Education Research and Perspectives
An international journal

Volume 51, 2024, Pages 82-101

How can Generative AI Benefit Educators in Designing Assessments in Computer Science?

Ling Wang[1†]

Shen Zhan[2]

1. University of Western Australia

2. University of Melbourne

Generative Artificial Intelligence (GenAI) is transforming education, with assessment design emerging as a crucial area of innovation, particularly in computer science (CS) education. Effective assessment is critical for evaluating student competencies and guiding learning processes, yet traditional practices face significant challenges in CS education. These include the growing need for personalised evaluation amidst increasing enrolments, the intensive practice demands of programming courses, and the rapid evolution of curricula aligned with emerging technologies. This paper examines the transformative potential of GenAI tools in addressing these challenges within CS education. Through a scoping review of existing literature, we explore how GenAI can assist educators in collecting relevant assessment materials, automating exercise creation, optimizing code testing, providing interactive feedback, and leveraging learning analytics. By synthesizing evidence-based insights, this study highlights the practical applications of GenAI, demonstrating its capacity to enhance efficiency, personalization, and impact in assessment practices, ultimately advancing teaching and learning in the era of GenAI.

Introduction

Generative Artificial Intelligence (GenAI), heralded by the public release of ChatGPT, has ushered in a new era of innovation and productivity,

[†]Address for correspondence: The University of Western Australia, 35 Stirling Hwy, Crawley WA 6009. Email: ling.wang@uwa.edu.au

particularly in the educational sector (Yan, Martinez-Maldonado, et al., 2024). Powered by state-of-the-art large language models (LLMs) developed using advanced technologies like Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) and Generative Pre-trained Transformer (GPT) (Brown et al., 2020), these AI systems leverage deep learning and self-attention mechanisms to model complex linguistic dependencies (Yan, Sha, et al., 2024).

By training on extensive text data, GenAI tools can generate human-like text content that goes beyond simple prediction, creating contextually relevant and coherent outputs across various applications such as natural language generation, summarization, and interactive dialogue systems (Yigci et al., 2024). Moreover, the capabilities of GenAI have expanded dramatically, now encompassing multiple modalities including text, image, video, and speech generation (Chiu, 2023; Pande & Mishra, 2023), which have attracted significant academic research interest and become transformative tools reshaping teaching and learning (Baidoo-Anu & Owusu Ansah, 2023).

GenAI has emerged from significant advancements in computer science (CS) and is profoundly transforming its education. Like other disciplines, such as social media education (Kim, 2024) and medical education (Totlis et al., 2023), GenAI is transforming CS education by enhancing pedagogical strategies through efficient grading and content summarization (Divasón et al., 2023), facilitating personalised learning experiences with real-time feedback (Xu et al., 2023), and reducing instructors' workloads through automation of tedious tasks (Becerra et al., 2024).

In fact, CS education is arguably more affected than other disciplines due to its unique characteristics, such as the necessity for strong programming skills (Wilson & Nishimoto, 2024), the rapidly evolving technological landscape (Ajani et al., 2024), and growing enrolments of students (Lehman et al., 2021). These dynamics create new opportunities for innovation while presenting challenges that educators must navigate with the integration of GenAI. Hazzan and Erez (2024) analysed pedagogical and cognitive theories and models in CS education within the context of GenAI, proving that GenAI enhances accessibility, affordability, and efficiency in improving CS education. GenAI's transformative potential underscores the need to rethink traditional approaches to assessment in CS education.

Assessment plays a pivotal role in evaluating student competencies while also providing teachers with valuable insights into the learning processes (Paiva et al., 2022). The integration of GenAI into CS assessment requires innovative strategies that align with its goals of enhancing teaching and learning.

With the increasing number of enrolments, providing personalised and adaptive assessments in educational contexts, where a one-to-one relationship between teacher and learner is often absent, has become increasingly challenging (Paiva et al., 2022). This difficulty is further amplified in courses like computer programming, which demand regular, intense practice and involve tasks with highly diverse solution pathways, making personalised assessment nearly unfeasible (K. Wang et al., 2017).

Furthermore, rapid technological advancements have necessitated the creation of entirely new curricula and assignments in emerging fields such as deep learning (McPhail, 2021) and quantum computing (Seegerer et al., 2021), further intensifying the complexities of designing effective CS assignments. GenAI tools offer promising solutions to address these challenges. For instance, they can provide instant, interactive feedback on students' work without overwhelming instructors (Xu et al., 2023), act as intelligent tutors in coding and debugging (Sun et al., 2024), and streamline the collection and synthesis of the latest technological developments (Atkinson, 2023). These enable learners to get enough practice while freeing instructors to focus on less repetitive tasks.

Currently, a wide variety of GenAI tools are designed and being used for assessment design, and research in this area is rapidly expanding. Many studies focus on evaluating GenAI tools developed for specific course assessments (Daun & Brings, 2023; Vassiliou et al., 2023; Yan, Zhao, et al., 2024; R. Zhang et al., 2023) and analysing their impact on assessment practices, including their successes and limitations (Grandel et al., 2024;

Kooli & Yusuf, 2024; Qureshi, 2023). However, there is a notable lack of research that reviews the role of GenAI tools in assessment design for CS education.

To address this gap, this paper explores the functionality and practical applications of GenAI tools in supporting assessment in CS education. We reviewed existing research on GenAI-assisted assessment in CS assignments, with a focus on five key facets: collecting relevant materials, automating the generation of exercises, optimising code testing, supporting feedback and grading, and analysing learning analytics. This study aims to bridge this research gap by analysing how GenAI tools can support assessment design in CS education, offering insights into their practical applications to enhance teaching and learning outcomes in the era of GenAI.

Innovations in GenAI-Assisted Assessment Design for CS Education

GenAI has become a powerful tool for educators in designing effective assessments in CS, offering innovative ways to automate processes, personalise learning experiences, and enhance assessment practices. Its applications range from discovering relevant literature to inform assessment design, to creating diverse and engaging exercises, and automating code testing for greater efficiency. GenAI has also transformed feedback and grading, making them more personalised. Moreover, GenAI-driven learning analytics provide educators with valuable insights to improve teaching strategies and assessment outcomes. This section highlights the most notable advancements reported in the literature, showcasing the transformative role of GenAI in CS education.

Finding Relevant Literature to Inform Assessment Design

GenAI streamlines the process of identifying, recommending, and summarizing relevant research and emerging topics to inform the design of assessment materials in CS education. With the boosting of CS technology, the increasing volume of reports and literatures presents a challenge for educators to keep pace with the growing output. Integrating new tools, such as computational methods and GenAI, can indeed assist in managing and analysing this vast amount of information (Atkinson, 2023). These technologies can enhance the efficiency, transparency, and rigor of information gathering, providing valuable support for educators in dealing with the expanding landscape of literature (Atkinson, 2024). For example, ChatGPT can provide relevant literature through direct inquiries; although the results may sometimes include non-existent references (Daun & Brings, 2023). While this highlighted the need for further advancements in generative AI to address these limitations, it demonstrated the potential of GenAI in assessment practices.

Moreover, educators can leverage GenAI to write scripts for collecting literature. Various online search engine, such as Google's Programmable Search Engine, support programmable scripts to gather targeted files or information. Researches have demonstrated GenAI's capability to create scripts for systematically gathering documents for systematic literature reviews (Atkinson, 2024) and constructing code for a Latent Dirichlet Allocation Topic Model (LDA-TM) to support such reviews (Atkinson, 2023).

GenAI excels at summarization. Educators can use GenAI to summaries key findings from large volumes of literature to inform the development of assessment materials. Research has proved that AI-generated high-quality plain language summaries can improve access to scientific information. (Anderson et al., 2023). Various GenAI-based tools have been developed to enhance summarization capabilities. For instance, (Chen, 2023) introduced ISum, a generative summarization framework for web data, which effectively performs intent-based summarization for diverse information-seeking tasks. Yang et al. (2020) proposed a cross-language summarization model that combines recurrent neural networks and attention mechanisms to enable direct multilingual summarization without translation.

Moreover, tools like ChatGPT can extract and process information from multiple files to generate summaries, and SummaryGPT (Vassiliou et al., 2023) leverages ChatGPT to create quotient summaries that provide overviews of entire knowledge graphs for visualization. Despite these advancements, comparisons between ChatGPT-generated and human-written abstracts reveal that AI-generated summaries often fall short in quality and accuracy, with some erroneous conclusions included in the content (Cheng et al., 2023). This highlights both the potential of GenAI in summarization and the need for careful validation to ensure the reliability of its outputs.

Generating Exercises to Meet Diverse Needs

Students have varying needs when it comes to the number and type of exercises available. While some students may feel overwhelmed or disengaged by an excess of repetitive exercises that diminish cognitive excitement, others may require additional practice to master challenging topics (Daun & Brings, 2023). This diversity is particularly pronounced in large-scale educational settings, such as Massive Open Online Courses (MOOCs), where students often exhibit varying learning objectives and abilities. To accommodate these differing needs, it is essential to generate individualised exercises tailored to students' varying skill levels and learning goals. Offering students the flexibility to select assessments that align with their personal learning objectives further supports their engagement and mastery of the material.

GenAI tools excel in creating personalised learning experiences by dynamically adapting question complexity and aligning exercises with individual cognitive abilities (Niu & Xue, 2023). In CS assessments, GenAI has shown remarkable potential, generating coding exercises, debugging tasks, and algorithm design problems that cater to varying difficulty levels and learning objectives. This capability is particularly valuable in high-enrolment settings such as MOOCs, where tools like ChatGPT have been employed to develop diverse and adaptive programming assessment formats (K. Wang et al., 2017). Research has also explored automatic question generation for younger students, such as multiple-choice generator for 9th and 11th-grade CS classes (Maheen et al., 2022).

More recent studies, (Logacheva et al., 2024) reports on personalised programming exercises created with GPT-4, while (Speth et al., 2024) highlights ChatGPT's ability to synthesise Unified Modelling Language (UML)-specific information to generate software engineering exercises. Meißner et al. (2024) compared various GenAI tools, including ChatGPT, Bing AI Chat, and Google Bard, in generating programming exercises, concluding that ChatGPT and Google Bard excel at producing quality exercises. However, Speth et al. (2023) emphasised that while these AI-generated exercises are generally effective, minor manual adjustments are often needed to improve their accuracy and ensure alignment with instructional goals, reflecting the complementary role of human oversight in leveraging GenAI for exercise generation.

Optimising Code Testing

Code functional validation is a time-intensive task for CS educators, but the emergence of GenAI offers promising opportunities to streamline this process through automation. Code testing typically involves running a program or function with specific test cases, providing inputs, and comparing the actual output to the expected output. GenAI has been employed in a wide range of testing tasks, including unit test case generation, test oracle creation, and system test input generation (J. Wang et al., 2024), enabling educators and students to efficiently automate the testing of code functionality and outcomes. For instance, recent research has experimentally demonstrated ChatGPT's effectiveness in generating unit test scripts for Python programs, showing performance comparable to Pynguin, an existing unit test generator, and even surpassing it in certain cases (Bhatia et al., 2024).

Moreover, GenAI also supports program debugging and program repair (J. Wang et al., 2024). For educators, these capabilities enable the delivery of precise, actionable feedback to students. For students, these insights facilitate iterative learning, allowing them to refine their programming skills and improve their understanding of code functionality. Studies like those by (Leotta et al., 2024) revealed that leveraging GenAI models, such as ChatGPT and GitHub Copilot, for web end-to-end testing can significantly reduce development time compared to fully manual script creation. Further innovations, such as AgoneTest (Lops et al., 2024), have automated the generation and evaluation of complex class-level test suites for Java projects, showcasing the potential of GenAI in advanced code analysis and defect detection. Moreover, Boukhlif et al. (2024) conducted a comparative study to identify the most used GenAIs in software testing, their interaction mechanisms, and the various testing types automated with these tools.

Complementing this, J. Wang et al. (2024) conducted a comprehensive review of GenAI applications in software testing, offering insights from both the software testing and GenAI perspectives. Their work also serves as a platform for sharing and hosting the latest research developments in this field. Collectively, these studies underscore the growing role of GenAI in enhancing the efficiency and accuracy of automated code testing.

Source code plagiarism is a significant ethical concern in CS education, where students may submit others' work as their own. The rise of publicly accessible GenAI tools introduces new risks, enabling students to misuse these technologies to solve assignments, thereby complicating traditional approaches to plagiarism detection. Recent research has explored machine learning (ML) models to detect ChatGPT-generated code submissions (Hoq et al., 2024). Tools like GPTZero (GPTZero, 2024), designed specifically to identify ChatGPT-generated text, can categorise content as human-written, AI-generated, or mixed, given textual inputs between 250 and 5000 characters.

Similarly, GPTSniffer (Nguyen et al., 2024) has demonstrated effectiveness in distinguishing between human-written and ChatGPT-generated code under various experimental conditions. Advanced approaches, such as using pre-trained models for detecting plagiarism in Java and C/C++ code, have also been proposed (Ebrahim & Joy, 2023). Moreover, GenAI models themselves can be applied to plagiarism detection, as demonstrated by Biörck & Eriksson (2023), who investigated the use of prompt engineering with ChatGPT to identify plagiarism in simple programming exercises.

Supporting Feedback and Grading

Feedback and grading are crucial components of a balanced learning program, significantly enhancing student performance and engagement, particularly in CS education where mastering programming languages and problem-solving skills is essential. However, providing continuous, high-quality feedback is challenging, especially in large-scale or online courses, due to the time, expertise, and resources required, which are often insufficient for growing student numbers (Lee, 2023).

Recent studies have explored the potential of GenAI to address these challenges. For instance, ChatGPT has been shown to generate formative feedback for Java programming assignments, with students identifying specific improvements that could make such feedback even more useful (Z. Zhang et al., 2024). Kusam (2024) further investigated the application of ChatGPT in providing feedback for project-based learning in undergraduate web technology courses and developed an approach to enhance the general-purpose GenAI for efficient feedback. Alyoshyna (2024) discussed both the benefits and limitations of GenAI in automated feedback systems, highlighting challenges such as maintaining feedback quality, potential for bias and difficulty understanding context. The study emphasised the need for ethical and responsible integration of AI-driven feedback systems into educational platforms, ensuring they complement human oversight.

Building on these efforts to improve feedback, a related and equally pressing challenge in CS education lies in the grading process. The increasing enrollment in CS courses has led to a surge in assignments, creating significant grading challenges for educators due to time constraints and the potential for biases (Messer et al., 2024). To address these issues, automatic grading tools have gained popularity for their ability to provide consistent grades and feedback for large cohorts. These tools not only offer instant feedback to students but also significantly reduce manual grading time for instructors (Messer et al., 2024).

GenAI models like ChatGPT present a promising alternative for automating grading processes in a scalable, consistent, and minimally biased manner. For instance, Divasón et al. (2023) proposed an AI-based methodology to evaluate complex projects in CS courses, uncovering hidden knowledge in the evaluation process and identifying discrepancies, inconsistencies, and biases in grading IT projects. Research (Kooli & Yusuf, 2024) demonstrated ChatGPT's potential to deliver reliable and consistent grading for student assessments.

Similarly, Alkafaween et al. (2024) evaluated GPT-4's effectiveness in generating test suites for CS1-level programming problems as part of an autograding workflow, highlighting its ability to enhance efficiency. Grandel et al. (2024) further investigated the accuracy, precision, and recall of ChatGPT-4 compared to human graders in identifying programming mistakes. Their results revealed that, when used appropriately with prompt engineering and feature selection, ChatGPT-4 could improve objectivity and grading efficiency. These advancements highlighted the growing role of GenAI in streamlining grading processes and improving the overall quality of assessment in CS education.

Enhancing Learning Analytics

Learning analytics (LA), the process of measuring, collecting, analysing, and reporting data about learners and their contexts to enhance learning, has emerged as a pivotal focus, especially in CS education (Paiva et al., 2022). As a data-driven discipline, LA harnesses learner data to optimise educational processes (Yan, Martinez-Maldonado, et al., 2024). The integration of GenAI tools has significantly expanded the capabilities of LA, enabling the analysis of unstructured data and enriching multimodal learner interactions. GenAI tools, such as text-to-code models like OpenAI Codex (OpenAI Codex, 2024), can automatically perform data analysis with natural language prompts, while speech-to-text models like OpenAI Whisper facilitate the transcription of educational speech, bridging the gap between verbal and visual communication (Pande & Mishra, 2023).

Furthermore, Yan, Zhao, et al. (2024) introduced VizChat, a GPT-4V-based chatbot prototype that enhances learning analytics dashboards by offering contextualised explanations for visualizations, and Busropan (2024) developed a GenAI-powered LA system to analyse programming submissions, identifying common patterns and issues in student work. Despite these advances, challenges remain in effectively integrating GenAI into LA practices, as mentioned by (Yan, Martinez-Maldonado, et al. (2024), who explored both the opportunities and the complexities of applying GenAI to transform LA and educational practices.

Conclusion

GenAI has emerged as a transformative tool in education, reshaping the way educators design, deliver, and assess learning experiences. This paper explored the innovative applications of GenAI in assessment design, demonstrating its capabilities in streamlining literature collection, generating tailored exercises, optimizing code testing, automating feedback and grading, and enhancing learning analytics. These advancements offer significant potential to address long-standing challenges in CS education, such as the growing demand for personalised evaluation amid increasing enrolments, intensive programming practice, and rapidly evolving curricula aligned with emerging technologies.

While GenAI offers efficiency in assessment design, it is important to recognise its role as an advanced tool that supports rather than replaces educators and tutors. Issues such as ensuring the accuracy and reliability of AI-generated content, mitigating biases in feedback and grading, and addressing ethical concerns like plagiarism require careful consideration. GenAI can significantly enhance teaching and learning processes, but its capabilities must be guided by human oversight to align with pedagogical goals and uphold the integrity of education. Educators remain indispensable in providing expertise on cutting-edge technologies, fostering critical thinking, and offering mentorship in areas like software modelling and system administration that AI tools cannot replicate.

The integration of GenAI into assessment and learning is inevitable, and avoiding its use is inadvisable in a rapidly evolving technological landscape. Future research should focus on how educators can effectively integrate GenAI into assessments, making the use of GenAI tools an essential component of completing assignments. For instance, GenAI can act as a tutor, with tools like ChatGPT or GitHub Copilot guiding students in debugging code and exploring diverse approaches to solving programming challenges. It can also serve as an assistant, with AI-driven platforms supporting students in data analysis by generating processing syntax. This allows students to concentrate on the analytical process and interpreting results, rather than getting bogged down in the intricacies of manual data processing.

By integrating GenAI into assessment practices, educators can foster a deeper understanding of its practical applications, empowering students to use these tools responsibly and effectively. This approach not only facilitates learning and mastery of GenAI but also promotes a collaborative dynamic between human creativity and technological innovation, preparing students for the demands of a technology-driven future.

Authors

Ling Wang holds a PhD in Computer Science from the Harbin Institute of Technology, China, specialising in the optimisation of computer architecture. Currently serving as a Research Officer at the University of Western Australia, Ling's research focuses on developing advanced machine learning methods and applications, as well as optimising computer systems for enhanced performance and efficiency.

Shen (Jason) Zhan holds an Honours degree in Engineering, majoring in Civil and Environmental Engineering, from the University of Auckland, New Zealand. He has several years of experience as a structural design engineer and has contributed to academia as a lecturer and tutor at the University of Auckland, specializing in construction management and engineering courses. Jason is currently a PhD student and graduate researcher at the Teaching & Learning Lab, University of Melbourne, Australia. His research focuses on exploring how universities can better support engineering students and graduates in unlocking their full employability potential.

References

Ajani, S. N., Khobragade, P., Jadhav, P. V., Mahajan, R. A., Ganguly, B., & Parati, N. (2024). Frontiers of computing - evolutionary trends and cutting-edge technologies in computer science and next generation application. *Journal of Electrical Systems*, *20*(1s), 28–45. https://doi.org/10.52783/jes.750

Alkafaween, U., Albluwi, I., & Denny, P. (2024). *Automating autograding: large language models as test suite generators* *for introductory programming* (No. arXiv:2411.09261). arXiv. https://doi.org/10.48550/arXiv.2411.09261

Alyoshyna, Y. (2024). *AI in programming education: automated feedback systems for personalized learning* [Essay (Bachelor), University of Twente]. http://essay.utwente.nl/101073/

Anderson, L. B., Kanneganti, D., Houk, M. B., Holm, R. H., & Smith, T. (2023). Generative AI as a tool for environmental health research translation. *GeoHealth*, *7*(7), e2023GH000875. https://doi.org/10.1029/2023GH000875

Atkinson, C. F. (2023). ChatGPT and computational-based research: Benefits, drawbacks, and machine learning applications. *Discover Artificial Intelligence*, *3*(1), 42. https://doi.org/10.1007/s44163-023-00091-3

Atkinson, C. F. (2024). *Generative artificial intelligence, python, and gathering grey literature for a systematic literature review with Google's programmable search engine*. Research Square. https://doi.org/10.21203/rs.3.rs-4007725/v1

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (ai): understanding the potential benefits of chatgpt in promoting teaching and learning. *Journal of AI*, *7*(1), 52–62. https://doi.org/10.61969/jai.1337500

Becerra, Á., Mohseni, Z., Sanz, J., & Cobos, R. (2024). A generative ai-based personalized guidance tool for enhancing the feedback to mooc learners. *2024 IEEE Global Engineering Education Conference (EDUCON)*, 1–8. https://doi.org/10.1109/EDUCON60312.2024.10578809

Bhatia, S., Gandhi, T., Kumar, D., & Jalote, P. (2024). Unit test generation using generative ai: a comparative performance analysis of autogeneration tools. *Proceedings of the 1st International Workshop on Large Language Models for Code*, 54–61. https://doi.org/10.1145/3643795.3648396

Biörck, J., & Eriksson, S. (2023). *Detecting Plagiarism with ChatGPT Using Prompt Engineering* [KTH Royal Institute of Technology]. https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-330996

Boukhlif, M., Kharmoum, N., & Hanine, M. (2024). LLMs for intelligent software testing: a comparative study. *Proceedings* *of the 7th International Conference on Networking, Intelligent Systems and Security*, 1–8. https://doi.org/10.1145/3659677.3659749

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Busropan, C. (2024). *Enhancing programming education through llm-based learning analytics and interventions* [Master Thesis]. Delft University of Technology.

Chen, H.-Y. (2023). Generative AI frameworks for web for good: Food pantry information seeking as an example. *2023 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 351–356. https://doi.org/10.1109/WI-IAT59888.2023.00057

Cheng, S.-L., Tsai, S.-J., Bai, Y.-M., Ko, C.-H., Hsu, C.-W., Yang, F.-C., Tsai, C.-K., Tu, Y.-K., Yang, S.-N., Tseng, P.-T., Hsu, T.-W., Liang, C.-S., & Su, K.-P. (2023). Comparisons of quality, correctness, and similarity between chatgpt-generated and human-written abstracts for basic research: cross-sectional study. *Journal of Medical Internet Research*, *25*(1), e51229. https://doi.org/10.2196/51229

Chiu, T. K. F. (2023). The impact of Generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 1–17. https://doi.org/10.1080/10494820.2023.2253861

Daun, M., & Brings, J. (2023). How ChatGPT will change software engineering education. *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, 110–116. https://doi.org/10.1145/3587102.3588815

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186.

Divasón, J., Martínez-de-Pisón, F. J., Romero, A., & Sáenz-de-Cabezón, E. (2023). Artificial intelligence models for assessing the evaluation process of complex student projects. *IEEE Transactions on Learning Technologies*, *16*(5), 694–707. IEEE Transactions on Learning Technologies. https://doi.org/10.1109/TLT.2023.3246589

Ebrahim, F., & Joy, M. (2023). Source code plagiarism detection with pre-trained model embeddings and automated machine learning. In R. Mitkov & G. Angelova (Eds.), *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing* (pp. 301–309). INCOMA Ltd., Shoumen, Bulgaria. https://aclanthology.org/2023.ranlp-1.34

Grandel, S., Schmidt, D. C., & Leach, K. (2024). Applying large language models to enhance the assessment of parallel functional programming assignments. *Proceedings of the 1st International Workshop on Large Language Models for Code*, 102–110. https://doi.org/10.1145/3643795.3648375

Hazzan, O., & Erez, Y. (2024). *Rethinking computer science education in the age of Genai* (SSRN Scholarly Paper No. 4992837). Social Science Research Network. https://doi.org/10.2139/ssrn.4992837

Hoq, M., Shi, Y., Leinonen, J., Babalola, D., Lynch, C., Price, T., & Akram, B. (2024). Detecting ChatGPT-generated code submissions in a cs1 course using machine learning models. *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 526–532. https://doi.org/10.1145/3626252.3630826

Kim, P. W. (2024). A framework to overcome the dark side of generative artificial intelligence (gai) like ChatGPT in social media and education. *IEEE Transactions on Computational Social Systems*, *11*(4), 5266–5274. IEEE Transactions on Computational Social Systems. https://doi.org/10.1109/TCSS.2023.3315237

Kooli, C., & Yusuf, N. (2024). Transforming educational assessment: insights into the use of ChatGPT and large language models in grading. *International Journal of Human–Computer Interaction*, *0*(0), 1–12. https://doi.org/10.1080/10447318.2024.2338330

Kusam, V. A. (2024). *Generative-AI Assisted feedback provisioning for project-based learning in cs education* [Master Thesis] University of Michigan–Dearborn. https://doi.org/10.7302/22651

Lee, A. V. Y. (2023). Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation. *Studies in Educational Evaluation*, *77*, 101250. https://doi.org/10.1016/j.stueduc.2023.101250

Lehman, K. J., Karpicz, J. R., Rozhenkova, V., Harris, J., & Nakajima, T. M. (2021). Growing enrollments require us to do more: perspectives on broadening participation during an undergraduate computing enrollment boom. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 809–815. https://doi.org/10.1145/3408877.3432370

Leotta, M., Yousaf, H. Z., Ricca, F., & Garcia, B. (2024). AI-generated test scripts for web e2e testing with ChatGPT and copilot: a preliminary study. *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering*, 339–344. https://doi.org/10.1145/3661167.3661192

Logacheva, E., Hellas, A., Prather, J., Sarsa, S., & Leinonen, J. (2024). Evaluating contextually personalized programming exercises created with generative AI. *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1*, *1*, 95–113. https://doi.org/10.1145/3632620.3671103

Lops, A., Narducci, F., Ragone, A., & Trizio, M. (2024). AgoneTest: automated creation and assessment of unit tests leveraging large language models. *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering*, 2440–2441. https://doi.org/10.1145/3691620.3695318

Maheen, F., Asif, M., Ahmad, H., Ahmad, S., Alturise, F., Asiry, O., & Ghadi, Y. Y. (2022). Automatic computer science domain multiple-choice questions generation based on informative sentences. *PeerJ Computer Science*, *8*, e1010. https://doi.org/10.7717/peerj-cs.1010

McPhail, G. (2021). The search for deep learning: A curriculum coherence model. *Journal of Curriculum Studies*, *53*(4), 420–434. https://doi.org/10.1080/00220272.2020.1748231

Meißner, N., Speth, S., & Becker, S. (2024). Automated programming exercise generation in the era of large language models. *2024 36th International Conference on Software Engineering Education and Training (CSEE&T)*, 1–5. https://doi.org/10.1109/CSEET62301.2024.10662984

Messer, M., Brown, N. C. C., Kölling, M., & Shi, M. (2024). Automated grading and feedback tools for programming education: A systematic review. *ACM Transactions on Computing Education*, *24*(1), 1–43. https://doi.org/10.1145/3636515

Nguyen, P. T., Di Rocco, J., Di Sipio, C., Rubei, R., Di Ruscio, D., & Di Penta, M. (2024). GPTSniffer: A CodeBERT-based classifier to detect source code written by ChatGPT. *Journal of Systems and Software*, *214*, 112059. https://doi.org/10.1016/j.jss.2024.112059

Niu, Y., & Xue, H. (2023). Exercise generation and student cognitive ability research based on ChatGPT and Rasch model. *IEEE Access*, *11*, 116695–116705. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3325741

OpenAI Codex. (2024). OpenAI Codex. https://openai.com/index/openai-codex/

Paiva, J. C., Leal, J. P., & Figueira, Á. (2022). Automated assessment in computer science education: a state-of-the-art review. *ACM Transactions on Computing Education (TOCE)*, *22*(3), No. 34, 1-40. https://doi.org/10.1145/3513140

Pande, A., & Mishra, D. (2023). The synergy between a humanoid robot and whisper: bridging a gap in education. *Electronics*, *12*(19), Article 19. https://doi.org/10.3390/electronics12193995

Qureshi, B. (2023). ChatGPT in computer science curriculum assessment: an analysis of its successes and shortcomings. *Proceedings of the 2023 9th International Conference on E-Society, e-Learning and e-Technologies*, 7–13. https://doi.org/10.1145/3613944.3613946

Seegerer, S., Michaeli, T., & Romeike, R. (2021). Quantum computing as a topic in computer science education. *Proceedings of the*

16th Workshop in Primary and Secondary Computing Education, 1–6. https://doi.org/10.1145/3481312.3481348

Speth, S., Meißner, N., & Becker, S. (2023). Investigating the use of ai-generated exercises for beginner and intermediate programming courses: a ChatGPT case study. *2023 IEEE 35th International Conference on Software Engineering Education and Training (CSEE&T)*, 142–146. https://doi.org/10.1109/CSEET58097.2023.00030

Speth, S., Meißner, N., & Becker, S. (2024). ChatGPT's aptitude in utilizing uml diagrams for software engineering exercise generation. *2024 36th International Conference on Software Engineering Education and Training (CSEE&T)*, 1–5. https://doi.org/10.1109/CSEET62301.2024.10663027

Sun, D., Boudouaia, A., Zhu, C., & 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. https://doi.org/10.1186/s41239-024-00446-5

Totlis, T., Natsis, K., Filos, D., Ediaroglou, V., Mantzou, N., Duparc, F., & Piagkou, M. (2023). The potential role of ChatGPT and artificial intelligence in anatomy education: A conversation with ChatGPT. *Surgical and Radiologic Anatomy*, *45*(10), 1321–1329. https://doi.org/10.1007/s00276-023-03229-1

Vassiliou, G., Papadakis, N., & Kondylakis, H. (2023). SummaryGPT: Leveraging ChatGPT for Summarizing Knowledge Graphs. In C. Pesquita, H. Skaf-Molli, V. Efthymiou, S. Kirrane, A. Ngonga, D. Collarana, R. Cerqueira, M. Alam, C. Trojahn, & S. Hertling (Eds.), *The Semantic Web: ESWC 2023 Satellite Events* (pp. 164–168). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-43458-7_31

Wang, J., Huang, Y., Chen, C., Liu, Z., Wang, S., & Wang, Q. (2024). Software testing with large language models: Survey, landscape, and vision. *IEEE Transactions on Software Engineering*, *50*(4), 911–936. IEEE Transactions on Software Engineering. https://doi.org/10.1109/TSE.2024.3368208

Wang, K., Lin, B., Rettig, B., Pardi, P., & Singh, R. (2017). Data-driven feedback generator for online programing courses.

Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale, 257–260. https://doi.org/10.1145/3051457.3053999

Wilson, S. E., & Nishimoto, M. (2024). Assessing learning of computer programing skills in the age of generative artificial intelligence. *Journal of Biomechanical Engineering*, *146*(51003). https://doi.org/10.1115/1.4064364

Xu, W., Meng, J., Raja, S. K. S., Priya, M. P., & Kiruthiga Devi, M. (2023). Artificial intelligence in constructing personalized and accurate feedback systems for students. *International Journal of Modeling, Simulation, and Scientific Computing*, *14*(1), 2341001. https://doi.org/10.1142/S1793962323410015

Yan, L., Martinez-Maldonado, R., & Gasevic, D. (2024). Generative artificial intelligence in learning analytics: contextualising opportunities and challenges through the learning analytics cycle. *Proceedings of the 14th Learning Analytics and Knowledge Conference*, 101–111. https://doi.org/10.1145/3636555.3636856

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & 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. https://doi.org/10.1111/bjet.13370

Yan, L., Zhao, L., Echeverria, V., Jin, Y., Alfredo, R., Li, X., Gaševi'c, D., & Martinez-Maldonado, R. (2024). VizChat: Enhancing learning analytics dashboards with contextualised explanations using multimodal generative AI Chatbots. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), *Artificial Intelligence in Education* (pp. 180–193). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64299-9_13

Yang, F., Cui, R., Yi, Z., & Zhao, Y. (2020). Cross-language generative automatic summarization based on attention mechanism. In G. Wang, X. Lin, J. Hendler, W. Song, Z. Xu, & G. Liu (Eds.), *Web Information Systems and Applications* (pp. 236–247). Springer International Publishing. https://doi.org/10.1007/978-3-030-60029-7_22

Yigci, D., Eryilmaz, M., Yetisen, A. K., Tasoglu, S., & Ozcan, A. (2024). Large language model-based chatbots in higher education. *Advanced Intelligent Systems*, *n/a*(n/a), 1–16. https://doi.org/10.1002/aisy.202400429

Zhang, R., Qiu, Y., & Li, Y. (2023). An empirical study on human-machine collaborative MOOC learning interaction empowered by generative AI. *2023 International Symposium on Educational Technology (ISET)*, 116–120. https://doi.org/10.1109/ISET58841.2023.00031

Zhang, Z., Dong, Z., Shi, Y., Price, T., Matsuda, N., & Xu, D. (2024). Students' perceptions and preferences of generative artificial intelligence feedback for programming. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(21), Article 21. https://doi.org/10.1609/aaai.v38i21.30372