

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection School Of Computing and  
Information Systems

School of Computing and Information Systems

---

7-2023

### ChatGPT, can you generate solutions for my coding exercises? An evaluation on its effectiveness in an undergraduate Java programming course

Eng Lieh OUH

Singapore Management University, elouh@smu.edu.sg

Benjamin GAN

Singapore Management University, benjamingan@smu.edu.sg

Kyong Jin SHIM

Singapore Management University, kjshim@smu.edu.sg

Swavek WLODKOWSKI

Singapore Management University, swlodkowski@smu.edu.sg

Follow this and additional works at: [https://ink.library.smu.edu.sg/sis\\_research](https://ink.library.smu.edu.sg/sis_research)



Part of the [Programming Languages and Compilers Commons](#), and the [Software Engineering Commons](#)

---

#### Citation

OUH, Eng Lieh; GAN, Benjamin; SHIM, Kyong Jin; and WLODKOWSKI, Swavek. ChatGPT, can you generate solutions for my coding exercises? An evaluation on its effectiveness in an undergraduate Java programming course. (2023). *ITICSE 2023: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education, Turku, Finland, July 10-12*. 54-60.

Available at: [https://ink.library.smu.edu.sg/sis\\_research/8068](https://ink.library.smu.edu.sg/sis_research/8068)

This Conference Proceeding Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [cherylids@smu.edu.sg](mailto:cherylids@smu.edu.sg).



# ChatGPT, Can You Generate Solutions for my Coding Exercises? An Evaluation on its Effectiveness in an undergraduate Java Programming Course.

Eng Lieh Ouh  
elouh@smu.edu.sg  
Singapore Management University  
Singapore

Kyong Jin Shim  
kjshim@smu.edu.sg  
Singapore Management University  
Singapore

Benjamin Kok Siew Gan  
benjamingan@smu.edu.sg  
Singapore Management University  
Singapore

Swavek Wlodkowski  
swlodkowski@smu.edu.sg  
Singapore Management University  
Singapore

## ABSTRACT

In this study, we assess the efficacy of employing the ChatGPT language model to generate solutions for coding exercises within an undergraduate Java programming course. ChatGPT, a large-scale, deep learning-driven natural language processing model, is capable of producing programming code based on textual input. Our evaluation involves analyzing ChatGPT-generated solutions for 80 diverse programming exercises and comparing them to the correct solutions. Our findings indicate that ChatGPT accurately generates Java programming solutions, which are characterized by high readability and well-structured organization. Additionally, the model can produce alternative, memory-efficient solutions. However, as a natural language processing model, ChatGPT struggles with coding exercises containing non-textual descriptions or class files, leading to invalid solutions. In conclusion, ChatGPT holds potential as a valuable tool for students seeking to overcome programming challenges and explore alternative approaches to solving coding problems. By understanding its limitations, educators can design coding exercises that minimize the potential for misuse as a cheating aid while maintaining their validity as assessment tools.

## CCS CONCEPTS

• **Social and professional topics** → **Computing education;**  
**Computer science education;** • **Software and its engineering**  
→ **Object oriented languages.**

## KEYWORDS

programming, Java, object-oriented, computer science education

## ACM Reference Format:

Eng Lieh Ouh, Benjamin Kok Siew Gan, Kyong Jin Shim, and Swavek Wlodkowski. 2023. ChatGPT, Can You Generate Solutions for my Coding Exercises? An Evaluation on its Effectiveness in an undergraduate Java Programming Course.. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2023)*, July 8–12, 2023, Turku, Finland. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3587102.3588794>

## 1 INTRODUCTION

ChatGPT, a large-scale deep learning-based natural language processing model, is developed by OpenAI [13]. Built on the transformer architecture, this neural network excels in natural language tasks such as language translation, text summarization, question answering, and even programming exercises. Trained on a diverse array of text data, including books, articles, and websites, ChatGPT can generate human-like text and respond effectively to contextual prompts. This makes it an valuable tool for natural language processing tasks like chatbots and virtual assistants, with the potential to revolutionize technology [6].

AI-powered language models like ChatGPT are transforming programming education by offering real-time assistance and personalized feedback, enabling students to learn at their own pace and in a manner tailored to their individual needs. However, using such models in assessments raises concerns about examination integrity [18]. When students access the model during an exam, they could exploit it to generate answers for questions they cannot solve, leading to cheating and invalidating test results.

Programming and object-oriented concepts serve as the cornerstone of computing programs. Numerous research studies have investigated techniques and teaching styles to improve student learning outcomes in programming. These include adopting flipped classrooms [12, 15, 17], enhancing interactivity in video-based programming tutorials [2, 8, 16], and how teachers can support students in refining their code [7] to guide novice programmers towards better coding practices. In the era of generative artificial intelligence (AI), Susnjak [18] investigates concerns regarding online examination integrity with ChatGPT. Susnjak's conclusion emphasizes the importance of educators and institutions acknowledging the potential for ChatGPT-assisted

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

ITiCSE '2023, July 10–12, 2023, Turku, Finland

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0138-2/23/07...\$15.00

<https://doi.org/10.1145/3587102.3588794>

cheating and exploring measures to address it, thus maintaining the fairness and validity of online assessments.

This study seeks to address the research question (RQ): *"How effective is ChatGPT in generating coding solutions for Java programming exercises?"*

We evaluate the extent to which ChatGPT-generated solutions meet the requirements of coding exercises in our introductory Java programming course for undergraduate computing students. Each week, students receive programming exercises accompanied by expected outcomes and sample output. During the evaluation, we input the exercise instructions into ChatGPT (versions 3.5 and 4, the latest versions at the time of writing) to generate a solution. The generated programming codes are then executed on our local machine, and the results are compared to the expected outcomes for each exercise.

In Section 2, we provide a literature review focusing on student engagement in programming exercises, followed by an outline of our course design in Section 3. Section 4 offers a detailed account of the weekly exercises and their respective analyses. In Section 5, we summarize the effectiveness of ChatGPT-generated solutions. Finally, Section 6 presents the conclusion of our work.

## 2 BACKGROUND AND LITERATURE REVIEW

Generating code from a high level description of what they should do is a long-standing task in computer science [10]. Most prior work has been limited to either restricted domain specific programming languages or short code snippets. In contrast, generating entire programs often relies on understanding the task, reasoning out the appropriate algorithm to solve it, and then writing the code to implement that algorithm. AlphaCode is a system applied to code generation for competitive programming that can generate novel solutions to unseen programming problems [3]. Solving competitive programming problems represents a big step forward. The submission is correct only if it has the correct output on all hidden tests, otherwise it is incorrect. Hidden tests are not visible to the submitter, who must instead write their own tests or rely on the trivial example tests for debugging. Recent AlphaCode researcher analyses indicate that their model was indeed able to solve problems it has never seen before (learning) even though those problems require considerable reasoning.

Both ChatGPT and AlphaCode are 'large language models' systems based on neural networks that learn to perform a task by digesting massive amounts of existing human-generated text [3]. ChatGPT and AlphaCode are trained on different data sets. While ChatGPT is a general-purpose conversation engine, AlphaCode is specialised and trained exclusively on CodeContest datasets [1, 4, 9]. Another code generation system is GitHub's Copilot [5] which uses the OpenAI Codex. ChatGPT and Codex are OpenAI's products, suggesting that ChatGPT has access to and trained on GitHub public code datasets.

Recent popularity of ChatGPT has brought much attention to its benefits (AI pair programming) or problem (cheating) in education. Teachers may have to adjust their assessments by testing in class where they can be monitored or writing questions that required deeper understanding. For example, at Sacred Heart

University in Connecticut, teachers will have to give assessments that judge analytical reasoning and not just facts that can be looked up [11]. While much more advanced, teachers comparing ChatGPT to programming is similar to the introduction of calculator to mathematics. Besides adjusting the assessments, teachers can educate students about the consequences of cheating during proctored exams and provide students with support so that students do not see the need to cheat [11]. In this research, we explore to what extent ChatGPT can be helpful in solving Java programming exercises. Findings from this research can help educators understand how ChatGPT as a "calculator" can 1) enhance Java programming learning for students and 2) craft Java coding exercises so that cheating is minimised.

## 3 COURSE DESIGN AND CONDUCT

This course is designed for undergraduate students enrolled in computer science and information systems programs. Students are provided with video links to watch and quizzes to self-assess their understanding before attending the class. During class sessions, the instructor conducts coding demonstrations to emphasize key concepts and complement the pre-class materials. Subsequently, students are assigned in-class programming exercises to complete asynchronously. These programming exercises adhere to the requirements of the Oracle Java Foundations Certification [14] and are organized into the following learning categories.

- (1) Java Basics (28 exercises) - Test knowledge on variables, data types, conditionals, loops and collections.
- (2) Object-Oriented Concepts (11 exercises) - Test knowledge on object-oriented programming and understanding of API documentation.
- (3) Java Classes and Methods (17 exercises) - Test knowledge on method parameters, overloading, inheritance and polymorphism. Understanding of class and sequence diagrams using Unified Modelling Language (UML).
- (4) Java Exceptions and I/O (16 exercises) - Test knowledge on how to incorporate exception mechanisms when coding and the use of text files as a persistent storage.
- (5) Java Packaging (8 exercises) - Test knowledge on environment variables such as classpath, organise classes into packages and how to use the appropriate modifier for access control.

## 4 EVALUATION OF JAVA PROGRAMMING EXERCISES

For each exercise, we copy the exercise instructions and execute them on the ChatGPT [13] online. We subscribed to ChatGPT Plus to use both ChatGPT 3.5 and 4 models for this evaluation. We explicitly modify the instructions to include the keyword Java so that ChatGPT does not output codes in other languages. For example, we change "create a method" to "create a Java method". The exercise instructions are in text format. Some questions are supplemented with UML diagrams, compiled class files and/or API documentation to answer the questions. The generated solution is compared with our expected results to evaluate the effectiveness of ChatGPT in answering these exercises. We present the results for each exercise category in the following sections to address our research question.

## 4.1 Java Basics

These exercises require students to declare variables and apply conditionals and loops to implement logic in the main method of the Java source file. We provided expected outputs based on sample input data for students' understanding. Below is an example of the exercise, and Listing 1 shows the two generated solutions from ChatGPT. Both answers fulfilled the exercise requirements.

*Exercise 1 - Implement the method `sumOfDigits` that returns the sum of all the digits of that number. For example, `sumOfDigits(123)` returns 6 ( $1 + 2 + 3$ ). For negative numbers, the sum is the same value as that for a positive number. For example, `sumOfDigits(-354)` returns 12 ( $3 + 5 + 4$ ).*

```
// Correct Answer 1
public static int sumOfDigits(int n) {
    int sum = 0;
    n = Math.abs(n); // make sure n is positive
    while (n > 0) {
        int digit = n % 10;
        sum += digit;
        n /= 10;
    }
    return sum;
}

// Correct Answer 2
public static int sumOfDigits(int n) {
    int sum = 0;
    n = Math.abs(n);
    String numString = String.valueOf(n);
    for (int i = 0; i < numString.length(); i++) {
        sum += Character.getNumericValue(numString
            .charAt(i));
    }
    return sum;
}
```

Listing 1: Java Basics - Answers to Exercise 1

On exercises to write efficient code, Listing 2 shows the correct output from ChatGPT using `StringBuilder`. We realise that there is no mention of `StringBuilder` or comparison between `StringBuffer`, `+` operator and `StringBuilder` append method. However, this shortfall can be fixed by asking ChatGPT a follow-up question to compare these 3 methods.

*Exercise 2 - Write a memory efficient Java Method to concatenate two Strings.*

```
// Memory efficient answer
public static String concatenate(String str1,
    String str2) {
    StringBuilder sb = new StringBuilder();
    sb.append(str1);
    sb.append(str2);
    return sb.toString();
}
```

Listing 2: Java Basics - Answer to Exercise 2

ChatGPT is capable of generating code that produces the expected output. The following example illustrates an exercise with its expected output, while Listing 3 presents a correct solution generated by ChatGPT.

*Exercise 3 - Write Java code that contains only 1 "for" loop to produce the following outputs. 1 12 123*

```
// Correct Answer
for (int i = 1; i <= 3; i++) {
    for (int j = 1; j <= i; j++) {
        System.out.print(j);
    }
    System.out.println();
}
```

Listing 3: Java Basics - Answer to Exercise 3

For exercises that required more complex output, such as longer instructions to provide a context of the exercise, ChatGPT-3.5 encountered with incorrect answer but ChatGPT-4 is able to understand the exercise instructions. Below is an example of an exercise and Listing 4 shows an correct output from ChatGPT-4.

*Exercise 4 - You need to implement a static method that attempts to align two strings. This method takes in two parameters which are both of type `String`. Let's call the first parameter `str1` and the second `str2`. You can assume that `str1` is longer than or of the same length as `str2`. The method should try to align `str2` with `str1` and insert spaces for additional characters in `str1`. In the end the method returns true if there is an alignment, and meanwhile the method prints out the alignment. If no alignment can be found, the method returns false and prints out nothing. For example, if `str1` is "Recess Week" and `str2` is "RcWk", then we can align `str2` to `str1` as follows:*

Recess Week  
R c W k

```
// Correct Answer
public static boolean matchStrings(String str1,
    String str2) {
    int str1Len = str1.length();
    int str2Len = str2.length();
    int str2Index = 0;
    StringBuilder alignedStr2 = new StringBuilder();
    for (int str1Index = 0; str1Index < str1Len; str1Index++) {
        if (str2Index < str2Len && str1.charAt(str1Index) == str2.charAt(str2Index)) {
            alignedStr2.append(str2.charAt(str2Index));
            str2Index++;
        } else {
            alignedStr2.append(' ');
        }
    }
    if (str2Index == str2Len) { return true; }
    return false;
}
```

Listing 4: Java Basics - Answer to Exercise 4

There can be inconsistencies in the interpretation of program codes in ChatGPT. In one exercise to test the logical operators (&, &&, ||, |) and increment/decrement operators (++). While ChatGPT-3.5 can generate solutions correctly for each type of operation, ChatGPT cannot generate a correct solution with both operations included. Below Listing 5 shows a simplified version of this example. ChatGPT-3.5 is able to learn once we indicate to ChatGPT. However, when we re-evaluate the exercise in a new chat session and modify the conditions, the inconsistency appears again. We understand that the learning requires time to be reflected in new chats. ChatGPT-4 gives the correct answers for this exercise.

*Exercise - 5 This exercise explores the differences between short-circuit and normal logical operators. Given the following:*

```
// Correct Answer - age is 10
int age = 9;
System.out.println(++age);
// Incorrect Answer for ChatGPT-3.5
// age is 11 but ChatGPT output 10 initially.
// After repeating the evaluations,
// ChatGPT acknowledge the output should be 11.
char c = 'a';
int age = 9;
if (c == 'a' & ++age == 10){
    age++;
}
System.out.println(age);
```

**Listing 5: Java Basics - Operators. Answers to Exercise 5**

## 4.2 Object-Oriented Concepts

These exercises require students to implement classes and methods based on the provided API documentation and, in some cases, class files. Since ChatGPT's input is text-based, we can only submit the exercise instructions. Consequently, ChatGPT is unable to generate fully correct solutions without comprehending the API documentation and class files. Nevertheless, ChatGPT can still produce partial solutions, allowing students a starting point to tackle the problem. Exercise 6 and Listing 6 demonstrate a nearly correct solution generated by ChatGPT. However, ChatGPT's performance declines when presented with exercises containing limited text-based information. Exercise 7 and Listing 7 display the incorrect outputs generated by ChatGPT alongside the expected solution we anticipate.

*Exercise 6 - Study the Shirt API documentation. Write a program called ShirtTest to do the following:*

*a. Create the following objects using the appropriate constructor:*

Variable Name	Brand	Price	Color
shirt1	SMUgger	85.50	Red
shirt2	SMUgger	90.40	Blue
shirt3	(no brand)	77.60	Green

*b. Checks whether shirt1 and shirt2 are of the same brand using an appropriate method.*

*c. Check whether shirt2 and shirt3 are of the same brand using an appropriate method.*

*d. Set the color of shirt2 to red.*

*e. Print to console the text representation of shirt1, shirt2 and shirt3.*

```
// Partially Correct Answer
public class ShirtTest {
    public static void main(String[] args) {
        // a. Create objects
        Shirt shirt1 = new Shirt("SMUgger", 85.50, "Red");
        Shirt shirt2 = new Shirt("SMUgger", 90.40, "Blue");
        Shirt shirt3 = new Shirt(null, 77.60);
        shirt3.setColor("Green");

        // b. Answer is incorrect
        // Method isSameBrand and not sameBrand
        if (shirt1.sameBrand(shirt2)) {
            System.out.println("Shirts_1_and_2_are_of_the_same_brand");
        } else {
            System.out.println("Shirts_1_and_2_are_not_of_the_same_brand");
        }

        // c. Same as b. Answer is incorrect
        // Method isSameBrand and not sameBrand
        // Codes left out due to limited pages.

        // d. Set the color of shirt2 to red
        shirt2.setColor("Red");

        // e. Print the textual representation of
            shirt1, shirt2, and shirt3
        System.out.println("Shirt_1:" + shirt1);
        System.out.println("Shirt_2:" + shirt2);
        System.out.println("Shirt_3:" + shirt3);
    }
}
```

**Listing 6: Object Oriented - Answer to Exercise 6**

*Exercise - 7 You are given the Substance API documentation and class file. Complete the code for the calculateMass method inside the given file called RadioActiveTest.java*

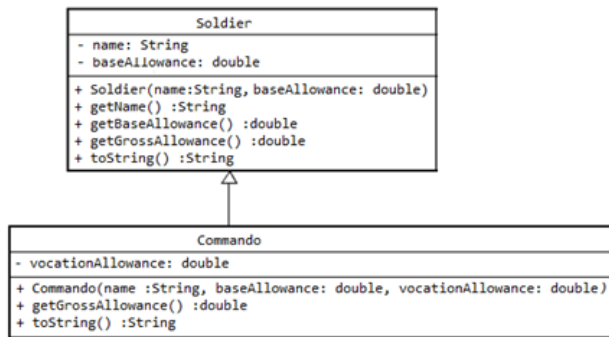
```
public class RadioActiveTest {
    // InCorrect Answer from ChatGPT
    public static double calculateMass(double atoms, double atomicWeight) {
        return atoms * atomicWeight;
    }
    // Some codes removed due to limited space.
    // A possible correct answer
    public static double calculateMass(double initialMass, double rateOfDecay, int years) {
        Substance sub = new Substance(initialMass, rateOfDecay);
        for (int i = 0; i < years; i++) {
            sub.decayOneYear();
        }
        return sub.getMass();
    }
}
```

**Listing 7: Object Oriented - Answers to Exercise 7**

### 4.3 Java Classes and Methods

These exercises require students to interpret Unified Modelling Language (UML) class and sequence diagrams to implement the required classes and methods. These exercises challenge ChatGPT with the instructions in the text but the diagrams as pictures. Below is an example of the exercise and Listing 8 shows the partially correct output from ChatGPT with missing attributes and methods, as shown in the class diagram.

*Exercise 8 - Most men who have completed their National Service with the military know that certain vocations (e.g. commandos) will be given an additional vocation allowance" as a reward for their higher risk. Study the class diagram below. Code the Soldier and Commando classes. The getGrossAllowance() methods should calculate salary using this formula:  
Soldier's gross allowance = baseAllowance  
Commando's gross allowance = baseAllowance + vocationAllowance*



```
// Partially Correct Answer from ChatGPT
class Soldier {
    private double baseAllowance;
    public Soldier(double baseAllowance) {
        this.baseAllowance = baseAllowance;
    }
    public double getGrossAllowance() {
        return baseAllowance;
    }
}

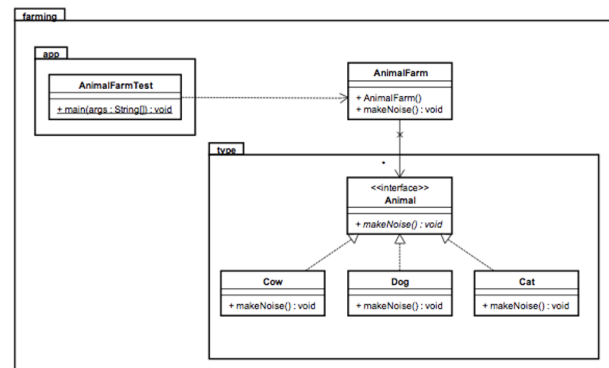
class Commando extends Soldier {
    private double vocationAllowance;
    public Commando(double baseAllowance, double
        vocationAllowance) {
        super(baseAllowance);
        this.vocationAllowance = vocationAllowance
    }
    public double getGrossAllowance() {
        return super.getGrossAllowance() +
            vocationAllowance;
    }
}
```

Listing 8: Classes and Methods - Answers to Exercise 8

### 4.4 Java Packaging

These exercises evaluate students' comprehension of packaging classes, organizing packaged classes into the proper folders, and defining the javac compile and java execution commands, which include classpath and output folders. When class packages are presented in a class diagram, ChatGPT is unable to determine the package for the class. In some exercises, the Java source file is provided, and we input the source code into ChatGPT. Despite this, the packaging is not accurately recognized. An example of this exercise is provided below.

*"Exercise 9 - Place the AnimalFarmTest.java in the proper sub-directories in the sourceFiles folder."*



```
package farming.app;
import farming.*;

public class AnimalFarmTest {
    public static void main(String[] args) {
        AnimalFarm farm = new AnimalFarm();
        farm.makeNoise();
    }
}
```

Listing 9: Packaging Sample

With both the exercise 9 instructions and Listing 9, ChatGPT recommends AnimalFarmTest.java to be placed in sourceFiles\farming\AnimalFarmTest.java instead of sourceFiles\farming\app\AnimalFarmTest.java

### 4.5 Java Exceptions and I/O

These exercises require the student to implement exception classes and codes to handle exceptions. If the exception classes are shown in a class diagram, as shown below, ChatGPT is unable to decipher the package for class, similar to earlier. On the other hand, if the exercise contains existing codes that generate exceptions, ChatGPT can regenerate solutions to handle exceptions. Below is an example of the exercise, and Listing 10 shows a correct output from ChatGPT.

*Exercise 10 - The program crashes when an invalid input (e.g. "abc") is entered. Include exception handling into the program to make the program robust, i.e. inform the user of invalid inputs and prompt the user to try again.*

```
// Correct Answer from ChatGPT
import java.util.InputMismatchException;
import java.util.Scanner;
public class Adder {
    public static void main(String[] args) {
        Scanner sc = new Scanner(System.in);
        boolean valid = false;
        int num1 = 0, num2 = 0;
        while (!valid) {
            try {
                System.out.print("Enter num_1> ");
                num1 = sc.nextInt();

                System.out.println();
                System.out.print("Enter num_2> ");
                num2 = sc.nextInt();
                valid = true;
            } catch (InputMismatchException e) {
                System.out.println("Invalid input.
                _Please enter a valid number."
                );
                sc.next();
            }
        }
        int sum = num1 + num2;
        System.out.println();
        System.out.println("The answer is " + sum);
    }
}
```

**Listing 10: Exception - Answer to Exercise 10**

## 5 DISCUSSION ON THE EFFECTIVENESS OF CHATGPT

In this section, we summarise the findings from our analysis of ChatGPT's effectiveness in generating coding solutions to address our research question (RQ).

- (1) ChatGPT's capability to generate coding solutions positively impacts student learning. Students can obtain guidance on coding solutions, explore alternative methods to solve problems, and compare solutions for efficiency. Educators can instruct students on understanding ChatGPT's features and capabilities from an educational perspective.
- (2) ChatGPT serves as a valuable resource for students in code debugging. Although assistance from instructors and peers remains helpful, ChatGPT can augment students' self-directed learning abilities by allowing them to independently seek solutions and explore alternative methods. This support is especially beneficial outside of lesson hours or in classroom environments with a high student-to-teacher ratio and also provides timely assistance to the student.

- (3) ChatGPT-generated solutions may not always be accurate in initial attempts, requiring students to critically evaluate the provided answers. As noted by ChatGPT, it may produce inaccurate information about people, places, or facts. Consequently, students must still rely on their subject knowledge to make their assessments. This evaluation process can contribute to their learning experience. As ChatGPT learns from its mistakes, its performance can improve over time, offering additional benefits to students.
- (4) As ChatGPT primarily supports text-based inputs, it is less effective for exercises that necessitate the interpretation of API documentation and UML diagrams, such as class and sequence diagrams. Educators can leverage non-text instructions to minimize the likelihood of students copying answers directly from ChatGPT, thus maintaining the integrity of the learning process. We do acknowledge with the advancement in AI technology, non-text inputs can be used. Together with the usage of prompt engineering, these exercises with diagrams can be handled in the future.
- (5) The ChatGPT-generated solutions provide a promising avenue for students to learn asynchronously. We acknowledge this benefit requires students' validation as part of future work.

## 6 CONCLUSION

The popularity of ChatGPT is growing quickly in recent months. In education, ChatGPT can potentially disrupt how students learn and how educators assess students. With the ChatGPT interactive mode, students can submit their exercise instructions for ChatGPT to answer. In this paper, we evaluate the effectiveness of ChatGPT's ability to generate coding solutions in an undergraduate Java programming course.

Using our course coding exercise instructions as input, ChatGPT generated coding solutions which we evaluate against the expected output. Our evaluation indicates that ChatGPT-generated solutions are effective when the instructions are clear and straightforward. ChatGPT generates partially incorrect solutions when there are complex instructions (e.g. producing a sophisticated representation as output). Due to its support for only text-based inputs, ChatGPT does not yield much benefit to exercises requiring students to interpret API documentation and UML diagrams.

Our analysis shows that ChatGPT can be an effective tool for students to learn asynchronously. They can overcome programming obstacles and learn alternative and more efficient ways to write codes. On the other hand, educators need to be aware of ChatGPT features and develop exercises as well as assessments that minimise students' chance of copying solutions from ChatGPT.

Our future plan involves integrating ChatGPT into our existing video-based programming tutorials [16], allowing students to receive timely guidance on programming challenges such as coding exceptions or logic issues. Additionally, we aim to utilize ChatGPT for summarizing assistance provided by comments in the StackOverflow postings, enabling students to quickly find help on specific topics.

## REFERENCES

- [1] AlphaCode. 2023. *AlphaCode data materials*. <https://zenodo.org/record/6975437#.Y8gBzy8RppQ> Accessed on 15.01.2023.
- [2] Lingfeng Bao, Zhenchang Xing, Xin Xia, and David Lo. 2018. VT-Revolution: Interactive programming video tutorial authoring and watching system. *IEEE Transactions on Software Engineering* 45, 8 (2018), 823–838.
- [3] Davide Castelvetti. 2022. Are ChatGPT and AlphaCode going to replace programmers? *Nature* (2022).
- [4] DeepMind. 2023. *DeepMind CodeContests dataset*. [https://github.com/deepmind/code\\_contests](https://github.com/deepmind/code_contests) Accessed on 15.01.2023.
- [5] GitHub. 2023. *GitHub Co-Pilot*. <https://github.com/features/copilot> Accessed on 15.01.2023.
- [6] Mubin Ul Haque, Isuru Dharmadasa, Zarrin Tasnim Sworna, Roshan Namal Rajapakse, and Hussain Ahmad. 2022. "I think this is the most disruptive technology": Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. *arXiv preprint arXiv:2212.05856* (2022).
- [7] Hieke Keuning, Bastiaan Heeren, and Johan Jeuring. 2019. How teachers would help students to improve their code. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*. 119–125.
- [8] Ada S Kim and Amy J Ko. 2017. A pedagogical analysis of online coding tutorials. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. 321–326.
- [9] Yujia Li, David Choi, Junyoung Chung, Nate Kushman, Julian Schrittwieser, Rémi Leblond, Tom Eccles, James Keeling, Felix Gimeno, Agustin Dal Lago, et al. 2022. Competition-level code generation with alphacode. *Science* 378, 6624 (2022), 1092–1097.
- [10] Zohar Manna and Richard J Waldinger. 1971. Toward automatic program synthesis. *Commun. ACM* 14, 3 (1971), 151–165.
- [11] Laura Meckler and Pranshu Verma. 2022. *Teachers are on alert for inevitable cheating after release of ChatGPT*. <https://www.washingtonpost.com/education/2022/12/28/chatbot-cheating-ai-chatbotgpt-teachers/> Accessed on 15.01.2023.
- [12] Abdallah Mohamed. 2020. Evaluating the Effectiveness of Flipped Teaching in a Mixed-Ability CS1 Course. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education (Trondheim, Norway) (ITiCSE '20)*. Association for Computing Machinery, New York, NY, USA, 452–458. <https://doi.org/10.1145/3341525.3387395>
- [13] OpenAI. 2023. *ChatGPT: Optimizing Language Models for Dialogue*. <https://openai.com/blog/chatgpt/> Accessed on 15.01.2023.
- [14] Oracle. 2023. *Java Foundations Exam Number: 1Z0-811*. [https://education.oracle.com/java-foundations/pexam\\_1Z0-811](https://education.oracle.com/java-foundations/pexam_1Z0-811) Accessed on 15.01.2023.
- [15] Eng Lieh Ouh and Benjamin Kok Siew Gan. 2022. Designing flipped learning activities for beginner programming course.. In *Proceedings of 28th AMCIS: Innovative Research Informing Practice*. Minneapolis, USA.
- [16] Eng Lieh Ouh, Benjamin Kok Siew Gan, and David Lo. 2022. ITSS: Interactive Web-Based Authoring and Playback Integrated Environment for Programming Tutorials. In *Proceedings of the ACM/IEEE 44th International Conference on Software Engineering: Software Engineering Education and Training (Pittsburgh, Pennsylvania) (ICSE-SEET '22)*. Association for Computing Machinery, New York, NY, USA, 158–164. <https://doi.org/10.1145/3510456.3514142>
- [17] Gina Sprint and Erik Fox. 2020. Improving student study choices in CS1 with gamification and flipped classrooms. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 773–779.
- [18] Teo Susnjak. 2022. ChatGPT: The End of Online Exam Integrity? *arXiv preprint arXiv:2212.09292* (2022).