LEARNING A FIRST PROGRAMMING LANGUAGE WITH LLM TUTOR

Milić Vukojičić *,
[0009-0002-1218-5893]

Ivana Korica
[0009-0009-1918-347X]

British International School,
Belgrade, Serbia

Abstract:

The paper presents the use of Large Language Models (LLM) and Chatbots as tutors when students need to learn novel concepts related to programming. Traditional methods of teaching and learning new programming concepts involve the students and teachers exploring one idea at a time – sequence, selection, and iteration – followed by more complex concepts. Traditional methods include using textbooks, computers, presentations, etc., and the teacher has the central role in the classroom. LLMs can help make complex topics more accessible for teachers to teach and for students to explore. Passive lectures and frontal teaching methods are replaced by the interactive use of LLM/chatbots, where the teacher has a central role in checking the information produced by the LLM/chatbot and is the key mediator between students and the LLM. The study was conducted with 30 students (n=30) in two groups. One group used LLM/chatbot as the core resource for first-time learning the programming language (n=15), while the second group used standard teaching and learning methods (n=15). The results show a significant improvement for first-time learning the programming language with the help of LLM – students showed better results in the assessment, besides being happier with the experience and the learning process. The study also demonstrates potential problems with this method, such as time, resources, accessibility issues, and initial preparation.

Keywords:

Large Language Models, Chatbots, Secondary Education, Programming Languages, Artificial Intelligence.

INTRODUCTION

The integration of Artificial Intelligence (AI) into computer science education presents a transformative potential for teachers and students alike. Modern curricula that are based on critical thinking and problem-solving can benefit from tools like Large Language Models (LLM) and chatbots as an opportunity for students to learn critical thinking and problem-solving and for teachers to implement these fields in their subjects easily. Modern lower secondary education often struggles to equip students with fundamental programming concepts such as programming syntax, control structures (e.g. loops, conditional statements), and the skill to debug their code. This is mainly due to the complexity of the tasks and the limited time available for more serious practice in the lessons.

Correspondence:
Milić Vukojičić

e-mail:
vukojicic.milic@gmail.com

DOI: 10.15308/Sinteza-2025-331-336

The approach of using LLM to help students generate code can be seen in various papers. Some papers [1] agree that standard methods of learning how to code can be demanding because of the way that instructions are presented, often with robust textual explanations. Studies also suggest that tools like LLM can help us with the power of the Integrated Development Environment and make a significant impact when it comes to writing computer programs.

Other studies suggest [2] that LLM can be used as a programming assistant by helping students write better and more effective code. This means that students can write code with a better understanding of what they are writing, and it also suggests that students appreciate other parts of code as very valuable, for example, comments. We can also see that in the comprehensive study, there are many benefits of using chatbots in education [3], as the study highlights the benefits of using chatbots as a virtual assistant.

The impact of LLMs within the domain of undergraduate education, secondary and lower secondary education is still limited. Some of the results are not in favour of LLMs in the domain of higher education [4] in addition to the domain of lower secondary education [5]. We can see a negative correlation between using LLMs and lower grades, in addition to lower use of critical thinking in undergraduate education. Similarly, when referring to resources, we can also see a negative correlation between students' grades and the use of LLMs in lower secondary education. From the work mentioned above, we can see several potential advantages and disadvantages in regard to the extensive use of LLMs in the education system. This is why we believe that teachers, tutors, and professors must play a crucial role in the education system – they need to monitor the output of LLM software and guide students towards their proper use. (Figure 1).

The research suggests that educational institutions must promote the use of LLM models [6] in addition to encouraging the best practices. Chatbots and LLMs can be used in order to promote the exploration of novel ideas and new content, along with problem-solving skills and critical thinking. With the rapid movement to the area of LLMs, schools and educators should reconsider resizing and redefining their assessment standards and student evaluation standards.

This experiment will introduce the idea of using LLMs for learning a programming language for the first time, as students will explore concepts like sequence, selection, and iteration. All of the students chose Python as the first programming language, and LLMs will be used as tutors, while the teacher will fill the role of a mediator between LLMs and students.

2. EXPERIMENT METHODOLOGY

This experiment was based on observing and analysing the work of 30 lower secondary students, aged between 10 and 14, of different genders and a variety of backgrounds, with zero experience in textual programming languages. The students were organised into two groups. The first group used LLM/chatbot as a tutor, and the second group used standard methods of education such as textbooks, notebooks, slides, and teacher presentations. This paper will compare the results of both groups on the assessment. The assessment was constructed of 9 questions, with 3 questions from each category, starting from low-level question (1 point), medium-level question (2 points) to high-level question (3 points):

- writing sequences with Python programming language;
- using Python programming language to define selection (if-statement);
- defining iterations in Python programming language (for and while loops).

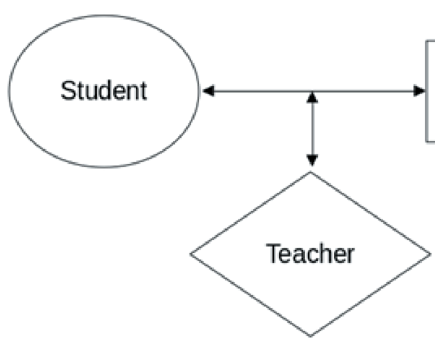

Figure 1. LLM, student and teacher relationship

Group 1 had the training on how to use LLM/chatbot in the domain of computing education. Students could choose between several LLM models: ChatGPT 4o-mini, Llama 3.3 70B, Claude 3 Haiku, and Mistral Small 3. The students also used the recommended prompting method: Persona, Aim, Recipients, Theme, and Structure, as suggested by Google and other researchers [7]. The different areas of prompt are defined as:

- Persona – Students will define who they are;
- Aim – Students will define a specific task for the prompt to solve;
- Recipients – they will define themselves because they are the recipient of the task;
- Theme – this is where LLM needs to set the voice, for example, to be more formal than usual;
- Structure – instruct the AI that it needs to behave as a tutor, with what it needs to explain and how it will ask questions about certain topics.

An example of a student's prompt: "I'm a Year 7 Computing student. Pretend that you are the Computer Science teacher. You want me to learn the Python programming language, you need to lead me to the key concepts like sequence, selection, and iteration, starting from the basic knowledge of Python language. After your explanation, you will need to ask me a question about the topic that you have explained. You will need to evaluate my answers. Each interaction should have an explanation of one concept and one question. After evaluation, you will teach me the next concept.". In Figure 2, we can see an example of the output of the LLM and the interaction with the user.

3. Results and discussion

Results are obtained from all 30 students after completing all 9 questions, 3 questions from each category, starting from low-level question (1 point), medium-level question (2 points) to high-level question (3 points) about the topics mentioned before. The data is shown in Table 1, with the results of Group 1 and Group 2.

In Table 2, we can see the results of both groups of students in the different question levels. We can see that group 1, which used an LLM/chatbot, has better results in questions about sequence and selection but lower results from iteration from low-level questions. In the medium-level question category, Group 2 had better results from questions related to sequence and lower performance with questions related to selection and iteration. In the domain of high-level questions, Group 1 achieved higher scores.

In Table 3, we can observe the lowest test score expressed in percentages from each group, which is 22% for Group 1 and 22.22% for Group 2. The highest score achieved in the test is 100% in both groups. The average score on the test for Group 1 is 71,48%, and for Group 2 is 65,18%.

From Table 1, Table 2, and Table 3, we can see that Group 1, the group that used an LLM/chatbot as a tutor, had better results on the test than Group 2, which used traditional methods of learning in the process of learning a textual programming language for the first time.

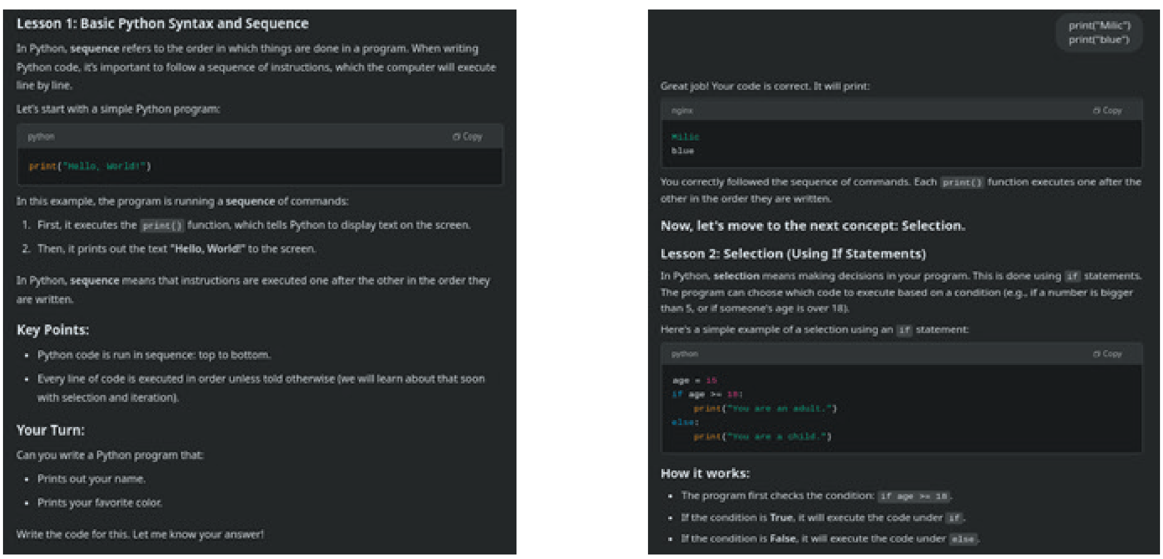

Figure 2. LLM output and the example of an interaction with the user

Table 1. Results of Students (S1- S15) separated into Group 1 -LLM (G1) and Group 2 – without LLM (G2), with different levels of question difficulty (low, medium, high) per each category (sequence, selection, iteration)

	Low -level question			**Medium-level question**			**High-level question**		
	Sequence	Selection	Iteration	Sequence	Selection	Iteration	Sequence	Selection	Iteration
S1 (G1)	0	0	0	1	1	1	1	0	0
S2 (G1)	1	1	1	2	2	2	3	3	3
S3 (G1)	1	1	1	2	2	2	3	3	3
S4 (G1)	1	1	0	2	1	1	3	0	1
S5 (G1)	1	1	1	2	0	1	3	2	2
S6 (G1)	1	0	0	2	1	1	3	2	3
S7 (G1)	1	0	0	2	0	0	3	1	2
S8 (G1)	1	1	1	2	2	1	3	3	1
S9 (G1)	1	1	0	1	2	2	2	1	1
S10 (G1)	1	0	1	2	0	0	3	1	1
S11 (G1)	1	1	1	2	2	2	3	3	3
S12 (G1)	1	1	0	2	2	1	3	1	1
S13 (G1)	1	1	0	2	2	2	3	2	2
S14 (G1)	1	1	1	2	2	2	2	3	2
S15 (G1)	1	0	1	2	1	1	2	2	2
S1 (G2)	1	0	0	2	1	1	3	1	0
S2 (G2)	1	1	0	2	1	0	2	1	1
S3 (G2)	0	1	1	2	2	2	3	2	3
S4 (G2)	1	0	1	2	1	2	3	3	3
S5 (G2)	1	1	1	2	2	1	3	2	1
S6 (G2)	1	1	1	2	2	2	3	3	3
S7 (G2)	1	0	1	2	0	0	3	1	1
S8 (G2)	1	1	1	2	1	0	3	2	0
S9 (G2)	1	0	0	2	0	0	1	0	0
S10 (G2)	1	0	0	2	1	1	2	2	2
S11 (G2)	1	0	1	2	2	2	2	1	2
S12 (G2)	1	0	0	2	1	2	1	1	2
S13 (G2)	1	1	1	2	2	1	3	2	1
S14 (G2)	1	1	1	2	1	1	2	2	2
S15 (G2)	0	1	0	2	1	0	3	1	1

Table 2. Comparison between results (Group 1 - G1, Group 2 – G2) of all questions

	Low -level question			**Medium-level question**			**High-level question**		
	Sequence	Selection	Iteration	Sequence	Selection	Iteration	Sequence	Selection	Iteration
Result (G1)	**14**	**10**	8	28	**20**	**19**	**40**	**27**	**27**
Result (G2)	13	8	**9**	**30**	18	15	37	24	22

Table 3. Test score criteria: Min, Max and Average (for all students) per group (Group 1 – G1, Group 2 -G2)

Test score criteria	%
Min (G1)	22
Min (G2)	22,22
Max (G1)	100
Max (G2)	100
Average (G1)	71,48
Average (G2)	65,18

Table 4. Overall satisfaction with learning (1-5) new concepts (sequence, selection, iteration)

	Overall satisfaction with learning (1-5) new concepts		
	Sequence	Selection	Iteration
S1 (G1)	5	5	5
S2 (G1)	5	4	5
S3 (G1)	5	5	5
S4 (G1)	4	4	4
S5 (G1)	5	5	5
S6 (G1)	5	5	5
S7 (G1)	1	1	1
S8 (G1)	5	5	5
S9 (G1)	3	4	5
S10 (G1)	1	1	1
S11 (G1)	5	5	5
S12 (G1)	4	5	5
S13 (G1)	5	2	2
S14 (G1)	3	3	3
S15 (G1)	5	3	3
S1 (G2)	2	2	2
S2 (G2)	4	2	2
S3 (G2)	3	3	3
S4 (G2)	4	4	4
S5 (G2)	4	4	4
S6 (G2)	5	5	5
S7 (G2)	5	5	3
S8 (G2)	3	3	3
S9 (G2)	1	0	0
S10 (G2)	1	0	0
S11 (G2)	1	0	1
S12 (G2)	1	0	0
S13 (G2)	1	1	1
S14 (G2)	1	1	1
S15 (G2)	0	1	0

Table 5. Overall satisfaction with learning score

	Overall satisfaction with learning score		
	Sequence	Selection	Iteration
Result (G1)	**64**	**61**	**64**
Result (G2)	56	51	**50**

From Table 4, we can see the overall satisfaction with learning, where each student rated their learning experience from 1 to 5 – 1 being "Not Satisfied at all" and 5 being "Completely Satisfied" related to the concepts of sequence, selection, and iteration. In Table 5, we can see the overall score for satisfaction, where the learning is better in group 1, which used LLM/chatbot as a tutor for learning new concepts.

4. CONCLUSION

In conclusion, this study provides fresh insights in the domain of using an LLM/chatbot as a tutor in the context of learning a new programming language. The study showed that there are significant differences between Group 1, which used LLM/chatbots as tutors and Group 2, which used traditional and ordinary methods of learning.

The average score of Group 1 is 71,48%, which is higher than the average score of Group 2, which is 65,18%. We can also see from the questionnaire that students find using an LLM/chatbot method more satisfying and that they had a better time and more fulfilment when we compared the results with those of Group 2.

The main problem that can occur is that student needs more time to adapt to a new method of learning and to the use of new tools in a meaningful way. Future work in the domain of LLMs and learning a programming language for the first time can be improved by looking at the methods for faster adaptation of students to their new learning environment and making students more independent in the domain of exploring new ideas, critical thinking, and problem-solving.

REFERENCES

[1] D. Nam, A. Macvean, V. Hellendoorn, B. Vasilescu, and B. Myers, "Using an LLM to help with code understanding," in *Proc. IEEE/ACM 46th Int. Conf. Softw. Eng. (ICSE)*, Lisbon, Portugal, Apr. 2024, pp. 97:1–97:13. doi: 10.1145/3597503.3639187.

[2] M. Vukojičić and J. Krstić, "ChatGPT in programming education: ChatGPT as a programming assistant," *InspirED Teachers' Voice*, vol. 2023, no. 1, pp. 7–13, 2023. [Online]. Available: https://www.teachers-voice.org/index.php/inspired/article/view/10

[3] C. K. Lo, "What is the impact of ChatGPT on education? A rapid review of the literature," *Educ. Sci.*, vol. 13, no. 4, p. 410, Apr. 2023. doi: 10.3390/educsci13040410.

[4] G. Jošt, V. Taneski, and S. Karakatič, "The impact of large language models on programming education and student learning outcomes," *Appl. Sci.*, vol. 14, no. 10, p. 4115, May 2024. doi: 10.3390/app14104115.

[5] M. Vukojičić, I. Korica, and M. Veinović, "The impact of LLM-based chatbots on secondary computing education," in Sinteza 2024 - *Int. Sci. Conf. Inf. Technol., Comput. Sci., Data Sci.*, Belgrade, Serbia: Singidunum Univ., 2024, pp. 31–37. doi: 10.15308/Sinteza-2024-31-37.

[6] M. M. Rahman and Y. Watanobe, "ChatGPT for education and research: Opportunities, threats, and strategies," *Appl. Sci.*, vol. 13, no. 9, p. 5783, May 2023. doi: 10.3390/app13095783.

[7] C. Chan, J. Cheng, X. Liu, Y. Yim, Y. Jiang, Z. Deng, H. Li, Y. Song, G. Y. Wong, and S. See, "Audience Persona Knowledge-Aligned Prompt Tuning Method for Online Debate," in P*roc. 27th Eur. Conf. Artif. Intell. (ECAI 2024)*, Santiago de Compostela, Spain, Oct. 2024, pp. 3851–3858. doi: 10.3233/FAIA240948.