# Solving Computer Science 2 tasks: Students' Reflections on the Use of ChatGPT

Inés Friss de Kereki Facultad de Ingeniería Universidad ORT Uruguay Montevideo, Uruguay kereki\_i@ort.edu.uy

*Abstract*— Computer Science II is a subject in the 2nd semester of the Bachelor's and Systems Engineering, Electrical, Electronics, and Telecommunications courses. As part of the evaluation, four tasks are included in the course to be performed individually by each student. From the beginning of the course, the use of Artificial Intelligence tools was introduced in different class activities, particularly ChatGPT. In this work, the activities are described and the solutions to the tasks presented by the students are evaluated. The surveys conducted with the students regarding their experience with these tools are examined.

Keywords—Computer Science 2, Programming, Artificial Intelligence, ChatGPT

# I. INTRODUCTION

Artificial Intelligence (AI) tools can be applied to various areas, in particular to Programming and its teaching.

Computer Science II is a second semester subject in the Bachelor's and Systems Engineering, Electronics, Electrical and Telecommunications courses at the Universidad ORT Uruguay.

In this study, 52 random students were selected and divided into two class groups. In the course taught in the second semester of 2023, the use of AI tools such as ChatGPT [1] was incorporated from the first weeks in both groups. ChatGPT is a model trained to follow an instruction in a prompt and provide a detailed response. Its use was demonstrated in class.

As part of the requirements to pass the course, the individual resolution of four exercises on the Hackerrank platform is included [2]. We assessed the outcomes of their assignments and analyzed their responses to a survey.

The article is structured as follows: it begins by providing an overview of the AI context, followed by a detailed description of the Computer Science II course. Subsequently, the experimentation is presented. It outlines the utilization of AI tools within the course and the tasks, which solutions were scrutinized taking various factors into account. Additionally, survey results are presented. Finally, the article presents conclusions and recommendations.

# II. OVERVIEW OF ARTIFICIAL INTELLIGENCE

ChatGPT is a model trained to follow an instruction in a prompt and provide a detailed response [1]. It is a large language model designed to facilitate natural and conversational interactions between individuals and computers. "GPT" stands for "Generative Pre-trained Transformer," representing a family of natural language models developed by OpenAI. This type of AI is often referred to as generative AI due to its capability to produce original and creative outputs [3]. Ismael Garrido Facultad de Ingeniería Universidad ORT Uruguay Montevideo, Uruguay ismael.garrido@fi365.ort.edu.uy

Stepanenko and Stupak [4] refer that the use of AI is facilitating the world of education and learning. Yu [5] highlights that the use of ChatGPT comes with significant advantages in enhancing learning efficiency and promoting communication. However, it also carries negative impacts and potential risks, such as completing academic tasks that could lead to behaviors like copying or academic dishonesty [5]. It is important to teach how to use AI tools responsibly [4], they indicate that institutions worldwide have had controversial approaches, ranging from restricting the use of ChatGPT to adopting and using it with students. At our university [6], educators were encouraged to use it, with instructional materials on how to integrate in.

Given its capacity to generate and evaluate information, ChatGPT can assume diverse roles within teaching and learning processes. Examples include serving as a co-designer, study buddy, and collaboration coach [3]. Welsh [7] points out that the "end of classical programming is coming". This refers to traditional programming courses where the objective is to translate an idea into a program manually written by an individual in languages like Java or Python. He suggests that the concept of manually writing a program will be replaced by AI systems that are trained rather than explicitly programmed. People will take on a supervisory role in this context.

Yu [5] points out that university students should learn how to leverage available resources and tools to efficiently complete tasks, thereby enhancing their opportunities for development. Ajlouni et al refer to that 73% of 623 undergraduates in its study, recognize the potential of ChatGPT to facilitate the learning process [8].

Prasad et al [9] refer that there is limited knowledge regarding whether and how students employ large language models to automatically create programs in response to textual prompts. As noted by Allam et al [10], further studies are needed to better understand how AI may improve the teaching and learning processes. In this regard, this work aims to contribute to visualizing how individuals perceive and utilize AI tools.

# III. COMPUTER SCIENCE II

The objectives of Computer Science II are to continue training in the area of programming, with strong emphasis on object-oriented programming and more advanced algorithms, and to develop object-oriented applications.

# A. Course syllabus and description

The main topics are: operators and statements, manipulation of array and matrix, classes, relationships between classes, graphic interface design, files, and exceptions. The language is Java, and the development environment is NetBeans [11]. The course is conducted inperson and consists of 4 hours of theoretical instruction per week in a classroom, along with 2 hours of practical lab sessions with another instructor. The course instructors are experienced. The course follows a flipped classroom format: "which is traditionally done in class is now done at home, and that which is traditionally done as homework is now completed in class" [12]. Prior to each class, students are required to watch specific short videos or review some materials, and the in-class sessions are structured around these materials. On the Moodle-based [13] course website, there are videos covering theoretical topics, the course textbook, exercises with solutions, and other materials. All classes are recorded using Panopto [14] and are available on the website within minutes of each class's conclusion.

# B. Evaluation

The course is passed by achieving 70% of the 100 available points and meeting the minimum requirements. It consists of four individual assignments referred to as tasks (each worth 2, 2, 3, and 3 points, without minimum requirements), 2 long-term team projects (worth 20 and 25 points each, with minimum requirements of 0 and 12 points, respectively), and an individual midterm exam (45 points with a minimum requirement of 20 points). The midterm is written manually on paper, without the use of a computer or additional materials. If the sum of the points exceeds or equals 86% of the total, there is no need to take the final exam. In the case of scoring between 70% and 85%, the right to take the final exam is earned. Scoring below 70% requires retaking the course.

In particular, the tasks are proposed in the first weeks of the course with the aim of fostering early engagement with the subject and clearly establishing the rules of the course. This includes guidelines for collaborative work (when allowed), and, or late submissions (which are not accepted). There is a two-week window provided for the completion of each task. They can be tested as many times as desired on Hackerrank [2], and for submission, the code must be uploaded to the course's platform to enable tracking of the student's progress.

Task proposals are selected and, or, adapted from websites such Hackerearth [15] or CodeChef [16]. In previous courses, tasks with similar format and complexity were included. Previously, points were awarded proportionally to the number of correctly solved cases. Starting from the 2023 course, it was decided that, given the availability of AI tools, points would only be awarded for the successful resolution of all test cases. Solution quality is not scored, but recommendations on how to improve style, structure, or other relevant observations are provided by the teachers.

Each of the tasks is individually tested by the instructor, and MOSS [17] is also used to identify potential cases of plagiarism. Students are informed at the beginning of the course about the use of this tool.

#### IV. EXPERIMENTATION

Two random groups of students were considered (Group "A"- morning group: 35 students, Group "B"-afternoon group: 17 students). The experimentation in both groups consisted of presenting ChatGPT, and then conducting a subsequent analysis of the tasks and a survey. The tasks will be compared with previous courses were ChatGPT was not available.

#### A. Presentation of AI tools

In the course, the use of ChatGPT version 3.5 was particularly shown in the first weeks. Only ChatGPT 3.5 was utilized (excluding later versions) due to its free availability. The methodology was, in some classes, an exercise was proposed and, after that, this tool was asked to solve the same problem, analyzing what was obtained. The need to provide context was discussed in class (example: "I am a 2nd semester Computer Science student and I am learning Java"), describing the preferred/desired outcome (example: "show me the code with comments"), including examples, among others points, as indicated by ChatGPT best practices [18]. Different roles of using AI were also discussed, for example, as a tutor to obtain explanations, as a colleague to find errors or to generate code. Some examples were presented too with Bard [19]. Bard [20] is an experimental conversational AI service, provided by Google.

#### B. Tasks

As presented, there are four tasks in the course. In all cases, test data is not displayed to the students.

Task 1 was focused on array manipulation. In brief, the task is as follows: 'You have an array of length N. You want to calculate and display the maximum number of numbers that contain the same digit. For example: '2 27 92 7777777.' The digit '2' appears in 3 numbers, the digit '7' appears in 2 numbers, and the digit '9' appears in 1 number. The maximum number of numbers that contain the same digit is 3. There are no leading zeros (e.g., 01). You receive N, which is the size of the array, and then the elements. Display the requested value."

The second task (related to string manipulation) was: "Given a sentence containing words separated by spaces, display the first longest word that can be formed by combining two consecutive words in the string and is lexicographically smaller (ignoring case). For example, if 'z y x' are received, 'yx' should be displayed".

Tasks 3 and 4 were related to matrix manipulation. Task 3 was: "In a certain algorithm, a matrix of positive numbers with N rows and N columns is received, and as a first step, for each row, the smallest element in that row is found and subtracted from each element in the row. Then, for each column, the smallest element in that column is determined, and it is subtracted from each element in the column. The program will receive two matrices and must verify if applying the algorithm to the first results in the second one.".

Task 4 was: "A company conducted a coding test to hire candidates. N candidates participated in the test, and each of them faced the same M problems. The results are stored in an NxM matrix where each position contains the candidate's result (row) on the problem (column) using 'N' (not solved), 'P' (partially solved), or 'C' (completely solved). To pass the test, each candidate must either solve a) X or more problems completely, or b) solve (X-1) problems completely and Y or more problems partially. As a result, display a line containing N integers. The i-th integer should be 1 if the i-th candidate has passed the test; otherwise, it should be 0. Also, indicate on the next line whether there are two or more problems with similar results, considering them similar if they have the same total count of 'C,' 'P,' and 'N'''.

The task statement was uploaded separately to both ChatGPT [1] and Bard [19] by the teachers, and the generated code was tested on Hackerrank [3]. As context, each tool was informed that "I am a 2nd-semester student, solve this exercise in Java.". Table I shows the results obtained by applying the code directly, without making any improvements or corrections. In other words, relying solely on ChatGPT's first solution is unlikely to fully solve the assignment. These results reinforce the idea that one must cover 100% of the cases to obtain the score of the task (not partially as previous courses). The student should either create the code or work on the solution obtained from ChatGPT or other sources (for example, StackOverflow [21]).

TABLE I. SOLUTIONS OF THE TOOLS

Task	ChatGPT	BARD		
1	Passes 1 out of 21 cases	Passes 1out of 21 cases		
2	Pass 10 out of 20 cases	Pass 8 out of 20 cases		
3	Pass 13 out of 27 cases	Pass 15 out of 27 cases		
4	Pass 9 out of 21 cases	The code does not compile (skips import among other errors)		

#### C. Analysis of solutions

Each of the submitted solutions was analyzed to identify patterns of resolution. The code was reviewed by two experienced teachers to evaluate whether each task is likely a typical task for a second-semester student or not, if it includes aspects that could be suggested by AI environments or other tools. For example, elements considered "typical" include style (mnemonic variables, use of braces, indentation, comments), and the use of variables names in Spanish. "Atypical" solutions include the use of "break" and "continue", and, or solutions that are clearly different from the majority of the solutions presented and what is typically expected at this stage of the course. Additionally, elements not covered in the course were examined. For example, the solution to Task 1 generated by ChatGPT, as shown in Fig. 1, includes advanced elements not covered at that point in the course, such as HashMap. This solution is classified as atypical.

This situation of failing to solve simple problems and using elements or constructions that novice programmers may not be familiar with is mentioned by Wermelinger in relation to Copilot, an artificial intelligence assistant [22]. Also, as Allam et al [10] refer, it might be possible for ChatGPT to provide accurate answers, but also results might contain errors. Ashraf and Imam [23] mention in their study that ChatGPT provided 49% of correct answers when coding questions were requested.

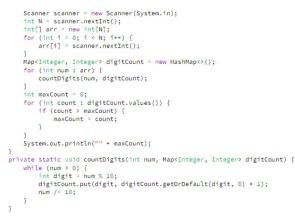


Fig. 1. ChatGPT solution of Task 1.

In Task 1, out of 50 submissions, in 18 cases (36%), there were elements not covered in class or with a coding style that was clearly different from what was taught in the course. For example, a typical solution for Task 1 could be like the one presented in Fig. 2 by one student. This solution can be classified as "typical" because it follows a coding style and notation similar to what is taught in the course.

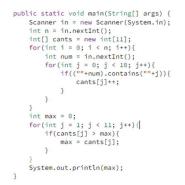


Fig. 2. Typical solution for Task 1.

In Fig. 3, part of the student's code is shown, where they load all the numbers into an array (instead of reading and processing each one, which was the recommended approach discussed in the class), and they use data structures like Set (initializing it with an object of the HashSet class). This element is not introduced in class and adds unnecessary complexity. This is classified as "atypical".



Fig. 3. Atypical solution of Task 1

In Task 2, there were 9 "atypical" cases out of 52. In Task 3, 18 cases out of 52, and in Task 4, 17 out of 50. Approximately 30% of the total solutions had "atypical" aspects. In a similar course from the previous year when ChatGPT was no available (but other sources such as [21] were), considering 50 randomly selected students, we observed 6 atypical cases in Task 1, 7 cases in Task 2, 6 cases in Task 3, and 5 cases in Task 4 (see Table II).

TABLE II. TASKS: TYPICAL AND ATYPICAL

Task	2022 typical	2022 atypical	2023 typical	2023 atypical
1	44 of 50	6 of 50	32 of 50	18 of 50
2	43 of 50	7 of 50	43 of 52	9 of 52
3	44 of 50	6 of 50	34 of 52	18 of 52
4	45 of 50	5 of 50	33 of 50	17 of 50
Total	176 of 200	24 of 200	142 of 204	62 of 204
	(88%)	(12%)	(70%)	(30%)

In total, there were 24 cases out of 200 tasks, which represents 12%. This indicates an increase in 'atypical' solutions in this course where the use of ChatGPT was encouraged. Considering binomial test and the null hypothesis that the rate of atypical tasks is the same in both semesters, and the alternative hypothesis that there is a difference, with alpha=0.05, the null hypothesis is rejected, and the difference is statistically significant.

# D. MOSS

The plagiarism detection program MOSS [17] was run with all task submissions (from both selected groups and all other students), and several cases of high similarity were detected (considering students of the selected groups: Task 1: 2 students, Task 2: 0 students, Taks 3: 0 students, Task 4: 5 students). The MOSS indicates similarity percentages. Cases with high percentages, significantly surpassing the others, are considered problematic. In addition to speaking with the students to analyze and understand the situation, they were informed about the seriousness of the matter and advised that it should not be repeated. In these cases, a grade of 0 was assigned and the task was disqualified.

In all of these cases, conversations were held with the students, and they indicated that they either worked on the tasks together or one completed the work and shared it with the other. Only one case was attributed by the student to the use of ChatGPT.

# E. Results of tasks

Table III displays the number of correct tasks (i.e. those that passed 100% of the cases). It also includes tasks with errors, disqualified tasks, and tasks that were not submitted.

TABLE III. TASKS RESULTS

	Results				
Task	ОК	With mistakes	Disqualified	Not submitted	Total students
1	44	4	2	2	52
2	50	2	0	0	52
3	50	2	0	0	52
4	43	2	5	2	52

Considering all the tasks, 7 students from group B and 26 from group "A" achieved the maximum score (10 points), which is 33 out of 52, or 63%. In a similar course in 2022, within another randomly selected group of 40 students, 17 students scored the maximum points (43%). While in previous courses points were awarded even if the solution was not complete, considering only the cases that achieved the maximum score among all tasks reveals an improvement in the results.

#### F. Survey

In order to analyze the students' perspective regarding the use of AI to solve tasks and their learning of programming, a survey was conducted. 46 students out of 52 responded. Among other aspects, they were asked about the frequency of use of AI for programming. 65% of the students (30 students) indicate "weekly" and 13% "daily" (6 students). That is, 78% of the students use the tools frequently on a daily or weekly basis. This would indicate that AI has become an important part of the work routine of the majority of students surveyed in the context of programming. It could be interpreted that its use would be useful and relevant.

They were asked about different roles for using AI, such as a tutor, colleague, code generation, among others, with open-ended responses, meaning they could indicate various roles or new options. 83% of the responses indicated the role of a tutor, 63% mentioned the role of a colleague, and 40% mentioned the role for code generation. Other responses included "to provide context to the problem," "to simplify the problem for me" and, "other" (23%).

When asked: "If you used ChatGPT for course activities, how was your experience?", the majority (94%) indicated a value of 3 or more, on a scale of 5 (1: Very bad, 5: Excellent)

(see Fig. 4). This indicates that the ChatGPT tool has been widely perceived as a useful and effective experience by students who used it in their course activities.

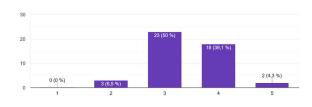
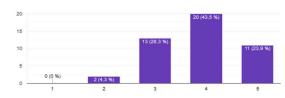
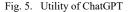


Fig. 4. Experience

For the question "How would you rate the utility of ChatGPT as a resource to support your programming studies?" on a scale from 1 to 5, where 1 is not very useful and 5 is very useful. Out of the 46 responses, 20 indicated that ChatGPT is "useful" (4 on the scale), and 11 rated it as "very useful" (5 on the scale) for supporting their programming studies (see Fig. 5).





The majority of students who responded to the survey either answered "no" or expressed uncertainty when asked, "Do you believe that using ChatGPT in your programming studies has improved your problem-solving skills?" Seventeen students indicated "yes" (37%), 21 students chose "I don't know" (46%), and seven students answered "no" (15%). See Fig. 6. One student does not use ChatGPT.

It can be interpreted that most students are unsure or do not believe that the use of ChatGPT has improved their problemsolving skills in their programming studies. It is important to note that the uncertainty in the "I don't know" response represents a significant proportion, suggesting that students may not have a clear perception of ChatGPT's impact on their skills. This could be due to various factors, such as a lack of specific feedback or the need for a more detailed assessment.

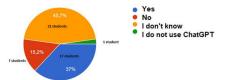


Fig. 6. Improvement of problem-solving skills

They were also asked to provide suggestions for advising a peer on how to use ChatGPT. Some suggestions emphasize the importance of being specific in questions and seeking clarification, while others caution against blind trust in the tool's responses. Many recommend using ChatGPT as a learning aid rather than a direct problem solver and stress the need to verify and validate the information provided. Independent thinking and creativity are encouraged, and the tool is suggested for understanding concepts, checking for errors, and obtaining examples related to class content. Overall, the guidance revolves around responsible and strategic utilization of ChatGPT as a valuable learning resource while remaining cautious and independent in problem-solving approaches. To the question: "What precautions do you take when using ChatGPT to ensure that the answers or solutions are correct and safe?", the responses highlight various strategies and approaches students take to interact with ChatGPT, which include validating results, learning, and understanding, and seeking information from diverse sources when in doubt, as an independent learner. Here's a summary of the responses categorized with examples:

- Validation and Verification. Examples: "Checking and verifying the results", "Running the code", "Asking about edge cases", "Comparing the answers with the course book."
- Understanding and Learning. Examples: "Requesting explanations when something is unclear.", "Requesting a change in strategy if the explanation is not understood", "Learning and understanding the solution before applying it", "Talking to ChatGPT in English for more information".
- Independent Learning. Examples: "Searching in multiple sources", "Seeking multiple alternatives", "Looking for information on other internet sources".

Finally, each student was asked to briefly describe how they would define ChatGPT. The most common response was "useful" (mentioned thirteen times), and "fast" was cited in 6 responses. They can be grouped into two categories. These categories reflect a range of opinions about ChatGPT, from its efficiency and, effectiveness, to concerns about its ambiguity and security, along with some mixed opinions. Those categories (with examples) are:

- Efficiency and Effectiveness. Examples: "Fast and effective", "Efficient, useful", "Fast and good", "Quick and efficient", "Useful, effective", "Timesaving", "useful", "fast".
- Ambiguity, security, and mixed opinions. Examples: "pseudo-help", "Questionable support", "Useful, but flawed.", "A double-edged sword", "Versatile and insecure", "efficient, but uncertain".

# V. CONCLUSIONS

This study describes an experiment conducted with 52 students in the Computer Science II course. The experiment involved introducing and discussing the use of ChatGPT at the beginning of the course, analyzing task resolutions to identify patterns, and collecting student responses to a survey regarding various aspects of their experience with ChatGPT.

From the analysis of the tasks, a statistically significant change in coding style and proposed solutions was observed, indicating influences from ChatGPT and other tools and sources. The similarities detected by MOSS corresponded to invalid situations, namely, students who worked collectively instead of individually and/or students who shared solutions, and not due to the use of AI tools. An improvement in the task results is observed in this course. Considering only the students who achieved the maximum score in all tasks, this course has a higher proportion of such students compared to the previous one.

Students were asked about their frequency of use (most indicated weekly or daily), their roles, with tutor and colleague roles standing out. The experience of using ChatGPT for course activities was positively rated. The utility of ChatGPT for programming study was also highlighted as useful or very useful. On one hand, most students indicate that they are not sure or do not believe that the use of ChatGPT improves their programming or problem-solving skills. But on the other hand, it is interesting that many students emphasize the importance of being specific when querying and validating answers. Students defined ChatGPT as "useful" and "fast," but also with uncertainty.

In summary, influences probably linked to the use ChatGPT are observed in the tasks. From the analysis of the surveys, it can be inferred that students tend to use ChatGPT critically, verifying the answers, comparing them with their own knowledge and seeking additional information in case of doubt. They also perform testing and verify the validity of the solutions provided. In general, they do not blindly trust ChatGPT answers and use it as a complementary tool to get suggestions or explanations. As recommendations for the next course, it is suggested to expand the experience of using ChatGPT and similar tools in the classes with more diverse activities, such as "Solve this exercise entirely using only ChatGPT.". Additionally, expanding the analysis to team projects is recommended.

#### REFERENCES

- [1] ChatGPT, https://openai.com/blog/chatgpt
- [2] Hackerrank, http://hackerrank.com
- [3] Sabzalieva, E. and Valentini, A., ChatGPT and artificial intelligence in higher education: quick start guide. UNESCO Int. Institute for Higher Education in Latin America and the Caribbean, 2023
- [4] Stepanenko, O. and Stupak, O., The use of GPT Chat Among Students in Ukrainian Universities. Scientific Journal of Polonia University, 58 (2023), 3. https://doi.org/10.23856/5828.
- [5] Yu, H., Reflection on whether ChatGPT should be banned by academia from the perspective of education and teaching. Frontiers in Psichology, doi 10.3389/fpsyg.2023.1181712, 2023
- [6] CAES, Enseñanza con el apoyo de ChatGPT, CAES. https://ie.ort.edu.uy/innovaportal/file/137511/1/ensenanza-con-elapoyo-de-chatgpt.pdf
- [7] Welsh, M., The End of Programming. Comm. of the ACM, Jan 2023 Vol 66 No. 1
- [8] Ajlouni, A., Abd-Alkareem, F., and Abdallah Salem A., Students' Attitudes Towards Using ChatGPT as a Learning Tool: The Case of the University of Jordan. Int. Journal of Interactive Mobile Technologies, 17(18), 2023, https://doi.org/10.3991/ijim.v17i18.41753
- [9] Prasad, S., Greenman, B., Nelson, T. and Krishnamurthi, S., Generating Programs Trivially: Student Use of Large Language Models, CompEd 2023
- [10] Allam, H., Dempere, J., Akre, V., Parakash, D., Mazher, N., and Ahamed, J., Artificial Intelligence in Education: An Argument of Chat-GPT Use in Education. 2023 9<sup>th</sup>. Int Conf on Inf. Technology Trends. https://doi.org/ 10.1109/ITT59889.2023.10184267
- [11] NetBeans, https://netbeans.apache.org/front/main/
- [12] Bergmann, J., and Sams, A., Flip Your Classroom: Reach Every Student in Every Class Every Day. Washington DC: Int. Soc. for Technology in Education, 2012.
- [13] Moodle, https://moodle.org
- [14] Panopto, https://www.panopto.com
- [15] HackerEarth, https://www.hackerearth.com/
- [16] CodeChef, http://codechef.com
- [17] MOSS, https://theory.stanford.edu/~aiken/moss/
- [18] ChatGPT Best Practices, https://platform.openai.com/docs/guides/gptbest-practices
- [19] Bard, https://bard.google.com/chat
- [20] Bard (2), https://blog.google/technology/ai/bard-google-ai-search-updates/
- [21] StackOverflow, https://stackoverflow.com/
- [22] Wermelinger, M., Using GitHub Copilot to Solve Simple Programming Problems. SIGCSE 2023, Canada.
- [23] Ashraf, A., and Imam, A., ChatGPT's Use case for Software Engineers. 8<sup>th</sup> Int Conf on Comp in Eng. and Tech.. India, 2023. https://doi.org/10.1049/icp.2023.15