GPT AI in Computer Science Education: A Systematic Mapping Study

Bruno H. Strik[1,2], André Menolli[1,3], Jacques Duílio Brancher[1]

[1]Departamento de Computação - Universidade Estadual de Londrina (UEL)
Londrina, PR - Brazil

[2]Instituto Federal do Paraná (IFPR)
Astorga, PR - Brazil.

[3]Universidade Estadual do Norte do Paraná (UENP)
Bandeirantes, PR - Brazil

bruno.strik@ifpr.edu.br, menolli@uenp.br, jacques@uel.br

Abstract. *With the advent of GPT-AI, new possibilities in education emerged. However, it is challenging to determine how and when to apply these new technologies and understand their actual impact on teaching and learning. This study conducts a systematic mapping to gather, include, and classify scientific papers that investigated the subject of generative AI in CS education. 31 relevant studies that conducted empirical evaluations of the application of GPT-AI tools in CS education were collected. Our findings highlight challenges regarding plagiarism, learning perception, and AI capability. The main contribution of this study is to present research opportunities and provide a background for future studies that address the application of GPT-AI in CS education.*

1. Introduction

The field of artificial intelligence (AI) applied to education is not new. As early as 1970, a study by Jaime R. Carbonell introduced a prototype running on a PDP-10 that interacted with students dialogically, assisting them in their studies [Carbonell 1970]. While the prototype had limitations due to the technology of its time, it already highlighted the potential of AI-assisted learning tools. Fifty years later, the extensive use of AI in education had not become a reality until the release of ChatGPT [Williamson 2024], which was made available to the general public in November 2022 [OpenAI 2022]. This event marked the first time an AI tool based on the GPT (Generative Pre-trained Transformer) architecture was freely accessible to the general public. The impressive characteristics of this tool, along with its capacity for human-like natural language interactions, led to its rapid adoption by the general public [Caldarini et al. 2022].

LLMs (Large Language Models) are a type of AI model developed to understand and generate natural language [Minaee et al. 2024]. They are trained on enormous amounts of textual data and based on Deep Neural Network architectures. GPT [Achiam et al. 2023] is a specific architecture of LLMs developed by OpenAI, which underpins the tools analyzed in the studies of this mapping, such as Copilot, Codex, and the most widely used and known, ChatGPT.

Although the full impact of ChatGPT is still unfolding, some studies [Hammad and Bahja 2023, Ghassemi et al. 2023, Hwang and Chang 2021,

DOI: 10.5753/sbie.2024.242103

Rotman 2023] indicate that its introduction may be as transformative as the internet. Just as the internet fundamentally changed the way we access and share information, generative AI like ChatGPT can change the way we work, learn, and interact with technology, which is already being observed in education [Mosaiyebzadeh et al. 2023].

The educational use of ChatGPT and other generative AI tools was not initially planned; rather, their use emerged exploratively by both students and educators [Dempere et al. 2023]. Initially unnoticed, their usage soon garnered attention and raised concerns among many educators and researchers regarding their influence on learning processes [Farhi et al. 2023]. The emergence and popularization of generative AI have been met with diverse reactions. Initially, concerns about its use led to bans or restrictions [Lau and Guo 2023], not only within education [BBC News 2023]. Regardless of the perception of various stakeholders, a revolution is happening, and its impact cannot be ignored. Significant educational reforms can be necessary, or at least profound changes may be considered, as combating or ignoring this technology does not seem to be a productive approach. Faced with the new challenges brought by generative AI tools, it is necessary to understand these technologies and consider the roles they can play in the teaching and learning of computer science.

Given the challenge of understanding and harnessing the potential of generative AI to enhance computer science education, this systematic mapping seeks to summarize the works that empirically explore the use of Generative AI tools based on neural networks with transformer architecture in computer science education published in the first 16 months after the release of ChatGPT[1], aiming to contribute to the understanding and construction of successful practices in integrating these new technologies in computer science education.

2. Methodology

This section outlines the formal guidelines adhered to, including the research questions, study selection criteria, search string, and quality assessment used in this systematic mapping. To meet the objectives of this study, we followed the latest guidelines for systematic mapping studies in software engineering by Petersen et al. [Petersen et al. 2015] according to the scheme presented in Figure 1.

2.1. Research questions

This mapping study aims to understand how generative AI has been explored in the educational context of computer science teaching. We intend to identify the most frequently studied subjects, analyze the findings and concerns raised, and assess the overall impact of its use. The following Research Questions (RQs) were formulated to guide this investigation:

- RQ01: How has generative AI been used in computer science education?
- RQ02: Which subjects have employed generative AI in their teaching activities?
- RQ03: What are the observations regarding the use of generative AI in computer science education?

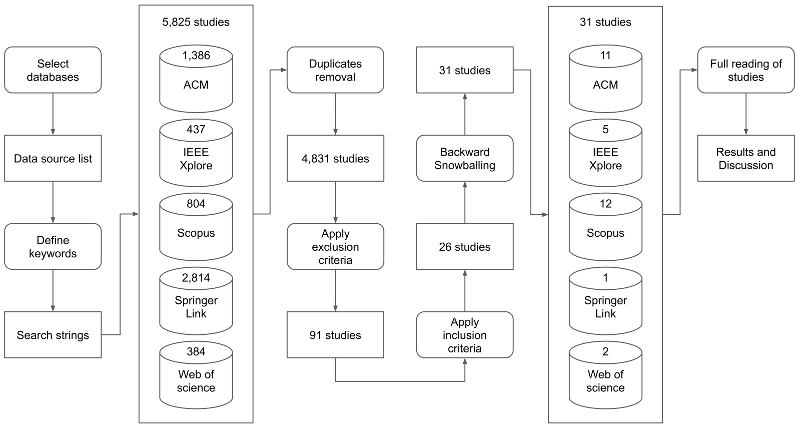

Figure 1. Research method representation inspired by Geraldi et al. [Geraldi et al. 2020]

2.2. Sources and search strategy

The PICO approach (Population, Intervention, Comparison, and Outcomes), suggested by Kitchenham and Charters [Kitchenham and Charters 2007], was employed to support the formulation of the search string.

- Population: refers to the study's target group, which may include computer science students, instructors, educational institutions, specific courses or subjects, researchers, or administrators. In this study, we included direct actors: students and instructors.
- Intervention: refers to the specific AI action, strategy, or tool being implemented in the educational context. Here, we focus on specifying generative AI tools.
- Comparison: refers to the reference group or condition used to evaluate the effectiveness of the AI intervention. Here we will include only empirical studies, whether comparative or not.
- Outcomes: represent the desired outcomes or effects of the AI intervention. There are no restrictions on the outcomes and effects obtained.

The search string employed was developed to combine various terms from three main topics. The first group contains keywords indicating that the studies should address the field of computing. The second group combines words indicating that the studies focus on education. The third group clusters words related to generative artificial intelligence. The keywords within each group are separated by the logical connector OR and each group by the connector AND, aiming to find articles encompassing all three topics, containing at least one word from each group in any combination.

The decision to explicitly include the terms GPT and Copilot in the search string was motivated by the fact that ChatGPT and Github Copilot were the most widely used generative AI tools at the time of the study's data collection [FlexOS 2024]. However, to

[1] https://chatgpt.com/

ensure that the identified studies were not limited to these two tools, additional terms representing other generative AI tools were also included, connected by the logical operator OR.

The search string employed is presented below, with adaptations made according to the formatting required by each database, when necessary:

(”computer” OR ”computing” OR ”software” OR ”programming” OR ”algorithm”) AND (”education” OR ”teaching” OR ”learning”) AND (”generative AI” OR ”GPT” OR ”copilot” OR ”generative artificial intelligence” OR ”Generative Pre-trained Transformer”)

The electronic databases searched as sources for the articles used in this study and the total number of articles found in each are presented in Figure 1.

2.3. Inclusion and exclusion criteria

Before applying the inclusion and exclusion criteria, duplicate articles were removed, reducing the sample by 24.78%.

To extract the articles that match the study's goal, we analyzed the articles based on their titles and abstracts and applied the exclusion criteria in the order presented below. Therefore, if an article met more than one exclusion criterion, its removal was motivated by the first applicable criterion. Table 1 presents the scope of exclusions according to the criteria.

Exclusion criteria	Total exclusions
1. Studies not presented in English	2
2. Books and grey literature	17
3. Conference summaries, editorials, or guidelines	322
4. Studies that do not address the Computer Science field	2,465
5. Articles that do not employ generative AI	380
6. Studies that do not apply generative AI in education	1,645

Table 1. Scope of exclusions by criteria

With a reduced sample, we applied the inclusion analysis. At this stage, we conducted a full-text reading of each article to determine if they were suitable according to the following inclusion criteria:

1. Studies must be empirical.
2. Studies must address teaching and learning-related experiences.
3. Studies must be conducted under formal courses of computer science and immediately related fields.
4. The research method must be clearly defined.

After removing duplicate articles and applying the exclusion and inclusion criteria, five new articles were added through backward snowball sampling. These articles were analyzed according to the same exclusion and inclusion criteria applied to the others. Figure 1 describes the entire process, which resulted in a final set of 31 articles.

2.4. Data extraction

Data from the included studies were extracted and organized in a spreadsheet according to the form presented in Table 2. This table was developed to facilitate the identification of key points in the articles and to assist in the process of classifying and extracting the necessary information to address the research questions. Using these key points, it is possible to classify the selected articles in various ways, making it important to detail some key data.

Key point	**Details**
ID	Article identifier
Article title	The title of article
Sample size	Number of participants in the empirical study
Educational level	Participants' stage of formal education.
Courses and subjects	Courses and subjects in which the empirical study was contextualized
Programming language	The programming language used in the study experiments, if applicable.
Country or region	Country or region in which the study was conducted
AI tools	AI tools used in the study
Teaching method	If the empirical study took place in a context where a specific teaching method was applied, this field is filled with the respective method
Methods	Data collection methods employed
Evaluated aspects	Aspects evaluated in the empirical study

Table 2. Form used for the extraction and organization of data from included studies.

3. Results

In this section, we present the detailed findings of our study, focusing on the evaluated aspects and the collected information. The objective was to understand how generative AI tools are being used in computer science education, the subjects that have employed these tools, and the general perceptions regarding their use.

3.1. Overview of included studies

The selected studies were conducted in various countries, as highlighted in Figure 2. Each country has different characteristics for classifying higher education levels. To facilitate organization and classification, the educational levels of the study participants were categorized into two groups: undergraduate and postgraduate. Among the articles selected for this mapping, 28 studies were conducted with undergraduate students and 3 with postgraduate students.

The majority of the studies (23) were conducted with samples consisting solely of students, with an average of 135.1 participants. Four studies were conducted with samples composed of both instructors and students, with averages of 3.5 and 63.5 participants, respectively. Four studies were conducted with instructors only, having an average of 8.2 participants.

Analyzing the publications from a chronological perspective, it is possible to observe an increase in the number of publications that meet the criteria of this study. Figure 2 and Table 3 details the publications by type, region and year. It's important to state that the timespan of the collected studies covers the first 16 months after the release of ChatGPT, the first Generative AI tool publicly available, and only the year 2023 was fully covered in the researched period, which help explain the disproportion in the number of works per year.

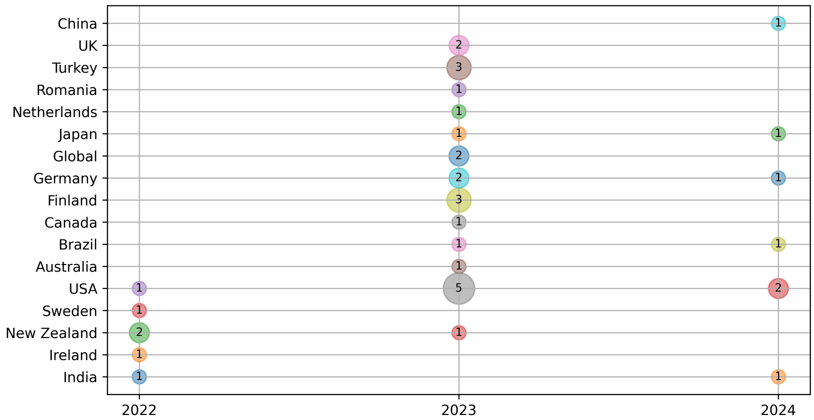

Figure 2. Number of publications by year and region

Type	2022	2023	2024
Conference proceedings	2	7	2
Journal	1	14	5

Table 3. Number of publications by year and type

3.2. AI tools utilization

Most of the selected studies specified the generative AI tools employed, with ChatGPT being the most used, present in more than half of the studies. Other tools, such as Codex[2], DaVinci[3], DALL-E[4], and Copilot[5], were also present. Some studies employed multiple technologies. Table 4[6] shows the variety of AI tools used in the studies, divided by subjects.

Among these studies, ChatGPT and Copilot were the most commonly used AI tools. Python was the most frequently used programming language, as shown in Figure 3, which presents the relationship between programming languages and AI tools.

[2]https://openai.com/index/openai-codex/
[3]https://davinci.ai/
[4]https://openai.com/index/dall-e-2/
[5]https://github.com/features/copilot
[6]Studies 16 and 17 employed image generation for the game-development activities associated with programming.

AI Tool	**Knowledge area**		
	Programming	Software Engineering	General
ChatGPT	20	3	1
Copilot	6	0	1
Codex	3	0	0
DaVinci	1	0	0
DALL-E	1	0	0
Unspecified	0	0	1

Table 4. AI tools utilized in the mapped studies.

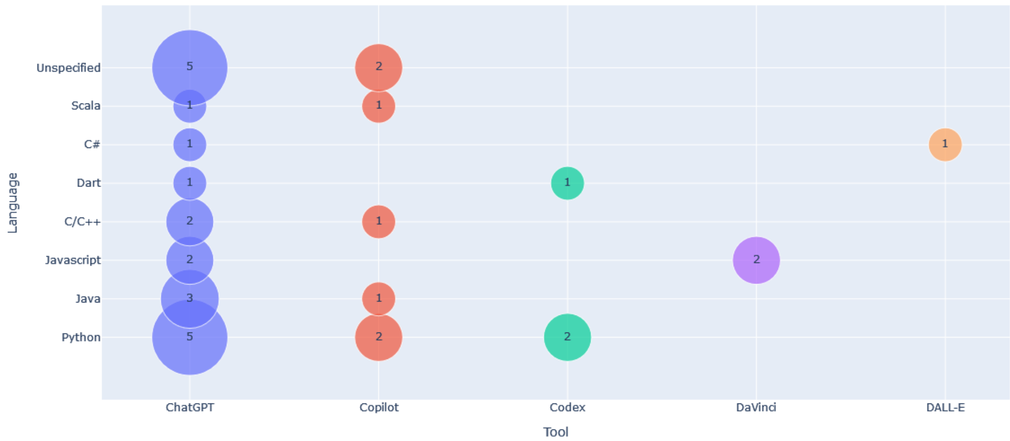

Figure 3. Programming languages used in the studies, by AI tool

The use of different generative AI tools in the selected studies is consistent with the overall proportion of these tools' general usage [FlexOS 2024]. It is important to note that, during the development of this mapping, new generative AI products were launched, and others were rebranded. For the classification presented here, the generative AI tools are identified by their names as of June 2024.

The detailing of activities supported by AI tools presented in the studies is shown in Figure 4, with the meaning of each application understood as follows:

- **Code generation:** Automatic code generation by AI.
- **Debugging:** Error identification and correction suggestions from faulty code or error message.
- **Data analysis:** Data analysis, detailing, and summarization.
- **Systems modelling:** Creation of representations of a system's components and interactions.
- **Refactoring:** Improvement of code structure without changing its external behavior.
- **Software architecting:** Design of structures and frameworks.
- **Requirements engineering:** Definition, documentation, and management of software requirements and specifications.
- **Code explanation:** Explanation of the purpose and functionality of functions and code snippets.

- **Technology suggestions:** Suggestions for the use of libraries, frameworks, and technologies to solve problems or add features.
- **Software testing:** Evaluation of software functionality and performance to identify defects.
- **Conceptual support:** Explanation and assistance in understanding certain concepts.

Application / Article Id	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	TOTAL
Code generation		■	■		■		■			■	■	■			■	■	■	■				■	■	■	■	■	■	■	■	■	■	21
Debugging	■	■			■					■		■			■	■		■	■	■				■	■	■			■	■	■	16
Data analysis											■														■							2
Systems modeling				■																					■							2
Refactoring																									■							1
Software architecting				■																					■							2
Requirements engineering																									■							1
Code explanation	■	■			■				■		■	■		■				■			■	■		■	■						■	13
Technology suggestions		■								■	■					■		■			■				■						■	8
Software testing						■			■																				■			3
Conceptual support				■				■	■		■		■	■				■			■								■		■	10

Figure 4. Activities supported by Generative AI tools in educational context present in studies

All selected articles in this mapping conducted empirical experiments, applying multiple and varied designs, as shown in Figure 5.

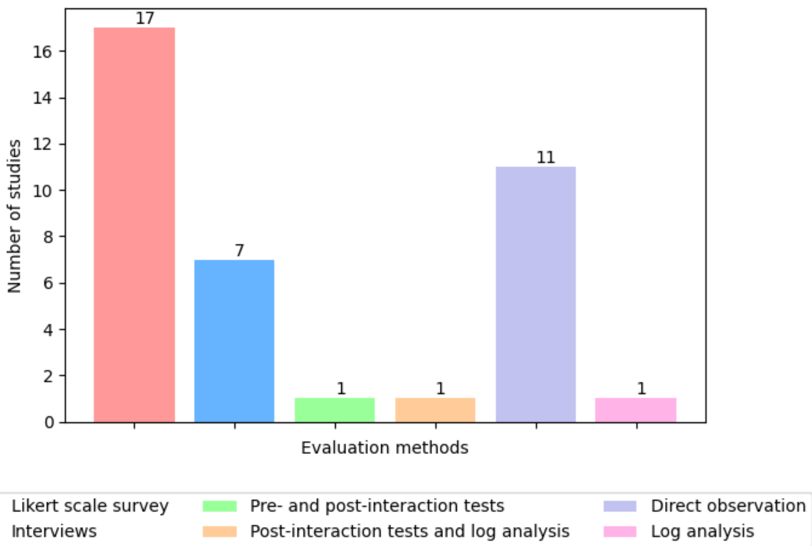

Figure 5. Distribution of evaluation methods in selected studies

3.3. Educational aspects

The vast majority of the studies in the sample did not specify the application of differentiated teaching methods in the context of their experiments. Among the two studies (26 and 29) that adopted a specific methodology, the flipped classroom model was applied.

Several studies addressed the observations of students and instructors regarding the applicability and usefulness of generative AI tools in their academic routines. Both students' and instructors' perceptions and concerns are presented in Figures 6 and 7.

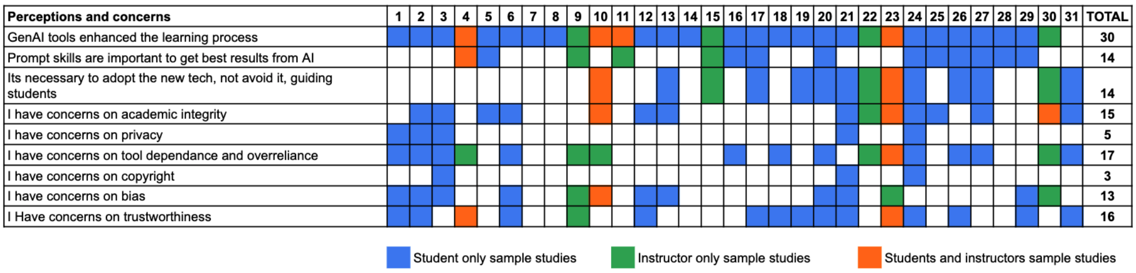

Figure 6. Perceptions and concerns reported in studies.

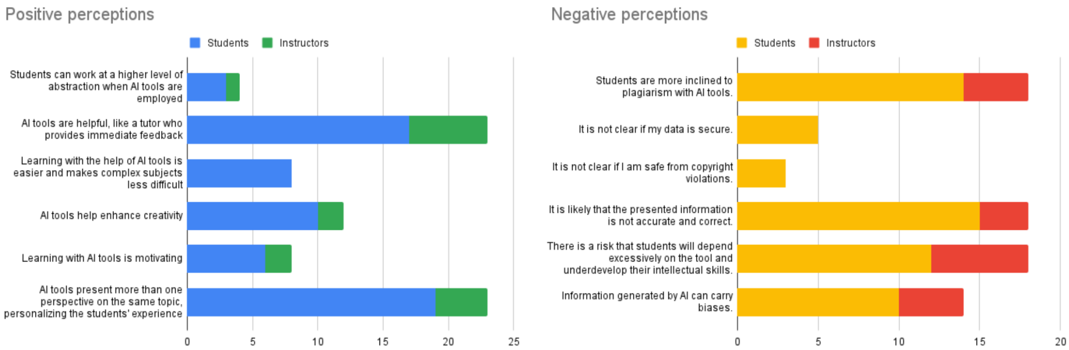

Figure 7. Number of works and pointed aspects

3.4. Quality assessment

As stated by Petersen et al. [Petersen et al. 2015], no quality assessment needs to be performed in systematic maps, as the analysis of exclusion and inclusion criteria is sufficient to evaluate the acceptance of articles in the mapping. However, the quality of the articles was considered during their selection and evaluated during the full-text reading conducted at the inclusion criteria analysis stage and the backward snowball sampling. Figure 8 shows the results of the quality assessment of all articles included in this study.

The three criteria described below were used to evaluate the quality of the articles and covered relevant aspects that needed to be clearly determined for them to be included.

1. **QA1:** There was an adequate description of the context in which the research was carried out.
2. **QA2:** There was an adequate description of the sample used and the methods for identifying and recruiting the sample.
3. **QA3:** There was an adequate description of the evaluated aspects, with the results being presented coherently.

The quality scale of the articles has four possible values: Insufficient (I), Sufficient (S), Good (G), and Excellent (E), applied to each of the three criteria. If an article was evaluated as insufficient in any criterion, it would not be considered suitable for inclusion in this mapping study.

The complete list of articles from this mapping study is available in an anonymous repository [7], with each article assigned a number used for referencing within this study.

[7]https://anonymous.4open.science/r/GPTAICSEdu/

QA Criteria	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
QA1	G	S	G	G	S	E	S	G	G	E	G	G	E	G	G	G	G	G	G	E	E	E	E	E	E	G	G	E	G	E	E
QA2	G	E	S	G	S	G	S	G	S	G	S	G	G	S	G	G	G	G	G	S	G	E	E	E	E	G	G	G	S	S	G
QA3	G	G	G	G	S	E	S	E	E	G	G	G	E	G	G	G	G	E	G	E	G	E	E	E	E	E	E	G	E	E	E

Figure 8. Results of the quality assessment application.

4. Discussion

Throughout the investigation, we observed where and how AI tools are used, the subjects that employ them, and the general perceptions and concerns of students and instructors regarding their applicability and impact. Next, we discuss in detail the results related to the three main research questions.

RQ01. How has generative AI been used in computer science education? Generative AI has been primarily used for code generation, debugging, code explanation, and conceptual clarification in programming tasks. Figure 4 provides more details on the uses of generative AI mentioned in the articles. Students interact with the AI by prompting topics, error messages, and code snippets, instructing tasks, seeking help, and asking questions.

RQ02. Which subjects have employed generative AI in their teaching activities? Generative AI tools have been studied in various computer science subjects, as shown in Table 4. Programming-related subjects are the most prevalent in the collected studies. Of these, 50% of the investigations focused on introductory programming courses, while the remaining 50% were divided among web programming, data science, game development, and other areas. Additionally, experiences of using these tools in Software Engineering were also reported, although in small numbers.

RQ03. What are the observations regarding the use of generative AI in computer science education?

Various impressions and concerns were identified in the mapped studies. From the collected results, the positive and negative aspects of using GPT in computer science education are mapped in Figure 7. In addition, general perceptions and concerns are listed in Figure 6.

Among the observations regarding the positive contributions of generative AI tools, most studies highlight that AI tools assist students in their academic activities similarly to a tutor, providing immediate feedback. However, the uncertainty of the correctness of the generated information is a common negative perception noted by studies, along with potential biases that can affect the precision and accuracy of the information.

The ability of generative AI tools to provide different perspectives and alternative explanations of topics is also commonly noted as a positive aspect, which can help students when they encounter difficulties in their studies.

Nevertheless, some studies have expressed concerns about the pedagogical influence of AI tools in computer science education. Beyond issues of bias and accuracy, studies point out that students might become overly dependent on AI tools, neglecting the development of their own skills and intellect.

Furthermore, the need for a generative AI tool that is aware of its role as a teaching aid has been emphasized in various articles. The natural tendency of these tools to directly produce answers may not be as beneficial for learning, as it allows students to bypass the process and focus excessively on the answer. Although this concern aligns with academic integrity issues, it is also noted by some studies that the tool neither exacerbates nor alleviates this pre-existing difficulty, as students who simply plagiarize will continue to do so regardless of the incorporation of AI tools.

Several studies have pointed out that the adoption of generative AI must be carefully planned to avoid the replacement of conceptual foundations and technical skills with the ability to effectively use the tool, thereby impairing the intellectual development of students. The risks of inadequate adoption of generative AI in teaching activities need to be mitigated through deep and careful methodological planning.

The accuracy of the statements produced by AI is also closely monitored, with a high level of incomplete or incorrect responses being reported. Some studies suggest that well-crafted prompts can mitigate this problem, but even so, generative AI should not be considered the sole source of information. All information provided by AI must be verified, as the stochastic nature of AI makes it vulnerable to biases and incorrect information.

Studies that addressed the perspective of instructors in the first months after the public release of ChatGPT and other AI tools pointed to more intense concerns and even a desire for the technology to be banned from educational processes. Subsequent studies did not present this desire as frequently, indicating that the familiarization of students and instructors with the tools allowed for a better understanding of the initial insecurities and a more positive view of their potential.

Despite the various concerns presented, the overall conclusion of the studies is that the use of generative AI is beneficial and positive. Its problems can be mitigated by refining its incorporation into teaching activities and by maturing the relationship between students and the tools. Studies have identified positive feelings regarding students' confidence when using generative AI tools, comparing their contribution to that of an ever-available tutor, addressing the same topic in various ways according to the student's learning convenience.

5. Limitations and threats to validity

One significant limitation of this study is the rapidly evolving nature of generative AI technology. As new features and tools are continually being introduced, the landscape of AI applications in education is in constant flux. For instance, ChatGPT Edu[8] was announced while this mapping was being written, illustrating how quickly advancements can occur. This rapid development makes it challenging to provide a comprehensive and up-to-date analysis, as some of the latest innovations may not be fully covered or evaluated in our study.

Another issue is the relatively small number and variety of studies available on this topic. Due to the novelty of generative AI, there is a limited body of research exploring its implications and effectiveness in educational settings, specifically in computer

[8]https://openai.com/index/introducing-chatgpt-edu/

science learning. This scarcity of studies focused on CS education not only restricts the depth of analysis but also makes it difficult to draw broad conclusions about the technology's impact. Some interesting studies recently published, such as those by Ansari et al. [Ansari et al. 2023], Yan et al. [Yan et al. 2024], Purnama et al. [Purnama et al. 2023], Baber et al. [Baber et al. 2024], and Lin et al. [Lin et al. 2024], go beyond the scope of this review by expanding the study beyond computer science education and provide valuable insights aligned with the findings of this mapping. Moreover, the fast-paced nature of AI advancements means that existing studies can quickly become outdated, further complicating efforts to keep up with the latest developments and accurately assess their validity.

6. Conclusion and Future Works

The integration of generative AI in computer science education presents both opportunities and challenges. While it facilitates progress and the understanding of concepts, it is essential to address their related concerns. Educators and students must adapt to these tools, ensuring they complement rather than compromise educational integrity and cognitive development.

The future perspectives of studying the application of GPT-AI in computer science education involve promising opportunities. One area of interest is the development of advanced, adaptive learning systems that can personalize students' experiences based on their individual needs and learning styles. Future research could explore teaching methodologies that effectively integrate GPT-AI, ensuring that these tools enhance students' intellectual development and support the role of educators, maximizing their potential. Additionally, it is necessary to study the long-term impacts of using GPT-AI tools on student outcomes, including knowledge retention, problem-solving skills, and career readiness. Finally, issues such as data privacy, bias, and copyright need to be addressed. These future perspectives aim to improve the effectiveness of computer science education and ensure that the implementation of generative AI tools is responsible and equitable.

Acknowledgements

The authors would like to acknowledge the use of ChatGPT-4o for text revision. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Adiguzel, T., Kaya, M. H., and Cansu, F. K. (2023). Revolutionizing education with ai: Exploring the transformative potential of chatgpt. *Contemporary Educational Technology*, 15(3):ep429.

Afjal, M. (2023). Chatgpt and the ai revolution: a comprehensive investigation of its multidimensional impact and potential. *Library Hi Tech*, ahead-of-print(ahead-of-print).

Albonico, M. and Varela, P. J. (2023). A report on the use of chatgpt in software engineering and systems analysis courses. In *Proceedings of the XXXVII Brazilian Symposium on Software Engineering*, pages 303–311.

Ansari, A. N., Ahmad, S., and Bhutta, S. M. (2023). Mapping the global evidence around the use of chatgpt in higher education: A systematic scoping review. *Education and Information Technologies*, pages 1–41.

Aruleba, K., Sanusi, I. T., Obaido, G., and Ogbuokiri, B. (2023). Integrating chatgpt in a computer science course: Students perceptions and suggestions. *arXiv preprint arXiv:2402.01640*.

Baber, H., Nair, K., Gupta, R., and Gurjar, K. (2024). The beginning of chatgpt–a systematic and bibliometric review of the literature. *Information and Learning Sciences*, 125(7/8):587–614.

Baker, J. (2000). The "classroom flip". using web course management tools to become the guide on the side. In *11th International Conference on College Teaching and Learning*, Jacksonville, FL.

BBC News (2023). Italy temporarily bans chatgpt over privacy concerns. `https://www.bbc.com/news/technology-65139406`. Accessed: 2024-06-01.

Budhiraja, R., Joshi, I., Challa, J. S., Akolekar, H. D., and Kumar, D. (2024). "it's not like jarvis, but it's pretty close!"-examining chatgpt's usage among undergraduate students in computer science. In *Proceedings of the 26th Australasian Computing Education Conference*, pages 124–133.

Caldarini, G., Jaf, S., and McGarry, K. (2022). A literature survey of recent advances in chatbots. *Information*, 13(1):41.

Carbonell, J. R. (1970). Ai in cai: An artificial-intelligence approach to computer-assisted instruction. *IEEE transactions on man-machine systems*, 11(4):190–202.

Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y.-S., Gašević, D., and Chen, G. (2023). Can large language models provide feedback to students? a case study on chatgpt. In *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)*, pages 323–325. IEEE.

Dempere, J., Modugu, K., Hesham, A., and Ramasamy, L. K. (2023). The impact of chatgpt on higher education. *Frontiers in Education*, 8.

Denny, P., Kumar, V., and Giacaman, N. (2023). Conversing with copilot: Exploring prompt engineering for solving cs1 problems using natural language. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2023)*, pages 1136–1142, New York, NY, USA. Association for Computing Machinery.

Farhi, F., Jeljeli, R., Aburezeq, I., Dweikat, F. F., Al-shami, S. A., and Slamene, R. (2023). Analyzing the students' views, concerns, and perceived ethics about chat gpt usage. *Computers and Education: Artificial Intelligence*, 5:100180.

Finnie-Ansley, J., Denny, P., Becker, B. A., Luxton-Reilly, A., and Prather, J. (2022). The robots are coming: Exploring the implications of openai codex on introductory programming. In *Proceedings of the 24th Australasian Computing Education Conference*, pages 10–19.

FlexOS (2024). Generative ai top 150: The world's most used ai tools. Accessed: 2024-06-02.

French, F., Levi, D., Maczo, C., Simonaityte, A., Triantafyllidis, S., and Varda, G. (2023). Creative use of openai in education: case studies from game development. *Multimodal Technologies and Interaction*, 7(8):81.

Gehringer, E. and Peddycord III, B. (2013). The inverted-lecture model: A case study in computer architecture. pages 489–494.

Geraldi, R. T., Reinehr, S., and Malucelli, A. (2020). Software product line applied to the internet of things: A systematic literature review. *Information and Software Technology*, 124:106293.

Ghassemi, M., Birhane, A., Bilal, M., Kankaria, S., Malone, C., Mollick, E., and Tustumi, F. (2023). Chatgpt one year on: who is using it, how and why? *Nature*, 624(7990):39–41.

Hammad, R. and Bahja, M. (2023). Opportunities and challenges in educational chatbots. *Trends, Applications, and Challenges of Chatbot Technology*, pages 119–136.

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. IEEE.

Hellas, A., Leinonen, J., Sarsa, S., Koutcheme, C., Kujanpää, L., and Sorva, J. (2023). Exploring the responses of large language models to beginner programmers' help requests. In *Proceedings of the 2023 ACM Conference on International Computing Education Research-Volume 1*, pages 93–105.

Hidalgo, C. G., Bucheli-Guerrero, V. A., and Ordóñez-Eraso, H. A. (2023). Artificial intelligence and computer-supported collaborative learning in programming: A systematic mapping study. *Tecnura*, 27(75):175–206.

Hou, I., Mettille, S., Man, O., Li, Z., Zastudil, C., and MacNeil, S. (2024). The effects of generative ai on computing students' help-seeking preferences. In *Proceedings of the 26th Australasian Computing Education Conference*, pages 39–48.

Hwang, G.-J. and Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 31(7):4099–4112.

Jalil, S., Rafi, S., LaToza, T., Moran, K., and Lam, W. (2023). Chatgpt and software testing education: Promises & perils. *arXiv*.

Jonsson, M. and Tholander, J. (2022). Cracking the code: Co-coding with ai in creative programming education. In *Proceedings of the 14th Conference on Creativity and Cognition*, pages 5–14.

Kazemitabaar, M., Chow, J., Ma, C. K. T., Ericson, B. J., Weintrop, D., and Grossman, T. (2023). Studying the effect of ai code generators on supporting novice learners in introductory programming. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–23.

Kiesler, N., Lohr, D., and Keuning, H. (2023). Exploring the potential of large language models to generate formative programming feedback. In *2023 IEEE Frontiers in Education Conference (FIE)*, pages 1–5. IEEE.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). Systematic literature reviews in software engineering - a systematic literature review. *Information and Software Technology*, 51(1):7–15.

Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical report EBSE-2007-01, EBSE Technical Report.

Kitchenham, B. A., Budgen, D., and Brereton, O. P. (2010). The value of mapping studies – a participant-observer case study.

Kokkoniemi, M. and Isomöttönen, V. (2023). A systematic mapping study on group work research in computing education projects. *Journal of Systems and Software*, 204:111795.

Lau, S. and Guo, P. (2023). From" ban it till we understand it" to" resistance is futile": How university programming instructors plan to adapt as more students use ai code generation and explanation tools such as chatgpt and github copilot. In *Proceedings of the 2023 ACM Conference on International Computing Education Research-Volume 1*, pages 106–121.

Leinonen, J., Denny, P., MacNeil, S., Sarsa, S., Bernstein, S., Kim, J., Tran, A., and Hellas, A. (2023). Comparing code explanations created by students and large language models. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, pages 124–130.

Lemke, C., Kirchner, K., Anandarajah, L., and Herfurth, F. N. (2023). Exploring the student perspective: Assessing technology readiness and acceptance for adopting large language models in higher education. In *22nd European Conference on e-Learning: ECEL 2023*. Academic Conferences and publishing limited.

Lin, M. P.-C., Chang, D., Hall, S., and Jhajj, G. (2024). Preliminary systematic review of open-source large language models in education. In *International Conference on Intelligent Tutoring Systems*, pages 68–77. Springer.

Lo, C. (2023). What is the impact of chatgpt on education? a rapid review of the literature. *Education Sciences*, 13(4):410.

Ma, B., Chen, L., and Konomi, S. (2024). Enhancing programming education with chatgpt: A case study on student perceptions and interactions in a python course. *arXiv preprint arXiv:2403.15472*.

Maher, M. L., Latulipe, C., Lipford, H., and Rorrer, A. (2015). Flipped classroom strategies for cs education. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, SIGCSE '15, page 218–223, New York, NY, USA. Association for Computing Machinery.

Maher, M. L., Tadimalla, S. Y., and Dhamani, D. (2023). An exploratory study on the impact of ai tools on the student experience in programming courses: an intersectional analysis approach. In *2023 IEEE Frontiers in Education Conference (FIE)*, pages 1–5. IEEE.

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., and Gao, J. (2024). Large language models: A survey. *arXiv preprint arXiv:2402.06196*.

Mosaiyebzadeh, F., Pouriyeh, S., Parizi, R., Dehbozorgi, N., Dorodchi, M., and Batista, D. (2023). Exploring the role of chatgpt in education: Applications and challenges. pages 84–89.

OpenAI (2022). Introducing chatgpt. `https://openai.com/index/chatgpt/`. Accessed: 2024-06-01.

Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1–18.

Popovici, M.-D. (2023). Chatgpt in the classroom. exploring its potential and limitations in a functional programming course. *International Journal of Human–Computer Interaction*, pages 1–12.

Prather, J., Denny, P., Leinonen, J., Becker, B. A., Albluwi, I., Craig, M., Keuning, H., Kiesler, N., Kohn, T., Luxton-Reilly, A., et al. (2023a). The robots are here: Navigating the generative ai revolution in computing education. In *Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education*, pages 108–159.

Prather, J., Reeves, B. N., Denny, P., Becker, B. A., Leinonen, J., Luxton-Reilly, A., Powell, G., Finnie-Ansley, J., and Santos, E. A. (2023b). "it's weird that it knows what i want": Usability and interactions with copilot for novice programmers. *ACM Transactions on Computer-Human Interaction*, 31(1):1–31.

Purnama, I., Edi, F., Agustin, R., Pranoto, N. W., et al. (2023). Gpt chat integration in project based learning in learning: a systematic literature review. *Jurnal Penelitian Pendidikan IPA*, 9(SpecialIssue):150–158.

Rahman, M. M. and Watanobe, Y. (2023). Chatgpt for education and research: Opportunities, threats, and strategies. *Applied Sciences*, 13(9):5783.

Rajala, J., Hukkanen, J., Hartikainen, M., and Niemelä, P. (2023). "call me kiran" – chatgpt as a tutoring chatbot in a computer science course. In *Proceedings of the 26th International Academic Mindtrek Conference*, Mindtrek '23, page 83–94, New York, NY, USA. Association for Computing Machinery.

Reiche, M. and Leidner, J. L. (2023). Bridging the programming skill gap with chatgpt: A machine learning project with business students. In *European Conference on Artificial Intelligence*, pages 439–446. Springer.

Rotman, D. (2023). How chatgpt will revolutionize the economy. we need to decide what that looks like. *MIT Technology Review*.

Sheese, B., Liffiton, M., Savelka, J., and Denny, P. (2024). Patterns of student help-seeking when using a large language model-powered programming assistant. In *Proceedings of the 26th Australasian Computing Education Conference*, pages 49–57.

Silva, C. A. G. d., Ramos, F. N., de Moraes, R. V., and Santos, E. L. d. (2024). Chatgpt: Challenges and benefits in software programming for higher education. *Sustainability*, 16(3):1245.

Sun, D., Boudouaia, A., Zhu, C., and Li, Y. (2024). Would chatgpt-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? an empirical study. *International Journal of Educational Technology in Higher Education*, 21(1):14.

Williamson, B. (2024). The social life of ai in education. *International Journal of Artificial Intelligence in Education*, 34(1):97–104.

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., and Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1):90–112.

Yilmaz, R. and Yilmaz, F. G. K. (2023a). Augmented intelligence in programming learning: Examining student views on the use of chatgpt for programming learning. *Computers in Human Behavior: Artificial Humans*, 1(2):100005.

Yilmaz, R. and Yilmaz, F. G. K. (2023b). 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.

Zastudil, C., Rogalska, M., Kapp, C., Vaughn, J., and MacNeil, S. (2023). Generative ai in computing education: Perspectives of students and instructors. In *2023 IEEE Frontiers in Education Conference (FIE)*, pages 1–9. IEEE.

Zheng, Y. (2023). Chatgpt for teaching and learning: An experience from data science education. In *Proceedings of the 24th Annual Conference on Information Technology Education*, pages 66–72.