A Review of Generative AI in Computer Science Education: Challenges and Opportunities in Accuracy, Authenticity, and Assessment

Iman Reihanian[1(✉)], Yunfei Hou[1], Yu Chen[2], and Yifei Zheng[3]

[1] School of Computer Science and Engineering, California State University, San Bernardino, USA
`iman.reihanian1166@coyote.csusb.edu`, `hou@csusb.edu`
[2] School of Information Systems and Technology, San Jose State University, San Jose, USA
`yu.chen@sjsu.edu`
[3] University of California, San Diego, USA
`yiz152@ucsd.edu`

Abstract. This paper surveys the use of Generative AI tools, such as ChatGPT and Claude, in computer science education, focusing on key aspects of accuracy, authenticity, and assessment. Through a literature review, we highlight both the challenges and opportunities these AI tools present. While Generative AI improves efficiency and supports creative student work, it raises concerns such as AI hallucinations, error propagation, bias, and blurred lines between AI-assisted and student-authored content. Human oversight is crucial for addressing these concerns. Existing literature recommends adopting hybrid assessment models that combine AI with human evaluation, developing bias detection frameworks, and promoting AI literacy for both students and educators. Our findings suggest that the successful integration of AI requires a balanced approach, considering ethical, pedagogical, and technical factors. Future research may explore enhancing AI accuracy, preserving academic integrity, and developing adaptive models that balance creativity with precision.

Keywords: Generative AI · Computer Science Education · AI Accuracy · AI Authenticity · AI Assessment

1 Introduction

With the rise of ChatGPT and large language models, a student in today's classroom can sit at their computer, not just writing code but co-developing

it with an AI partner—a virtual assistant that suggests code snippets, refines algorithms, and troubleshoots errors. The lines between what the student creates and what the AI generates become blurred. How, then, do we ensure that the student is truly mastering the core concepts of programming? Are the algorithms being created as accurate as they appear? And most critically, how do we assess a student's progress in an educational environment where AI can both aid and overshadow the human creative process? These questions represent real challenges in today's computer science education, where concerns about AI's accuracy, authenticity, and assessment are no longer just emerging but are core issues shaping the integration of generative AI (GenAI) into the classroom.

Unlike traditional AI systems that analyze existing data to predict outcomes, Generative AI uses deep learning algorithms to generate new content, such as text, images, and code, to assist and enhance educational activities. By leveraging the ability of GenAI to generate, refine, and troubleshoot code, students can use AI models to explore creative solutions, receive instant feedback, and benefit from personalized learning. Tools like ChatGPT [1] and Claude [2] exemplify this shift by providing real-time assistance that enhances the learning experience. Retrieval-Augmented Generation (RAG) [3] is often used to add context to the large language model (LLM) based on course materials, instructor instructions, and conversation history. It is a proven approach that allows AI systems to offer better explanations and context-aware tutoring.

However, the accuracy of AI-generated content is not guaranteed, particularly when dealing with novel problems or edge cases, complicating the learning process [4,5]. Moreover, GenAI introduces concerns about the authenticity of student work. How can educators ensure that the solutions reflect the student's understanding rather than the AI's output? Educators must therefore redefine assessment frameworks to evaluate both AI-supported and independent work fairly, which requires a nuanced understanding of the interaction between technology and pedagogy [6,7].

In the context of computer science education, GenAI tools provide exciting opportunities for enhancing efficiency and creativity in teaching foundational skills such as coding, computational thinking, and problem-solving. However, the accuracy of AI-generated solutions remains a significant concern, as AI hallucinations can lead to incorrect or irrelevant information, but students may lack the knowledge to fact-check AI responses. GenAI also raises concerns about authenticity, where it is unclear whether students are truly learning to solve problems independently or relying on AI-generated solutions. Additionally, the auto-generation of assignments complicates the learning environment and raises questions about the depth of student engagement [8,9]. These emerging shifts in AI use in computer science teaching and learning pose challenges for traditional assessment methods like exams or project-based evaluations, which may no longer suffice in measuring student comprehension and skill development [10,11]. From a teacher's perspective, AI-generated content can seamlessly blend with human effort, raising concerns about the authenticity of student work. This blending also complicates the verification of accuracy, particularly in complex

programming scenarios, requiring new levels of scrutiny from educators [12,13]. As a result, designing assessment frameworks that reflect individual student skills while accounting for AI support becomes an increasingly complex task [14,15].

This survey synthesizes recent research on GenAI in computer science education, focusing on the themes of accuracy, authenticity, and assessment. While many studies have explored these themes individually, there is a notable gap in the literature regarding their interconnection and integration. By examining how these three issues interact, this paper seeks to provide a better understanding of how GenAI can transform teaching and learning in computer science. Additionally, this study summarizes areas for further research [4,5].

While this study focuses on accuracy, authenticity, and assessment, these themes are part of a broader and ongoing discussion about GenAI's impact on education. We aim to explore the relationship between AI technologies and the evolving educational landscape, particularly within computer science. The goal is to highlight both the transformative potential and the significant challenges these technologies introduce into teaching and learning processes [16,17], fostering a more informed approach to their integration in computer science curricula [18,19]. Other critical considerations, such as ethical implications, student engagement, and the development of new technical literacies, also demand attention. We propose these as areas for future work.

2   Literature Review

We reviewed 52 papers published between 2019 and 2024, sourced from academic databases such as IEEE Xplore, ACM Digital Library, and SpringerLink. These studies were categorized into three primary themes: accuracy, authenticity, and assessment. Most of the papers focused on accuracy, particularly addressing AI hallucinations and bias, while fewer discussed the challenges of assessment in AI-driven educational settings.

2.1   Accuracy

The accuracy of generative AI (GenAI) in computer science education has garnered significant attention as educators and researchers strove to incorporate AI technologies into learning environments. Accuracy is crucial for generating reliable content, offering effective educational feedback, and supporting student learning. However, there were significant concerns about AI hallucinations (i.e., the generation of factually incorrect or nonsensical information), error propagation (i.e., when initial mistakes persist and compound in subsequent outputs), bias, and the delicate balance between creativity and precision.

Implications of AI Hallucination and Error Propagation for Accuracy.
AI hallucinations and error propagation were central issues in the adoption of

GenAI models in CS education, particularly when AI-generated outputs accumulated inaccuracies over time. AI hallucinations referred to instances where AI systems produced incorrect or nonsensical information with high confidence. Wu et al. [20] highlighted the prevalence of hallucinations, reporting that 53% of the errors in their study were classified as hallucinations, which could mislead both students and educators.

Hallucinations compromised the reliability of AI-generated content, leading to concerns about student learning outcomes. For instance, Butler [21] noted that hallucinations detracted from the overall quality of AI-generated reports, potentially eroding trust in AI tools as educational aids. These errors could accumulate over time, propagating through multiple layers of content generation and significantly affecting the educational process.

Prather et al. [22] reinforced these concerns, particularly for novice learners, who might not yet have had the skills to critically assess AI-generated outputs. Lin et al. [23] further noted that models trained on biased or incomplete datasets often generated erroneous content, complicating the learning process. Becker et al. [24] offered a counterpoint, suggesting that robust feedback systems could mitigate the effects of error propagation by allowing students to cross-reference AI outputs with reliable sources.

Sharun [25] proposed that improving the diversity and quality of training datasets was a crucial step toward reducing hallucinations and enhancing the accuracy of AI systems. Additionally, Aditya [26] emphasized the role of human-in-the-loop (HITL) systems, where human oversight was integrated into the AI output generation process to provide real-time corrections and ensure accuracy. This approach aligned with the need to reduce error propagation and mitigate the negative impact of hallucinations in educational settings.

Challenges and Mitigation Strategies of Bias. Bias in AI-generated content was another critical issue affecting accuracy. Mahaini [27] discussed how GenAI tools, while adept at solving complex programming tasks, often introduced biases that could distort accuracy, particularly in assessments. This bias could disproportionately impact students from diverse backgrounds, as biased models might overlook cultural or contextual differences in problem-solving approaches. Emenike and Emenike [28] argued that reliance on biased AI-generated assessments could exacerbate educational inequalities.

To address these challenges, Gupta et al. [29] proposed hybrid models that combined human oversight with AI-generated content to enhance accuracy and mitigate bias. Walter [30] emphasized the importance of AI literacy and critical thinking, enabling students and educators to recognize and correct biased outputs. These strategies underscored the essential role of human involvement in addressing the challenges posed by bias in AI-generated content.

Challenges in Measuring and Evaluating Accuracy. Measuring and evaluating the accuracy of AI-generated content posed ongoing challenges. Francisco and Silva [31] pointed out that current accuracy metrics often failed to account

for the contextual nature of educational tasks, where solutions might not be strictly right or wrong. Similarly, Chan and Hu [32] suggested that traditional metrics focusing solely on error rates overlooked the broader educational value of AI-generated content. Becker et al. [24] proposed a holistic approach to accuracy, where metrics encompassed not only the correctness of the output but also how well AI-generated content supported learning goals. Liu [33] proposed hybrid metrics that combined traditional accuracy measures with qualitative assessments of student engagement and problem-solving skills, offering a more comprehensive picture of AI's educational effectiveness.

Creativity vs. Accuracy: A Careful Balance. The trade-off between creativity and accuracy was another common concern. Pack and Maloney [34] noted that while GenAI was celebrated for its creative solutions, this creativity often came at the expense of accuracy. In computer science education, where precise solutions were paramount, this trade-off could hinder learning. Francisco and Silva [31] advocated for a balance between creativity and accuracy, suggesting the implementation of context-aware AI systems that adjusted their outputs based on the task. For instance, while creative solutions might be encouraged in brainstorming exercises, accuracy should take precedence in tasks requiring code generation.

Zastudil et al. [35] cautioned that an overemphasis on accuracy might stifle creativity, particularly in fields like computing where innovation was crucial. They argued that students should be encouraged to experiment with AI-generated outputs, even if these were not always accurate, as such experimentation fostered deeper engagement and problem-solving skills. This highlighted the ongoing discussion of promoting creativity while maintaining accuracy in AI-generated content.

Proposed Solutions for Improving Accuracy. Several strategies had been proposed to enhance the accuracy of GenAI models in computer science education. Kung et al. [36] suggested that improving the diversity and quality of training datasets could significantly enhance model accuracy, particularly in coding tasks where precision was essential. Incorporating diverse datasets helped AI models generate outputs that were less prone to errors and biases. Liu [33] advocated for employing reinforcement learning techniques to fine-tune AI models for educational purposes, allowing models to learn from their mistakes and improve accuracy over time. This aligned with the findings of Lee and Song [37], who emphasized the role of iterative feedback in refining AI-generated outputs, particularly in programming tasks.

Human-AI interaction was crucial for improving accuracy. Mahaini [27] argued that while AI tools could generate highly accurate content, their effectiveness was enhanced when used in conjunction with human oversight. Educators could provide contextual understanding and correct errors, thereby improving the overall quality of AI-generated outputs. Walter [30] emphasized the importance of AI literacy among educators and students, suggesting that

informed users could guide AI inputs and interpret outputs more effectively, thus enhancing accuracy. However, Lee and Song [37] cautioned against over-reliance on human intervention, noting that AI models should be designed to operate as accurately as possible independently, to avoid overburdening educators.

Future Research Directions. Despite advancements in improving accuracy, several research gaps persisted. One significant gap was the lack of longitudinal studies examining the long-term impact of AI-generated inaccuracies on student learning outcomes. Chan and Hu [32] noted that while short-term studies demonstrated the benefits of GenAI in education, little was known about how inaccuracies in AI-generated content affected students' academic performance over time. Additionally, the development of bias detection and mitigation techniques for AI-generated assessments remained underexplored. While some strategies, such as those proposed by Gupta et al. [29], aimed to reduce bias, comprehensive frameworks were needed to integrate these techniques into educational systems effectively.

The balance between creativity and accuracy also required further investigation. Adaptive AI models that dynamically adjusted between fostering innovation and maintaining precision based on task requirements had been proposed, but empirical work was needed to refine these approaches. Furthermore, developing nuanced metrics to evaluate accuracy beyond simple error rates would provide educators with better insights into how AI-generated content impacted student learning.

Summary of GenAI Accuracy. Table 1 summarizes key findings on the theme of accuracy in generative AI for computer science education.

2.2   Authenticity

The theme of authenticity in the context of GenAI within computer science education has gained prominence as AI-generated outputs become increasingly prevalent. Key concerns revolve around the originality of student work, the ethical implications of AI-generated content, and the challenges educators face in distinguishing between human-authored and AI-generated submissions. This review synthesizes key findings from the literature, exploring how the integration of GenAI affects the authenticity of student work in educational settings.

AI-Generated Content and Challenges to Student Authorship. A primary concern is the risk of students presenting AI-generated content as their own, raising issues of authorship and academic integrity. Prather et al. [24] argued that as tools like ChatGPT evolved, differentiating between human-authored and AI-generated work became increasingly difficult, particularly in coding assignments where correctness is paramount. Becker et al. [24] suggested that the accessibility of AI-generated code diminished the authenticity of student

Table 1. Key Findings on Accuracy in Generative AI for Computer Science Education

Challenges	Ref. Cited	Opportunities
AI hallucinations can mislead students and educators, undermining content reliability.	[21–23, 26]	Conduct long-term studies on the cumulative effects of AI hallucinations on learning outcomes.
Error propagation amplifies inaccuracies, especially in AI-assisted programming.	[23–25, 27]	Developing real-time mitigation strategies for error propagation, particularly benefiting novice learners.
Bias in AI affects the accuracy of assessments, particularly for diverse student groups.	[28–30]	Development frameworks for bias detection and mitigation in AI-driven assessments.
Creativity vs. accuracy trade-off in AI-generated content hinders precision in coding tasks.	[32–34]	Explore adaptive models that dynamically balance creativity and accuracy in AI systems.
Human-in-the-loop (HITL) systems improve AI accuracy and support real-time corrections.	[25, 27, 30, 31]	Scaling HITL models for large educational contexts.
AI struggles with adapting to educational contexts where correctness is subjective, highlighting the need for holistic accuracy measures.	[32–34]	Developing AI metrics that better capture the contextual nature of educational tasks.

submissions. Lin et al. [23] noted that students may unintentionally incorporate AI-generated outputs into their work, blurring the line between original thought and AI assistance, which could hinder the development of critical thinking and problem-solving skills. Conversely, Zastudil et al. [35] acknowledged that AI tools could enhance learning by providing immediate feedback but emphasized the need for clear guidelines to prevent misuse.

Ethical Concerns Surrounding AI-Generated Content. Ethical considerations regarding AI-generated content are central to discussions about authenticity. Emenike and Emenike [28] highlighted that AI-generated content, particularly in written assignments, posed significant challenges for educators in assessing genuine student engagement. Over-reliance on AI tools may have diminished students' ability to produce original work. Gupta et al. [29] emphasized the difficulty of detecting AI-generated text, noting that traditional plagiarism detection

tools often failed to recognize it. They advocated for institutions to invest in AI-detection technologies to uphold academic integrity. Francisco and Silva [31] adopted a more optimistic perspective, arguing that AI could be used ethically if students were transparent about its usage, framing AI as a learning aid rather than a substitute for original work.

The Sociotechnical Perspective on AI Authenticity. Addressing the ethical implications of AI-generated content and authorship in education requires a sociotechnical perspective, where the interaction between social dynamics and technological tools is key to understanding authenticity issues. In computer science education, AI tools like ChatGPT and GitHub Copilot present unique challenges to the authenticity of student work, raising concerns about academic integrity and necessitating a holistic approach.

The authenticity of student work is increasingly questioned as AI tools generate text and code. Reliance on AI tools like ChatGPT can lead to reduced emphasis on original thought, with some researchers warning that such tools might affect creativity and critical thinking skills [38]. Educators must design assignments that promote deeper engagement and critical thinking, moving beyond tasks easily completed by AI [38].

The ethical implications of authorship are complex, as AI blurs the lines between student-authored and AI-generated content. This raises concerns about plagiarism and academic integrity [39,40]. Clear guidelines are essential to ensure students understand how to responsibly use AI, including educating them about AI's limitations and the consequences of over-reliance [38,41,42].

A sociotechnical approach can help educators distinguish between human and AI-generated content by leveraging AI's capabilities while preserving student work integrity. Assessment methods like oral presentations and reflective essays can ensure students demonstrate personal understanding [43,44]. Integrating AI literacy into curricula can also empower students to use these tools responsibly [45].

Transparency and fairness are critical as AI tools become more prevalent in education. Students with limited access to technology may face disadvantages [46,47], so educators must provide resources and training to ensure equitable access to AI technologies [46]. Ethical considerations must guide AI's implementation, preparing students to navigate technology responsibly while maintaining academic integrity [48].

Summary of GenAI Authenticity. Table 2 summarizes key findings on the theme of Authenticity in generative AI for computer science education.

2.3   Assessment

The integration of GenAI in education has significantly transformed assessment practices, particularly within computer science. AI tools are increasingly employed for grading, providing feedback, and evaluating student performance,

Table 2. Key Findings on Authenticity in Generative AI for Computer Science Education

Challenges	Ref. Cited	Opportunities
AI blurs the line between student-authored and AI-generated work, raising academic integrity concerns.	[23–25,39,43,44]	Develop clear policies to distinguish AI-assisted work from student-authored work.
Ethical concerns arise due to the difficulty in detecting AI-generated content in student submissions.	[29–31,41,42]	Create better plagiarism detection tools and institutional guidelines to ensure academic integrity.
Sociotechnical approaches can help preserve authenticity in student work.	[45–49]	Explore a combination of sociotechnical approaches to promote authenticity.
Transparent use of AI tools, combined with AI literacy education, can help mitigate ethical concerns while preserving student engagement and creativity.	[30,32,36]	Opportunity for more empirical studies on the effectiveness of AI literacy programs.
Responsible use of AI supports learning without compromising authenticity.	[23,24,36]	Explore the balance between ethical AI usage and maintaining originality in student work.

streamlining the assessment process while raising concerns about fairness, reliability, and objectivity. This section reviews key discussions on the role of AI in assessment, offering a thematic analysis of its impact on computer science education.

AI-Driven Assessment Tools: Grading and Feedback. AI-driven grading tools have gained attention for their ability to automate grading and deliver timely feedback. Lin et al. [23] noted that these tools could provide real-time feedback on programming tasks, enabling students to correct mistakes as they worked, which was particularly beneficial in large classrooms where human grading was challenging. Prather et al. [22] emphasized that AI grading systems evaluated code for correctness and efficiency, alleviating the burden on educators during routine assessments.

However, Becker et al. [24] cautioned that AI-driven grading tools might struggle with complex or creative problem-solving tasks that required deeper contextual understanding. Francisco and Silva [31] argued that while AI effectively assessed straightforward coding tasks, it fell short in evaluating innovative algorithms or creative solutions. In contrast, Walter [30] suggested that AI

assessments enhanced objectivity and consistency by minimizing human biases, such as favoritism and fatigue.

Fairness, Objectivity, and Ethical Challenges in AI-Driven Assessments. A significant debate in the literature centered on the fairness and objectivity of AI-driven assessments. Zastudil et al. [35] expressed concerns that AI tools could perpetuate biases if trained on skewed data, leading to unfair evaluations. For instance, AI systems might favor specific coding styles based on their training data. Kung et al. [36] further highlighted that students from diverse backgrounds might face unfair assessments if AI tools were not trained on inclusive datasets.

Prather et al. [22] advocated for a hybrid approach that combined AI systems with human assessments to ensure both objectivity and nuanced judgment. This model balanced the strengths of AI evaluations with the contextual understanding human graders provided. However, Mahaini [27] emphasized that human oversight was crucial, particularly for summative assessments where creativity and critical thinking had to be evaluated.

Evaluation of AI-Generated Content: Metrics and Limitations. The limitations of traditional grading metrics in assessing AI-generated content were a major concern. Mahaini [27] noted that while AI systems excelled at evaluating code correctness, they struggled with assessing creativity and originality—key learning outcomes in computer science. González-Calatayud et al. [49] argued that metrics focused solely on correctness and efficiency failed to capture essential critical thinking and problem-solving skills.

Several authors, including Liu et al. [33], advocated for new assessment frameworks that evaluated both technical and creative aspects of student work. Becker et al. [24] suggested training AI systems to recognize diverse coding styles and problem-solving approaches to avoid penalizing students for innovative thinking.

Formative and Summative Assessment: The Role of AI in Supporting Learning. AI tools could enhance both formative and summative assessments. Lee [50] highlighted that AI could provide formative feedback throughout the learning process, helping students identify areas for improvement before final evaluations, thereby reducing pressure on summative assessments. However, Mahaini [27] cautioned against relying solely on AI for summative assessments, stressing the necessity of human judgment in evaluating creativity and critical thinking. Pack and Maloney [34] also emphasized the importance of human involvement in assessing complex tasks, such as innovative coding projects, which might exceed the capabilities of AI-driven assessments.

Human-AI Collaboration in Assessments. There was a consensus that AI should complement, not replace, human involvement in assessments. Mahaini [27]

and Becker et al. [24] asserted that while AI could efficiently handle routine grading tasks, human oversight was essential for more complex evaluations, such as coding projects and creative problem-solving exercises. Francisco and Silva [31] proposed a hybrid assessment model where AI conducted an initial evaluation, followed by human review to ensure that the assessment accurately reflected the student's understanding and creativity. This model combined AI's efficiency with human insight, facilitating a more holistic evaluation process.

Research Gaps and Future Directions. Despite the growing body of literature on AI-driven assessments, several gaps remained. Notably, there was a lack of long-term studies examining the impact of AI-driven assessments on student learning outcomes. While short-term studies demonstrated the benefits of AI in providing timely feedback, limited research existed on how these tools affected students' development of critical thinking and problem-solving skills over time [22,23].

Another significant gap was the impact of AI-generated feedback on student motivation. Gupta et al. [29] noted that while AI could provide instant feedback, its effectiveness in motivating students compared to human feedback remained unclear. Future research should explore how students perceive and respond to AI-generated feedback.

Additionally, more research was needed to ensure fairness, transparency, and objectivity in AI-driven assessments. Strategies for mitigating bias in AI systems, as discussed by Zastudil et al. [35] and Kung et al. [36], required further exploration. Addressing these research gaps was essential to ensure that AI-driven assessments enhanced the learning process while upholding educational integrity.

Summary of GenAI Assessment. Table 3 summarizes key findings on the theme of Assessment in generative AI for computer science education.

3   Discussion

This review highlights the significant opportunities and challenges posed by the integration of generative AI in computer science education, particularly in terms of accuracy, authenticity, and assessment. While AI tools such as ChatGPT and GitHub Copilot offer benefits like increased efficiency, real-time feedback, and creative assistance, they also introduce critical issues like AI hallucinations, error propagation, and biases. Ensuring the accuracy of AI-generated content remains a significant challenge, particularly for novice learners, where errors and biases could adversely affect learning outcomes. Moreover, the authenticity of student work is jeopardized by the blurring of lines between AI-assisted and student-generated content, raising concerns about academic integrity. To address these challenges, educators must implement hybrid assessment models that combine human judgment with AI tools, promote AI literacy, and develop

Table 3. Key Findings on Assessment in Generative AI for Computer Science Education

Challenges	Ref. Cited	Opportunities
AI-driven grading tools improve efficiency and provide timely feedback in large classes.	[19,22,27,31]	Incorporate AI for evaluating creativity and problem-solving with human oversight.
AI grading enhances objectivity and consistency, reducing human biases such as favoritism.	[9,31]	Improve AI training with diverse datasets to minimize bias in assessments.
Hybrid AI-human evaluation balances efficiency with nuanced judgment for complex tasks.	[3,22,27]	Develop scalable hybrid models for broader use in diverse educational contexts.
New frameworks are needed to evaluate both technical and creative aspects of student work.	[11,21,22]	Create comprehensive frameworks for assessing creativity and accuracy.
AI enhances formative assessments by providing real-time feedback to guide learning.	[9,12,17,25]	Study long-term effects of AI-driven formative feedback on student learning outcomes.
Human oversight is essential for summative assessments requiring creativity and critical thinking.	[3,22,25]	Explore optimal combinations of AI and human input in summative assessments.

clear policies on the ethical use of AI. Future research should focus on refining AI systems to balance creativity with precision, enhance bias detection frameworks, and study the long-term effects of AI-generated inaccuracies on student learning.

Research Questions for Future Studies To address these gaps, future research could focus on the following questions:

- How does long-term exposure to AI-generated inaccuracies affect students' problem-solving abilities and coding skills?
- What strategies can be developed to more effectively detect and mitigate bias in AI-generated educational content?
- How can adaptive AI models be designed to balance the need for creativity and accuracy in different educational contexts?
- What are the ethical implications of using AI detection tools in academic assessments, and how can institutions ensure that these tools are used transparently and fairly?

References

1. Zhang, P.: A systematic review of ChatGPT use in k-12 education. Eur. J. Educ. **59**(2), e12599 (2023)
2. AlAli, R.: Opportunities and challenges of integrating generative artificial intelligence in education. Int. J. Religion **5**(7), 784–793 (2024)
3. Chen, L., Chen, P., Lin, Z.: Artificial intelligence in education: a review. IEEE Access **8**, 75264–75278 (2020)
4. Chan, K., Zary, N.: Applications and challenges of implementing artificial intelligence in medical education: integrative review. JMIR Med. Educ. **5**(1), e13930 (2019)
5. Niu, S., Luo, J., Niemi, H., Li, X., Lu, Y.: Teachers' and students' views of using an AI-aided educational platform for supporting teaching and learning at chinese schools. Educ. Sci. **12**(12), 858 (2022)
6. Zhai, X., Nehm, R.: AI and formative assessment: the train has left the station. J. Res. Sci. Teach. **60**(6), 1390–1398 (2023)
7. Pinto, P.H.R., et al.: Assessing the psychological impact of generative AI on computer and data science education: an exploratory study. Preprints (2023)
8. Namjoo, F., Liaghat, E., Shabaziasl, S., Modabernejad, Z., Morshedi, H.: Students' experience on self-study through AI. AI Tech Behav. Soc. Sci. **1**(3), 35–42 (2023)
9. Akgün, S., Greenhow, C.: Artificial intelligence in education: addressing ethical challenges in k-12 settings. AI Ethics **2**(3), 431–440 (2021)
10. Subaveerapandiyan, A., Paladhi, M.M., Maruthaveeran, V.: Evaluating AI literacy proficiency among LIS researchers in ASEAN. Library Hi Tech News **41**(4), 6–8 (2024)
11. García-Martínez, I., Batanero, J., Fernández-Cerero, J., León, S.: Analysing the impact of artificial intelligence and computational sciences on student performance: systematic review and meta-analysis. J. New Approaches Educ. Res. **12**(1), 171–197 (2023)
12. Cooper, G.: Examining science education in ChatGPT: an exploratory study of generative artificial intelligence. J. Sci. Educ. Technol. **32**(3), 444–452 (2023)
13. Qasem, F.: ChatGPT in scientific and academic research: future fears and reassurances. Libr. Hi Tech News **40**(3), 30–32 (2023)
14. Sapci, A., Sapci, H.: Artificial intelligence education and tools for medical and health informatics students: systematic review. JMIR Med. Educ. **6**(1), e19285 (2020)
15. Idroes, G.: Student perspectives on the role of artificial intelligence in education: a survey-based analysis. J. Educ. Manage. Learn. **1**(1), 8–15 (2023)
16. Akhmadieva, R.: Artificial intelligence in science education: a bibliometric review. Contemp. Educ. Technol. **15**(4), ep460 (2023)
17. Crompton, H., Burke D.: Artificial intelligence in higher education: the state of the field. Int. J. Educ. Technol. High. Educ. **20**(1), 22 (2023)
18. Henrick, E., et al.: Assessing the effectiveness of computer science RPPS: the case of cafécs. In: Proceedings of the 2019 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT) (2019)
19. Jang, Y.: Development of AI liberal arts curriculum for the general public. Int. J. Adv. Sci. Eng. Inf. Technol. **13**(5), 1978–1983 (2023)
20. Wu, K.: Intersection of artificial intelligence and medical education (student abstract). In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 23684–23685 (2024)

21. Butler, J.: From technical to understandable: artificial intelligence large language models improve the readability of knee radiology reports. Knee Surg. Sports Traumatol. Arthrosc. **32**(5), 1077–1086 (2024)

22. Prather, J., et al.: 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, pp. 108–159, Turku, Finland. ACM (2023)

23. Lin, X., et al.: Teachers' perceptions of teaching sustainable artificial intelligence: a design frame perspective. Sustainability **14**(13), 7811 (2022)

24. Becker, B.A. et al.: Generative AI in introductory programming. In: Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education, pp. 108–159, Turku, Finland. ACM (2023)

25. Sharun, K.: ChatGPT and artificial hallucinations in stem cell research: assessing the accuracy of generated references – A preliminary study. Ann. Med. Surg. **85**(10), 5275–5278 (2023)

26. Aditya, G.: Understanding and addressing AI hallucinations in healthcare and life sciences. Int. J. Health Sci. **7**(3), 1–11 (2024)

27. Mahaini, D.: Generative AI in computer science education: a study on academic performance. J. Comput. Sci. Educ. (2024)

28. Emenike, M.E., Emenike, B.U.: Was this title generated by ChatGPT? Considerations for artificial intelligence text-generation software programs for chemists and chemistry educators. J. Chem. Educ. **100**(4), 1413–1418 (2023)

29. Gupta, P., Ding, B., Guan, C., Ding, D.: Generative AI: a systematic review using topic modelling techniques. Data Inf. Manage. **8**(2), 100066 (2024)

30. Walter, Y.: Embracing the future of artificial intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. Int. J. Educ. Technol. High. Educ. **21**(1), 15 (2024)

31. Francisco, R., Silva, F.: Intelligent tutoring system for computer science education and the use of artificial intelligence: a literature review. In: Proceedings of the 14th International Conference on Computer Supported Education. SCITEPRESS (2022)

32. Chan, C., Hu, W.: Students' voices on generative AI: perceptions, benefits, and challenges in higher education. Int. J. Educ. Technol. High. Educ. **20**(1), 43 (2023)

33. Liu, Y.: Design and optimization of human-computer interaction system for education management based on artificial intelligence. Int. J. Interact. Mob. Technol. **18**(7), 107–124 (2024)

34. Pack, A., Maloney, J.: Using generative artificial intelligence for language education research: insights from using Openai's ChatGPT. TESOL Q. **57**(4), 1571–1582 (2023)

35. Zastudil, C., Rogalska, M., Kapp, C., Vaughn, J., MacNeil, S.: Generative AI in computing education: perspectives of students and instructors. arXiv Preprint 2023. Accessed 16 Sep 2024

36. Kung, T., et al.: Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit. Health **2**(2), e0000198 (2023)

37. Lee, S., Song, K.-S.: Teachers' and students' perceptions of AI-generated concept explanations: implications for integrating generative AI in computer science education. Comput. Educ. Artif. Intell. **7**, 100283 (2024)

38. Er, E.: Students' preference in using chatbots for academic writing. Indian J. Sci. Technol. **16**(36), 2912–2919 (2023)

39. Crawford, J., Cowling, M., Allen, K.A.: Leadership is needed for ethical ChatGPT: character, assessment, and learning using artificial intelligence (AI). J. Univ. Teach. Learn. Pract. **20**(3), 1–9 (2023)
40. Foltynek, T., Bjelobaba, S., Glendinning, I., Khan, Z., Santos, R., Pavletic, P., Kravjar, J.: ENAI recommendations on the ethical use of artificial intelligence in education. Int. J. Educ. Integr. **19**(1) (2023). https://doi.org/10.1007/s40979-023-00133-4?trk=public_post_comment-text
41. Perera, P.: AI in higher education: a literature review of ChatGPT and guidelines for responsible implementation. Int. J. Res. Innov. Soc. Sci. **8**(6), 306–314 (2023)
42. Lawrie, G.: Establishing a delicate balance in the relationship between artificial intelligence and authentic assessment in student learning. Chem. Educ. Res. Pract. **24**(2), 392–393 (2023)
43. Thanh, B.: Race with the machines: assessing the capability of generative AI in solving authentic assessments. Australas. J. Educ. Technol. **39**(5), 59–81 (2023)
44. Yu, H., Guo, Y.: Generative artificial intelligence empowers educational reform: current status, issues, and prospects. Front. Educ. **8**, 1183162 (2023)
45. Delcker J.: First-year students' AI-competence as a predictor for intended and de facto use of AI-tools for supporting learning processes in higher education. Int. J. Educ. Technol. High. Educ. **21**(1) (2024). https://doi.org/10.1186/S41239-024-00452-7
46. Williams, R., et al.: AI + ethics curricula for middle school youth: lessons learned from three project-based curricula. Int. J. Artif. Intell. Educ. **33**(2), 325–383 (2022)
47. Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O.M., Păun, D., Mihoreanu, L.: Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. Sustainability **13**(18), 10424 (2021)
48. Halaweh, M.: ChatGPT in education: strategies for responsible implementation. Contemp. Educ. Technol. **15**(2), ep421 (2023)
49. González-Calatayud, V., Prendes-Espinosa, P., Roig-Vila, R.: Artificial intelligence for student assessment: a systematic review. Appl. Sci. **11**(12), 5467 (2021)
50. Lee, J.: Development of a content framework of artificial intelligence integrated education considering ethical factors. Int. J. Adv. Sci. Eng. Inf. Technol. **14**(1), 205–213 (2024)