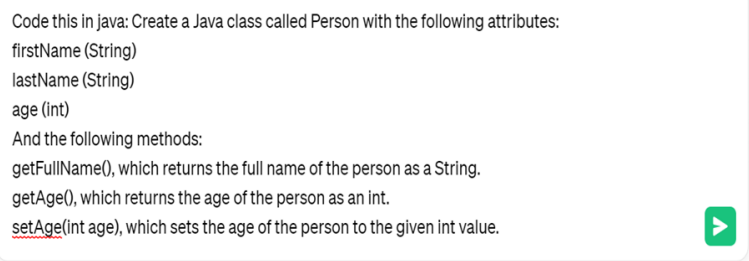
- firstName (String)
- lastName (String)
- age (int)

And the following methods:

- getFullName(), which returns the full name of the person as a String.
- getAge(), which returns the age of the person as an int.
- setAge(int age), which sets the age of the person to the given int value.”

The students overall performed decently on this question without ChatGPT initially with the average score being 3.75 points. After they were able to ask ChatGPT questions, the average score was 4.75 points.

For this question, the following images show what a student typed into ChatGPT and what was given as a response:



The screenshot shows a chat window with a white background and rounded corners. The user's prompt is in a light gray box on the left, and the AI's response is in a white box on the right. The response text is as follows:

```
Code this in java: Create a Java class called Person with the following attributes:  
firstName (String)  
lastName (String)  
age (int)  
And the following methods:  
getFullName(), which returns the full name of the person as a String.  
getAge(), which returns the age of the person as an int.  
setAge(int age), which sets the age of the person to the given int value.
```

Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts. [ChatGPT May 24 Version](#)

Sure! Here's the Java code for the 'Person' class with the specified attributes and methods:

```
java Copy code

public class Person {
    private String firstName;
    private String lastName;
    private int age;

    public String getFullName() {
        return firstName + " " + lastName;
    }

    public int getAge() {
        return age;
    }

    public void setAge(int age) {
        this.age = age;
    }

    // Constructor
    public Person(String firstName, String lastName, int age) {
        this.firstName = firstName;
        this.lastName = lastName;
        this.age = age;
    }
}
```

Figure 1: ChatGPT Response Example

The second question was based on creating a Java class to extend the “Person” class, to test the students on inheritance. The following prompt was given:

Inheritance (5 points): Create a Java class called Employee that extends the Person class and adds the following attributes:

- employeeId (int)
- salary (double)

And the following methods:

- getEmployeeId(), which returns the employee ID as an int.
- getSalary(), which returns the salary as a double.
- setSalary(double salary), which sets the salary to the given double value.
- Override the getFullName() method to include the employee ID in the full name.

The students performed well on this question initially with an average of 3.87 points. Using ChatGPT, the average became 5 points, getting a perfect score.

The third question was based on drawing a UML diagram to represent the two classes created in the two previous assessments. The following prompt was given to the students:

UML Diagram (5 points): Draw a UML class diagram for the Person and Employee classes, including all attributes and methods. Include inheritance and any necessary relationships.” The students got an initial score of 4.37 points. After using ChatGPT the score slightly changed, going up to an average of 4.67 points, because the AI does not have capabilities to draw UMLs. However, ChatGPT can provide help understanding the type of inheritance and consequently the type of arrow that has to be used.

The fourth prompt was based on a basic programming task to test the students’ proficiency with the Java language. The following question was given to the students:

Programming Task (5 points): Write a Java program that takes a list of integers from 0 to 9 as a user input and returns the sum of all even numbers in the list.” The initial score without using ChatGPT was 3.3 points. However, the attempt with the help of the AI reported an average score of 93%.

The fifth task was a theoretical question on one of the main concepts of the class, which is based on object oriented programming. The following prompt was given to the students: **Theory (5 points):** What is the meaning of object oriented programming?.” Without the help of ChatGPT the students earned an average score of 3.13 points. The help of ChatGPT made the score increase up to 5 points.

The sixth task was another theoretical question: **Theory (5 points):** What is data encapsulation?.” The first try reported an average score of 4 points, while the attempt with ChatGPT increased the score up to 5 points.

Results

After the students took the test the first time, unaided by ChatGPT, the average score was 71.25%. Then, after they were able to use ChatGPT as a tutor, the average score was 93.325%. The individual results can be seen in the following table, where the “a” questions are the first attempt at the test before ChatGPT and the “b” questions are the second attempt after they were able to ask ChatGPT questions.

Student #	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	
1	40%	100%	40%	100%	90%	90%	60%	20%	60%	20%	100%	40%	100%
2	90%	100%	90%	100%	100%	100%	100%	70%	100%	80%	100%	80%	100%
3	100%	100%	100%	100%	80%	80%	80%	90%	80%	80%	100%	100%	100%
4	70%	80%	80%	100%	90%	80%	80%	70%	100%	80%	100%	100%	100%
Average	75%	95%	78%	100%	90%	80%	63%	85%	65%	100%	80%	100%	

Table 1: Summary of Scores

The first and section questions’ average scores increased by 20% and 22%, respectively, after the students were allowed to use ChatGPT as a tutor. Since the questions were fairly similar, these similar results make sense. The questions should have been easy to answer regarding the content taught in the course, however the C average on these questions says otherwise. This result shows that the lack of examples being provided in class is negatively affecting students’ understanding of the concepts and students’ abilities to apply those concepts. After ChatGPT provided the students' example code, the scores increased significantly.

The third question required students to draw a UML diagram. Initially, the students seemed to understand what was being asked and, overall, succeeded. They did not need ChatGPT as much for this question as they did for the other ones. This result was convenient because ChatGPT can not produce images, therefore it can not draw a UML diagram. It does have the ability to explain what a UML diagram is and what the different symbols and arrows mean, but since it can not produce an actual diagram, it is not entirely useful in this sense. This is one of the most relevant limitations of ChatGPT.

The fourth question had an average increase of 23%, which was the second highest increase overall. This question asked the students to write a simple program, which was completely within the scope of the course content. Initially, most of the students knew what to do, but had

trouble implementing that in actual code syntax. However, after they were able to view sample code from ChatGPT, they were much more successful.

The fifth and sixth questions were theory questions about things they would have learned in class. The students did significantly better after asking ChatGPT for help. ChatGPT tends to perform well when asked questions that require explanations and also performs well with follow up questions.

After the students finished the second iteration of testing, they were all in agreement regarding how much better they felt about the second exam as opposed to the first. They all claimed that they learned more in our singular tutoring section than they learned in the entire duration of the class thus far. They claimed that the class was very lecture heavy, and often, no examples were given. They said that seeing ChatGPT provide example code that included comments describing what the code did was extremely beneficial to their learning and helped them understand the concepts taught in the lectures on a deeper level, so at that point, they were then able to apply the concepts taught in class to actual code.

Conclusion

Though ChatGPT is still a fairly new resource for students, it has proved to be very useful in the field of academics. Some theorize that it will cause students to lack critical thinking or problem solving skills, however others are able to see it as a beneficial thing to aid students' learning. However, it is necessary that ChatGPT is used carefully and lightly because sometimes it can give incorrect answers. In the article, "Collaborating With ChatGPT," the authors explain that "There are substantial limitations to the capabilities of AI, including its range and depth of knowledge and capacity to think critically or creatively. Yet, ChatGPT has an impressive level and range of knowledge" [2]. Even though ChatGPT houses some limitations that must be considered when using it in the field of academics, it is still an extremely useful tool for students. This can be seen in the test results taken in this study. They show that when students are able to use ChatGPT as a tutor, the scores increased greatly. Overall, ChatGPT is on the peer level with students, and should be used as a tutor, especially in the computer science field. ChatGPT has the potential to serve as a tutor in a programming class, as it is available 24/7.

Overall, ChatGPT's strengths outweigh its weaknesses in many ways. The results shown here show that after ChatGPT was used to tutor students, their performance increased significantly.

References

- [1] *Chatgpt*. ChatGPT. <https://openai.com/chatgpt>
- [2] J. V. Pavlik, "Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education," *Journalism & Mass Communication Educator*, vol. 0, no. 0, 2023, <https://doiorg.ezproxy.libproxy.db.erau.edu/10.1177/10776958221149577>
- [3] C. Elder, G. Pozek, S. Horine, A. Tripaldelli and B. Butka, "Can Artificial Intelligence Pass a Sophomore Level Digital Design Laboratory?," *SoutheastCon 2023*, Orlando, FL, USA, 2023, pp. 861-868, doi: 10.1109/SoutheastCon51012.2023.10115116.
- [4] J. Warner, "How about We Put Learning at the Center?: Inside Higher Ed," *Just Visiting*, 2023, <https://www.insidehighered.com/blogs/just-visiting/how-about-we-put-learning-center>
- [5] *CHATGPT outperforms 85% of the 4 million programmers assessed on python on linkedin*. CFTE. (2022, December 13). <https://courses.cfte.education/chatgpt-outperforms-4m-programmers/>

Contact emails: tripalda@my.erau.edu
butkab@erau.edu
elderc2@my.erau.edu