Use of ChatBots in Programming Education: A Scoping Review

João Victor Barroso Alves[1], **Yanna Torres Gonçalves**[1], **Henrique Barbosa Silva**[1]

[1] Instituto Universidade Virtual
Universidade Federal do Ceará (UFC) – Fortaleza, CE – Brazil

{joaovba2002, yannatorres}@alu.ufc.br, henriquesilva@virtual.ufc.br

Abstract. *Proficiency in programming languages is essential across various sectors. As technology advances, AI-driven chatbots are increasingly recognized for their potential in education. This paper presents a scoping review of chatbots in programming education, examining existing research to identify benefits, challenges, and user perceptions. Findings reveal positive impacts on learning and widespread student adoption. However, gaps exist, such as limited focus on teachers' perspectives and issues with student engagement and accessibility.*

1. Introduction

In the modern era, proficiency in programming languages, bug-solving abilities, and effective code comprehension are essential skills not only for future IT professionals, but also for various sectors requiring logical problem-solving and task automation. As technological innovation drives our world forward, there is a growing expectation regarding the role of Artificial Intelligence (AI) in education, particularly through chatbots and generative AI. These technologies, capable of simulating human conversations naturally and coherently, are increasingly seen as valuable allies in enhancing programming education.

Chatbots, defined as computer programs that mimic intelligent human-like responses in text or voice conversations [Adamopoulou and Moussiades 2020], have gained significant prominence in contemporary digital society. Since 2016, interest in these virtual agents has surged across diverse applications—from educational contexts to customer support and e-commerce (see Figure 1). Notably, ChatGPT, launched by OpenAI in late November 2022, stands out as a leading example, leveraging the GPT-3.5 architecture trained using Reinforcement Learning with Human Feedback (RLHF) to excel in engaging conversations and providing responses that closely resemble human interactions in terms of coherence, naturalness, and contextual understanding[1].

Recent advancements in artificial intelligence, particularly in Generative AI technologies like ChatGPT, Mistral, and Gemini, show significant progress in creating natural and coherent human-like conversations. Generative AI, as defined by [Labadze et al. 2023], employs computational techniques to generate new content such as text, images, or audio based on existing data.

These technologies are viewed as having transformative potential in education, with the expectation of facilitating broader access to learning, personalized teaching experiences, and optimized methods to achieve specific educational objectives

[1] https://openai.com/blog/chatgpt/

DOI: 10.5753/sbie.2024.242473

[Zhang and Aslan 2021]. Noteworthy usage examples, as highlighted by [Vicari 2021], include intelligent tutoring systems capable of interpreting text and voice inputs while generating concise summaries.

In this context, chatbots leveraging Generative AI emerge as promising educational tools, enhancing the learning experience by harnessing machine learning algorithms to deliver tailored content that meets students' unique demands and capabilities. Additionally, these technologies can improve social benefits, such as supporting students with special needs and fostering collaboration among peers. Therefore, it is crucial to further explore the specific impacts of these tools, particularly in programming education, considering the distinctive challenges and requirements of this domain.

This scoping review aims to provide a comprehensive understanding of the current landscape regarding using chatbots in programming education. By examining existing applications and research, identifying potential risks and limitations, and assessing user perceptions, this study seeks to contribute valuable insights to the ongoing discourse on integrating AI technologies effectively into educational practices.

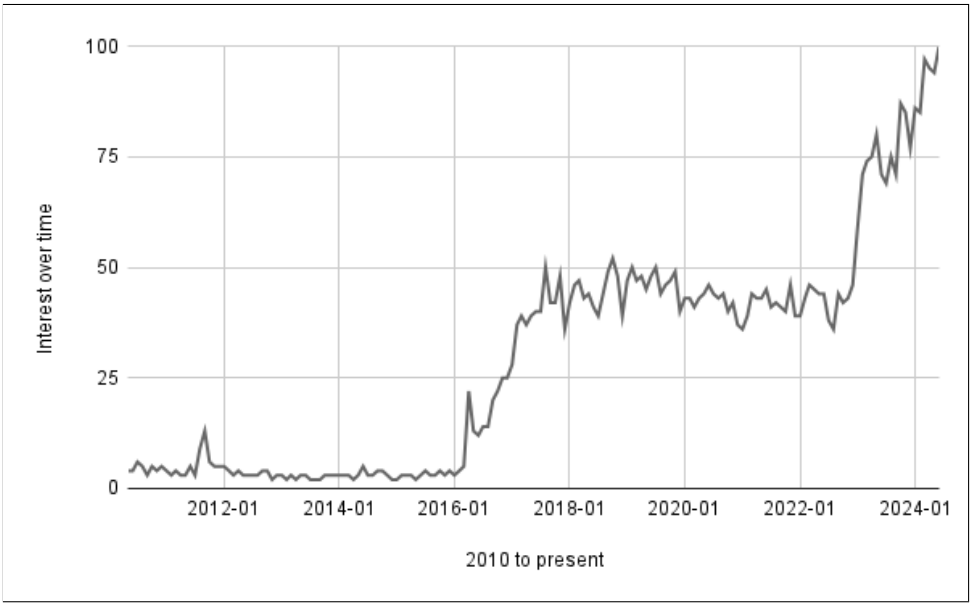

Figure 1. Interest overtime on the word "chatbot" worldwide. Data source: Google Trends.

2. Methods

Considering the recent nature of the topic of interest in this study, the use of chatbots with LLMs, a scoping review was conducted. A Scoping Review, or Systematic Mapping Study, can be defined as a method to provide an overview of a research area, assessing the existence and quantity of evidence on a certain topic [Kitchenham and Charters 2007]. To conduct this study, we followed the guidelines established by [Dermeval et al. 2020] and utilized the management tool *Parsifal*[2].

Primarily, the research questions (RQs) on which this study is focused were established. A total of five RQs were defined:

[2]https://parsif.al/

1. **RQ1.** In what context have chatbots been utilized in the teaching and learning process of programming?
2. **RQ2.** What educational benefits, if any, can be derived from using chatbots in the process of teaching and learning programming?
3. **RQ3.** What are the challenges and limitations, if any, of using chatbots in the process of teaching and learning programming?
4. **RQ4.** What social benefits, if any, arise from using chatbots in the process of teaching and learning programming?
5. **RQ5.** How widely accepted is this technology among teachers and students?

2.1. Search and Study Selection

The primary sources for study collection were IEEE *Xplore*[3] and ScienceDirect[4]. Additionally, Google Scholar[5] was utilized to identify relevant studies across various other databases. To find as many relevant studies as possible, the study employed a two-round search process to comprehensively gather relevant literature. Initially, a broad search string (chatbot AND programming AND (learning OR teaching OR education)) was used to identify foundational studies. In the second round, synonyms and related terms were included (conversational agents OR chat agents OR learning assistants and coding OR developing) to capture additional relevant literature. The search was limited to studies published from 2020 onwards to ensure inclusion of the most recent research.

To ensure the relevance and quality of the studies included in our review, we established a set of inclusion and exclusion criteria. Four inclusion criteria were defined and six exclusion criteria. These criteria were designed to filter the search results and select only those studies that directly pertain to our research objectives and theme. The detailed inclusion and exclusion criteria are presented in Table 1.

In total, 7,173 studies were retrieved through both rounds of searching. After removing duplicates and evaluating the abstracts, titles, and keywords, two reviewers independently analyzed the remaining full-text papers. Any discrepancies, such as disagreements over inclusion or exclusion criteria or concerns about the level of detail in an article, were discussed and resolved collaboratively by re-evaluating the texts together. This rigorous process resulted in the identification of 17 eligible studies for inclusion in this review. The final list of studies and their abbreviations is provided in Table 2. Additionally, the detailed selection process, including the handling of reviewer disagreements, is illustrated in Figure 2.

Given that this study is focused primarily on identifying the current state-of-the-art usage of chatbots as a tool in programming education, the quality assessment step was not executed. As [Dermeval et al. 2020] states, in scoping reviews, the quality assessment can be optional depending on the focus and objective of the review.

2.2. Data Analysis

To conduct the data analysis, specific questions were formulated for each research question, forming a data extraction form described in Table 3. During the review of the papers,

[3]https://ieeexplore.ieee.org/
[4]https://sciencedirect.com/
[5]https://scholar.google.com

Inclusion	Exclusion
Chatbots included in the study must use Generative AI/LLM	Any study that is not primary (e.g., scoping review, systematic review)
The study should be centered around teaching or learning programming	Any study that is not a paper (e.g., book, book chapter)
The study must be published in peer-reviewed journals or conferences	Any study that is not in English
The study must be a complete work (e.g., not a work in progress)	Any study that discusses the design, development, or deployment of a chatbot but does not include data related to students or teachers' usage of the tool
	Any study on other tools (e.g., wearable devices, voice assistants)
	Any study on teaching or learning of topics other than programming (e.g., deep learning, language learning)

Table 1. Inclusion and Exclusion Criteria

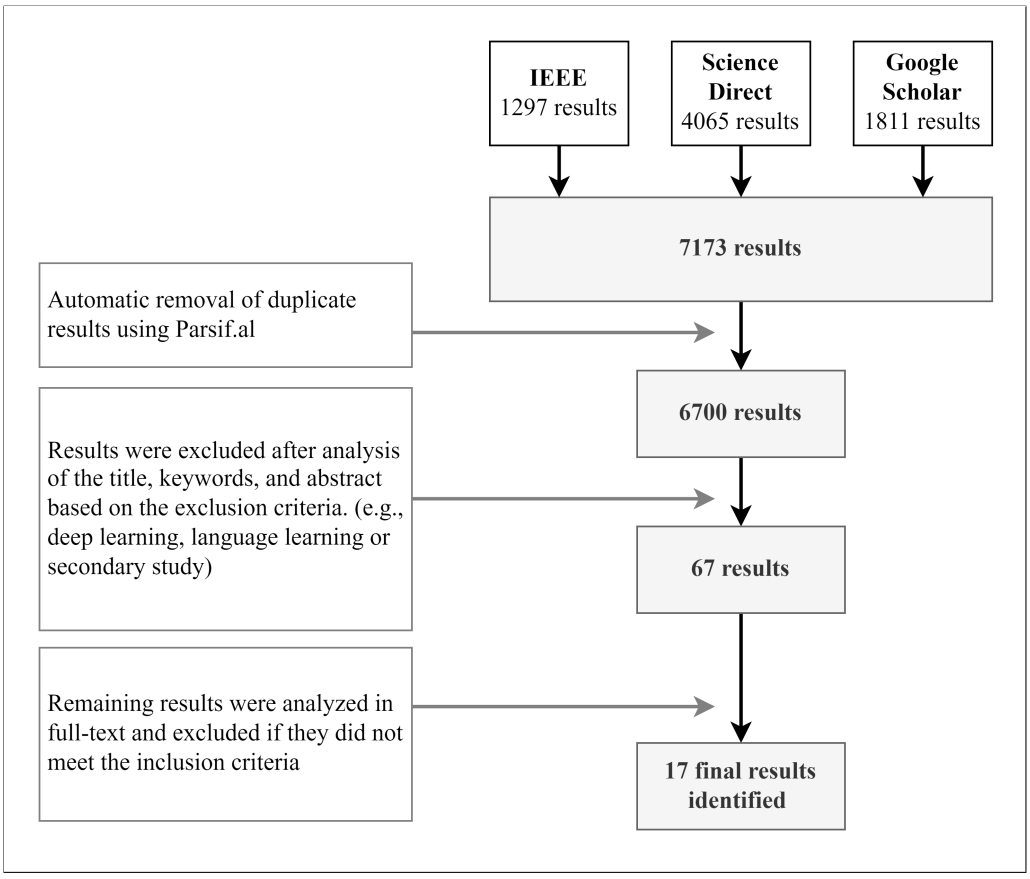

Figure 2. Study Retrieval and Selection Process

ID	Study	Reference
S01	Teach AI How to Code: Using Large Language Models as Teachable Agents for Programming Education	[Jin et al. 2024]
S02	Codeaid: Evaluating a classroom deployment of an LLM-based programming assistant that balances student and educator needs	[Kazemitabaar et al. 2024]
S03	Teaching CS50 with AI: leveraging generative artificial intelligence in computer science education	[Liu et al. 2024]
S04	The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation	[Yilmaz and Karaoglan Yilmaz 2023]
S05	Pair Programming Education Aided by ChatGPT	[Banić et al. 2023]
S06	Challenging the Confirmation Bias: Using ChatGPT as a Virtual Peer for Peer Instruction in Computer Programming Education	[Dos Santos and Cury 2023]
S07	Explicitly Introducing ChatGPT into First-year Programming Practice: Challenges and Impact	[Hu et al. 2023]
S08	Language AI in Programming: A Case Study of ChatGPT in Higher Education Using Natural Language Processing	[Padilla et al. 2023]
S09	Scaffolding Computational Thinking With ChatGPT	[Liao et al. 2024]
S10	Exploring the Role of AI Assistants in Computer Science Education: Methods, Implications, and Instructor Perspectives	[Wang et al. 2023]
S11	Using GPT-4 to Tutor Java Programming in Amharic	[Butgereit and Egu 2023]
S12	Investigating the Use of AI-Generated Exercises for Beginner and Intermediate Programming Courses: A ChatGPT Case Study	[Speth et al. 2023]
S13	Students' Experiences of Using ChatGPT in an Undergraduate Programming Course	[Haindl and Weinberger 2024]
S14	Assessing the Impact of ChatGPT in a PHP Programming Course	[Hajj and Sah 2023]
S15	Using ChatGPT Standard Prompt Engineering Techniques in Lesson Preparation: Role, Instructions and Seed-Word Prompts	[Spasić and Janković 2023]
S16	Toward AI-assisted Exercise Creation for First Course in Programming through Adversarial Examples of AI Models	[Chan et al. 2023]
S17	On ChatGPT: Perspectives from Software Engineering Students	[Hanifi et al. 2023]

Table 2. Included Studies

each author conducted an exploratory reading and extracted data according to the form.

#	Study Data	Relevant RQ
1	Authors, Year, Title	Overview
2	Study Source	Overview
4	How many participants	RQ1
5	Who participated in the research	RQ1
6	Education Level	RQ1
7	For who is the intended use	RQ1
8	Primary Goal	RQ1
9	Where was the research conducted	RQ1
10	Programming language	RQ1
11	Which LLM was used	RQ1
12	Was there an improvement in learning	RQ2
13	Was there an improvement in engagement	RQ2
14	Was there an improvement in feedback/support	RQ2
15	It anyway made the work of educators easier	RQ2
16	Were there other benefits	RQ2
17	Were there any technical challenges	RQ3
18	Were there any pedagogical challenges	RQ3
19	Were there any usage challenges	RQ3
20	Were there any other limitations	RQ3
21	Was there an improvement in inclusion and/or accessibility	RQ4
22	Was there an improvement in collaboration and/or interaction	RQ4
23	What were the teachers' and professors' perceptions	RQ5
24	What were the students' perceptions	RQ5

Table 3. Data Extraction Form

Subsequently, the extracted data were discussed in a group, resolving any disagreements at this stage.

To answer RQ1 — Investigating the contexts of chatbot use in teaching and learning programming — we analyzed descriptions of chatbot implementations across 17 studies, focusing on pertinent details. This included geographic data where research was conducted, participant demographics, purposes of use, intended users, utilized LLMs, and programming languages involved.

For RQ2 — Exploring educational benefits from chatbot use in teaching and learning programming — we assessed potential improvements in learning outcomes, enhanced student engagement, improved feedback and support mechanisms, and the facilitative impact on teachers' workload. We evaluated the presence of quantitative evidence supporting these benefits and insights gleaned from participant feedback.

To address RQ3 — Identifying challenges and limitations in using chatbots for teaching and learning programming — it examined technical complexities such as implementation challenges and AI limitations, along with integration issues with other educational tools. Pedagogical challenges encompassed alignment with curricula, teacher acceptance, and adaptation of teaching materials, while usage challenges focused on interaction difficulties and language barriers.

To answer RQ4 — Exploring social benefits arising from chatbot use in teaching and learning programming — we investigated inclusivity aspects, including support for students with special needs and addressing educational disparities. Additionally, we explored collaborative benefits, such as promoting teamwork and enhancing communication among students through chatbot interactions.

Finally, for RQ5 — Assessing the technology's acceptance among teachers and students — we analyzed studies presenting feedback from participants, including their opinions and adoption rates of chatbot technology in educational settings.

3. Results and Discussion

This section presents and discusses the findings from the review of 17 studies on the use of chatbots in programming education. Each subsection addresses these aspects based on the available literature, highlighting both the observed impacts and areas where further research is needed.

3.1. In what context have chatbots been utilized in the teaching and learning process of programming?

Among the 17 articles, 14 focused on students and 3 on teachers. The perspective of teachers has been underrepresented in the literature included in this scoping review. Of all the studies, 15 conducted direct research with participants.

The geographical distribution of the studies indicates a variety of locations where initiatives using LLM-based chatbots in programming education are taking place. Among the 17 studies, the distribution is well-balanced, covering countries such as the USA, Turkey, New Zealand, the Philippines, Ethiopia, North Cyprus, and China. Only the United States was repeated twice, demonstrating a diversity of locations.

In terms of educational level, the research demonstrated the initiative to use chatbots in teaching programming across the three main levels: middle school, high school, and undergraduate. However, among the 17 studies, 13 focused on undergraduate education, only 1 on high school, and 1 on middle school. Therefore, we can see their primary use is in the academic environment, but there is a lack of studies focused on other educational levels.

In terms of intended users, the studies indicate a focus on students, with 14 studies involving them compared to only 3 involving teachers. This highlights a greater emphasis on student engagement and interaction with chatbot technology in programming education, while there is a comparatively limited exploration of its use among teachers. This disparity suggests a need for further investigation into how chatbots can be effectively integrated into teaching practices to benefit educators as well as students.

All selected studies used GPT as the underlying language model for their chatbots. Among them, 3 exclusively utilized GPT-4, 2 used GPT-3.5, 1 used both versions, and the remaining studies referenced using ChatGPT without specifying the model. This indicates a preference for GPT as the primary language model for the task of teaching and learning programming.

In the studies, various programming languages were addressed, including JAVA, PHP, Python, C/C++/C#, with JAVA being the most used in 4 studies. This demonstrates the use of chatbots in teaching specific, current, and modern programming languages, as well as more traditional ones. Additionally, 7 studies did not specify a particular programming language.

3.1.1. Goal

The usage objectives were analyzed across the 17 studies and classified according to their purposes. Initially, two general categories were defined: "Learning" for studies where students use the chatbot to learn programming, and "Teaching" for studies where teachers use the chatbot to assist in programming education. Following this, specific subcategories were defined within the "Learning" category: "Assisting" for chatbots that help clarify concepts, assist in problem-solving, and aid in understanding; "Coding" for chatbots specifically involved in generating code, debugging, and suggesting code improvements; and "Teachable" where students learn by teaching the chatbot. In the "Teaching" category, specific subcategories were not created due to the limited number of studies, but the studies typically involved chatbots used for generating exercises and lesson planning.

The categories contribute to establishing a mapping system for the studies based on their usage objectives. Each study may fall into multiple subcategories; however, we observed that a study cannot belong to more than one general category. In terms of quantity, most studies are categorized under "Learning", indicating a predominant use of chatbots by students rather than teachers. The specific distribution, considering the subcategories, is detailed in Table 4.

Intended Use	Subcategory	Studies	Freq	%
Learning	Assisting	S02, S03, S04, S05, S06, S09, S10, S11, S13, S14	10	58.82%
Learning	Teachable	S01	1	5.88%
Learning	Coding	S02, S04, S05, S06, S07, S08, S09, S17	8	47.05%
Teaching	N/A	S12, S15, S16	3	17.64%

Table 4. Studies over primary goal

3.2. What educational benefits, if any, can be derived from using chatbots in the process of teaching and learning programming?

To analyze the educational benefits, we evaluated whether chatbots promoted improvements in learning, student engagement, student support, and the facilitation of teachers'

work. In terms of learning, 9 studies indicated improvements, 1 study suggested that the use was not effective, and 7 studies did not address results in this aspect. Regarding engagement, support, and facilitation for teachers, only 2 articles evidenced improvements in each of these aspects.

Overall, the literature suggests that the use of chatbots can improve programming learning. However, most articles considered only the users' opinions, predominantly students, without resorting to quantitative analyses. Additionally, the aspects of engagement, student support, and facilitation of teachers' work were not well covered in the literature, indicating the need for further research on these topics.

3.3. What are the challenges and limitations, if any, of using chatbots in the process of teaching and learning programming?

Regarding the challenges and limitations of using chatbots, the results were analyzed in three aspects: technical, pedagogical, and usage. In terms of technical challenges, 7 articles highlighted issues related to the malfunctioning of chatbots, such as hallucination, generation of incorrect code, limitation in textual comprehension, privacy, and implementation challenges. In the pedagogical aspect, 6 articles pointed out that the use of chatbots in programming education can lead to dependence on the chatbot, limitation of critical thinking, and demotivation. Lastly, in relation to usage challenges, 5 studies raised concerns about the lack of reliability, as the chatbot can provide incorrect information convincingly. Another point is the difficulty in making the chatbot understand the queries, especially when dealing with complex and abstract subjects.

Overall, the technical challenges are well-known in the field of LLM, without presenting any new issues. The pedagogical challenges and limitations, despite the concerns raised, have not been sufficiently explored, with most observations coming from students' opinions.

3.4. What social benefits, if any, arise from using chatbots in the process of teaching and learning programming?

Regarding the social benefits of using chatbots in programming education, aspects of inclusion and accessibility, as well as the promotion of collaborative use and interaction among students, were analyzed. No article addressed these aspects. Specifically regarding collaboration, all studies demonstrated an individual use of the chatbot, both by students and teachers. Overall, these aspects still represent a gap in the literature.

3.5. How widely accepted is this technology among teachers and students?

We examined the acceptance of technology in programming education, focusing on analyzing the opinions of students and teachers in interviews about the use of chatbots. Most teachers expressed concerns about the potential negative impacts of chatbots, but due to the limited number of studies (only 3 articles addressed their opinion), we cannot state with certainty. Of the 12 articles that explored students' views, 8 showed positive acceptance, 2 reported negative opinions, and 2 presented balanced views. Overall, students demonstrated receptivity to the use of chatbots for programming learning, though they are aware of its limitations and potential adverse effects.

In summary, the use of chatbots in programming education is well received by students and considered by teachers, albeit with reservations and concerns about its potential negative effects. However, teachers' perspectives remain a gap in the literature and require further research.

4. Conclusion and Future Works

This study conducted a systematic mapping of the literature on the use of chatbots in teaching and learning programming between 2020 and 2024. We analyzed 17 studies that revealed a widespread adoption of chatbots by students and teachers across different educational levels and countries such as the USA, Turkey, China, and Ethiopia, among others. ChatGPT was the most used language model.

Chatbots have shown positive impacts on programming learning and have been well-received by students. However, significant gaps were identified in the literature, such as the lack of studies focusing on teachers' perspectives, as well as issues related to student engagement, accessibility, and collaboration. These areas provide opportunities for future research, including testing other language models such as Mistral and Google's Gemini, comparisons among different LLMs, quantitative studies on learning impacts, ethical considerations regarding the use of these technologies, more detailed investigations on student engagement and accessibility in chatbot usage, and exploring the use of chatbots by teachers as a tool to facilitate and reduce repetitive workload.

Acknowledgement

This document was reviewed using ChatGPT to ensure clarity, coherence, and grammatical accuracy. ChatGPT reviewed the text but did not create it.

References

Adamopoulou, E. and Moussiades, L. (2020). An overview of chatbot technology. In Maglogiannis, I., Iliadis, L., and Pimenidis, E., editors, *Artificial Intelligence Applications and Innovations*, pages 373–383, Cham. Springer International Publishing.

Banić, B., Konecki, M., and Konecki, M. (2023). Pair programming education aided by chatgpt. In *2023 46th MIPRO ICT and Electronics Convention (MIPRO)*, pages 911–915.

Butgereit, L. and Egu, A. G. (2023). Using gpt-4 to tutor java programming in amharic. In *2023 First International Conference on the Advancements of Artificial Intelligence in African Context (AAIAC)*, pages 1–5.

Chan, W. K., Yu, Y. T., Keung, J. W., and Lee, V. C. (2023). Toward ai-assisted exercise creation for first course in programming through adversarial examples of ai models. In *2023 IEEE 35th International Conference on Software Engineering Education and Training (CSEE&T)*, pages 132–136.

Dermeval, D., Coelho, J. A. P. d. M., and Bittencourt, I. I. (2020). Mapeamento sistemático e revisão sistemática da literatura em informática na educação. In Jaques, P. A., Siqueira, S., Bittencourt, I., and Pimentel, M., editors, *Metodologia de Pesquisa Científica em Informática na Educação: Abordagem Quantitativa*. SBC, Porto Alegre.

Dos Santos, O. L. and Cury, D. (2023). Challenging the confirmation bias: Using chatgpt as a virtual peer for peer instruction in computer programming education. In *2023 IEEE Frontiers in Education Conference (FIE)*, pages 1–7.

Haindl, P. and Weinberger, G. (2024). Students' experiences of using chatgpt in an undergraduate programming course. *IEEE Access*, 12:43519–43529.

Hajj, J. A. and Sah, M. (2023). Assessing the impact of chatgpt in a php programming course. In *2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS)*, pages 1–10.

Hanifi, K., Cetin, O., and Yilmaz, C. (2023). On chatgpt: Perspectives from software engineering students. In *2023 IEEE 23rd International Conference on Software Quality, Reliability, and Security (QRS)*, pages 196–205.

Hu, M., Assadi, T., and Mahroeian, H. (2023). Explicitly introducing chatgpt into first-year programming practice: Challenges and impact. In *2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, pages 1–6.

Jin, H., Lee, S., Shin, H., and Kim, J. (2024). Teach ai how to code: Using large language models as teachable agents for programming education. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA. Association for Computing Machinery.

Kazemitabaar, M., Ye, R., Wang, X., Henley, A. Z., Denny, P., Craig, M., and Grossman, T. (2024). Codeaid: Evaluating a classroom deployment of an llm-based programming assistant that balances student and educator needs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA. Association for Computing Machinery.

Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report.

Labadze, L., Grigolia, M., and Machaidze, L. (2023). Role of ai chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1):56.

Liao, J., Zhong, L., Zhe, L., Xu, H., Liu, M., and Xie, T. (2024). Scaffolding computational thinking with chatgpt. *IEEE Transactions on Learning Technologies*, 17:1668–1682.

Liu, R., Zenke, C., Liu, C., Holmes, A., Thornton, P., and Malan, D. J. (2024). Teaching cs50 with ai: Leveraging generative artificial intelligence in computer science education. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, SIGCSE 2024, page 750–756, New York, NY, USA. Association for Computing Machinery.

Padilla, J. R. C., Montefalcon, M. D. L., and Hernandez, A. A. (2023). Language ai in programming: A case study of chatgpt in higher eduation using natural language processing. In *2023 IEEE 11th Conference on Systems, Process & Control (ICSPC)*, pages 276–281.

Spasić, A. J. and Janković, D. S. (2023). Using chatgpt standard prompt engineering techniques in lesson preparation: Role, instructions and seed-word prompts. In *2023 58th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*, pages 47–50.

Speth, S., Meißner, N., and Becker, S. (2023). Investigating the use of ai-generated exercises for beginner and intermediate programming courses: A chatgpt case study. In *2023 IEEE 35th International Conference on Software Engineering Education and Training (CSEE&T)*, pages 142–146.

Vicari, R. M. (2021). Influências das tecnologias da inteligência artificial no ensino. *Estudos Avançados*, 35:73–84.

Wang, T., Díaz, D. V., Brown, C., and Chen, Y. (2023). Exploring the role of ai assistants in computer science education: Methods, implications, and instructor perspectives. In *2023 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 92–102.

Yilmaz, R. and Karaoglan Yilmaz, F. G. (2023). The effect of generative artificial intelligence (ai)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, 4:100147.

Zhang, K. and Aslan, A. B. (2021). Ai technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2:100025.