Mediterranean Smart Cities Conference (MSCC 2024)
2-4 May, 2024, Martil, Morocco

Rethinking Assessment Methodologies in the Era of Artificial Intelligence: Expanding Beyond ChatGPT's Scope

Muhammad Nadeem
College of Engineering and Technology
American University of the Middle East
Egaila, Kuwait
muhammad.nadeem@aum.edu.kw

Wael A. Farag
College of Engineering and Technology
American University of the Middle East
Egaila, Kuwait
wael.farag@aum.edu.kw

Magdy Helal
College of Engineering and Technology
American University of the Middle East
Egaila, Kuwait
magdy.helal@aum.edu.kw

***Abstract*— Historically and in the present world, humans tend to seek innovations and new gadgets for convenience and to enhance daily life. ChatGPT is an advanced AI language model that is based on the Generative Pretrained Transformer (GPT) architecture fabricated to produce human-like text by forecasting subsequent words in a sequence, given all the previous words within some text. ChatGPT is a double-edged sword, potentially enhancing the learning process on one hand and impeding it on the other. This paper examines challenges and limitations linked to students' reliance on ChatGPT in solving assessments, urging a reconsideration of assessment methodologies. Acknowledging the growing influence of artificial intelligence (AI), the study recommends a reevaluation of these methodologies to enable instructors to design assessments that restrain students from dependence on ChatGPT. It explores constraints in ChatGPT's ability to offer comprehensive and contextually accurate responses, emphasizing the necessity of redesigning assessments to minimize the potential misuse of the model by students. Through an exploration of these limitations, the paper advocates for a nuanced understanding of the dynamic between AI and assessment practices. The findings underscore the significance of adapting assessment methodologies to empower instructors in crafting evaluations that discourage reliance on ChatGPT, providing insights into potential strategies for achieving this objective.**

Keywords*— *ChatGPT, Assessment methodologies, Evaluation strategies, Academic integrity, Artificial intelligence (AI).

I. INTRODUCTION

Historically and in the present world, humans tend to seek innovations for convenience, resulting in various inventions and diverse public reactions. Throughout history and contemporary times, people consistently seek new gadgets to enhance daily life. This pursuit has resulted in numerous inventions, sparking diverse public reactions, ranging from simple tools like calculators to complex innovations such as self-driving cars. The introduction of the scientific calculator in classrooms in 1975 elicited mixed responses. While many educators supported its use, claiming it fostered a better problem-solving attitude and logical thinking among students, a survey indicated that 72% of teachers and parents vehemently argued that it would erode their children's basic mathematics skills. Despite ongoing doubts about the extensive use of calculators in classrooms, educational organizations favored their integration, emphasizing their importance in curriculums and assessments. However, as late as 1993, 40% of parents remained hesitant It was not until the late 1990s that schools and various organizations fully embraced calculators, as exemplified by changes in standardized testing like the SAT math portion, which was adjusted in 1994 to accommodate the use of calculators. The American National Council of Teachers of Mathematics [1] promoted the use of calculators in middle schools since the 1980s, nonetheless, the discussions continued for decades after that [2].

ChatGPT [5] is an advanced AI language model that is based on the Generative Pretrained Transformer (GPT) architecture fabricated to produce human-like text by forecasting subsequent words in a sequence, given all the previous words within some text. Swiftly, ChatGPT found its way to academia and students used it to provide "help" in their assigned assessments and other works. This obviously hurts the student's learning and interrupts the integrity of the assessment process. Although plagiarism applications have developed tools to identify GPT-generated text, the threat is still serious. ChatGPT is a double sword where on one hand it may become handy in the learning process and, on the other hand, it may hinder this process. Academicians should adapt and develop their ways to continue to perform assessments of the student learning effectively and correctly.

OpenAI's CEO was quoted to have said, in response to concerns about plagiarism based on using ChatGPT: "We adapted to calculators and changed what we tested for in math class" [6]. It is easily asserted that ChatGPT offers unprecedented opportunities for supporting course development as well as student assessments and evaluation. Educators must adapt and embrace technology to have it seamlessly integrated into the pedagogy of the teaching and learning processes. The takeaway is that students are determined to embrace artificial intelligence (AI) technology, even in the face of resistance from academia. This technology is here to stay, and educators would benefit from proactively adapting to these changes as soon as possible. The good news is that the resilient nature of human beings has helped them cope with many such 'disruptive' technologies in higher education in the past.

Several research studies have reported the significance of assessment and its effectiveness as a learning tool in higher education as it enhances motivation and performance and fosters learning [7]. Depending on the goal of evaluation in the course's learning environment, there are several ways to conduct an assessment. It might be an electronic assessment, formative, summative, or diagnostic. Written assignments are academic projects that call for students to use critical thinking abilities, effectively explain ideas via writing, and express their grasp of a subject. These assignments have a variety of formats and functions within the educational setting. some common types of written assignments include essays, research papers, literature reviews, reports, and case studies.

Evaluations are essential for assessing students' subject-matter knowledge and skills and determining how well they have learned the desired material.

There is debate around the use of ChatGPT in academic settings, with serious questions about academic integrity and the possibility of AI-assisted cheating [8]. There are now no established guidelines in academic communities on how these innovations should be used in the classroom. They have threatened the traditional assessment method by unveiling the vulnerability when attacked by AI generative tools. A major threat to the integrity of traditional university assessment artifacts, which are mostly written documents, is the widespread use of text-generating AI tools such as ChatGPT. Academic integrity and AI-assisted cheating are key concerns as AI essay-writing platforms abet researchers and students create essays, articles, research, and assignments on their behalf. Distinguishing whether a specific artifact is crafted by the student or generated by AI is particularly challenging. This makes it hard to measure student understanding accurately posing a great threat to the validity of the assessment. It has been stated that using ChatGPT impacts students' development of critical thinking, problem-solving, imagination, and research abilities in a negative way [9].

There is a pressing need for methods or techniques that could ensure that learners and graduates have learned the skills so that they become competent and secure professionals. The ability of ChatGPT to produce high-quality work that can even evade plagiarism detection technologies is marking the beginning of the end of the era of written evaluation. A suggested remedy is to bring back the traditional paper-based exam written in isolation where only calculators are allowed, which will offer an effective defense against cheating. It will bring back issues such as logistics, commuting to the exam venue, creating a silent examination-friendly environment, and creating rubrics for uniform marking issues, etc. – reasons for moving away from written exams. Furthermore, they promote cramming instead of fostering critical thinking and assess only limited knowledge. Such techniques can never be an alternative to written assessments [11]. In short, there is no way back, and designing more sophisticated assessments instead is the only solution. Now, in the era of AI, educators must create exams that both promote and rigorously measure learning.

Every time there is a breakthrough in technology that can help students, they are quick to embrace it despite all the hurdles created by the teachers, and in the end, they are the ones who must adapt, not the students. Therefore, the sooner we learn to live with it, the better it is for us. Using proctored tests, which forbid utilizing the internet, with a high percentage allocated to it will discourage the use of ChatGPT. Although this is good for assessing students, this restricts the use of assessment for learning. Furthermore, the main purpose of university education is to produce a skilled workforce who can communicate and work as a team. These types of skills cannot be assessed through isolated exams. The transformative impact of such AI technologies has underlined the necessity for a paradigm shift in assessment practice moving to a more dynamic and adaptive approach to evaluate students in the presence of technologies capable of producing answers for the students.

Research on the use and potential misuse of AI in academia is still in its early stages, necessitating further contributions to enrich the existing body of literature. While much attention has been given to literature discussing assessment redesign utilizing AI, there remains a gap in research focusing on identifying and addressing the challenges and limitations of students relying on ChatGPT for assessments. Emphasizing the potential drawbacks of excessive dependence on AI technology is essential for comprehensive understanding and informed decision-making in educational contexts. This study fills this gap and complements the existing literature by providing recommendations or ideas on how to design written assessments to minimize the potential for misuse of AI tools. The authors propose adjustments that would address or mitigate the risks associated with the inappropriate use of AI in the assessment process. The intention is to offer constructive suggestions for redesigning assessment frameworks, ensuring they are less susceptible to the negative impacts or manipulation that can arise from the use of AI.

II. Assessment Design

Assessments are measures used for inculcating learning as well as evaluating students' mastery of the knowledge and skills. These measures can be used directly or indirectly at the program level or class level. Course-level measures include exams, tests, projects, essays, portfolios, presentations, etc. Furthermore, these assessments can be objective, have right or wrong answers, or be subjective and have varied responses with more than one correct response. These are normally used to assess or evaluate the complicated and qualitative aspects of students' performance. A careful approach is needed as outlined in [12] to achieve the goals of educating and improving the performance of the learners.

An assignment is a task or piece of academic work that gives students the chance to practice, study, and demonstrate that they have met learning objectives. It provides instructors with concrete proof to determine whether pupils have accomplished the objectives. The use of assignments as an assessment tool is very beneficial for assessing higher-order cognitive capabilities and the real-world application of certain knowledge or skills [13].

When designing an assignment, it is important to ensure reliability, transparency, and validity. Reliability means that the assessment is graded in a consistent way that can be ensured by creating a rubric that should be followed by evaluators all the time. Transparency means what they need to provide as evidence when solving assignments and how they will be assessed. This information can be provided along with the assignment. The final factor is validity which ensures that assignment results are aligned with the learning outcomes and deviation will threaten the validity of the assignment. A careful design of the assignment may ensure that these threats are mitigated fully. However, the advent and use of generative AI technology have considerable consequences for the methods employed in evaluating and training students and trainees. It has threatened the validity as we can no longer rely on submitted "artifacts" to assess student-learning outcomes [14].

III. ChatGPT and LLMs

Generative AI is a technology that produces text, images, audio, synthetic data, etc. based on the context provided in the form of prompts. It is powered by pre-trained language models known as foundation models (FM). Large language models (LLMs) are a subset of FM trained across many different languages. Several writing tools such as ChatGPT and Google Bard have been released recently that are getting more and more accessible generating text in response to the user's request.

OpenAI [15] is an American AI research organization that surprised the world late in 2022 with the launching of the ChatGPT platform, also known as Chat Generative Pre-trained Transformer. It is trained on a broad spectrum of internet text and can perform a variety of language tasks such as answering questions, writing essays, summarizing text, translating languages, and even engaging in casual conversation (ChatGPT described by itself). ChatGPT is built upon specific GPT foundation models such as GPT-3.5 and GPT-4, which have been fine-tuned to cater specifically to conversational use cases. A ChatGPT subscription plan, ChatGPT Plus, provides priority access to new features, quicker response times, and availability even during periods of heavy demand. The ChatGPT can be integrated into user applications through APIs. As of November 2023, it is possible to customize ChatGPT for specific purposes. All these variants review the information entered by the user thus privacy of the contents is not guaranteed. A standard GPT-4 model provides up to 8,000 tokens for the context extendible to 32000. Also, it can interpret static images (in .png, .jpg, or .gif format) up to the size of 20MB but not the videos.

ChatGPT is not the only language model, many other such models are available in the market such as Bingo AI, Jasper AI, Bard AI, and Copilot to name a few, each having its unique feature which makes it better than others. Bard has access to the most recent studies, giving him additional data to collect information in real time.

IV. Proposed Methodologies for Design Assessments

To reduce students' reliance on ChatGPT in solving assessments, instructors can implement various assessment methodologies that encourage critical thinking, and application of knowledge, and discourage dependence on AI-generated content. In this paper, some proposed strategies are being crafted as follows:

A. Context-based question

Context-based questions are those that depend on the details or context of the question to yield an appropriate and pertinent answer. One of ChatGPT's drawbacks is that it has trouble understanding context which can be used to our advantage. Adding more context makes it harder for ChatGPT to generate accurate responses. For example, sarcasm in a question might cause ChatGPT to misinterpret its intended meaning and respond in a way that is not consistent with the user's expectations. Besides, there is a limit on the token which restricts the size of the response. Through the integration of extra context, you may create useful assessments that mimic real-world development jobs inside existing codebases and make them authentic [16]. This approach significantly improves contextual comprehension while limiting the duration of the tasks that candidates must complete. Authentic assessments impact deep learning and higher-order cognitive skills in a positive way [17].

B. Present scenarios using large images and video

Scenario-based assessments involve presenting individuals with realistic situations or scenarios that simulate authentic, context-specific challenges. These assessments are designed to evaluate a person's ability to apply their knowledge, skills, and decision-making abilities in practical or real-world contexts, instead of relying solely on theoretical or abstract questions, scenario-based assessments immerse individuals in scenarios that mirror the situations they might encounter in their professional or academic endeavors. The assessment typically requires the individual to analyze the situation, make informed decisions, and justify their choices based on relevant knowledge and skills. GPT-4 (Plus and ChatGPT Enterprise plans) has the capability of interpreting images of up to the size of 20MB. Presenting a scenario in the form of a video or an image of a size larger than 20MB will restrict the use of ChatGPT to solve the assessments.

C. Use assessment task requires students to draw figures

Designing an assessment task that requires students to draw a figure is an effective way to encourage active engagement and prevent reliance on ChatGPT. However, it involves the definition of clear learning objectives and the selection of appropriate drawing tools. Precise instructions aligned with course content should be provided, and real-world context should be integrated to make the task relevant. Evaluation criteria focusing on accuracy, creativity, and adherence to instructions must be set. Submission formats are usually specified, and time constraints are considered while encouraging creativity within individual efforts. Optionally, resources and references could be provided to guide students, with an emphasis on the learning process alongside the final drawing. Constructive feedback and reflections are offered after the assessment, and the incorporation of peer review for diverse perspectives can be considered. Authenticity is ensured in evaluating drawn figures, aligning assessments with learning objectives, and promoting active student engagement while minimizing reliance on external sources like ChatGPT. In an interesting study conducted by Beisler and Medaille, students were asked to draw a figure depicting the full process of completing the assignment aiming to study their behaviors in carrying out the task [18]. This will help in finding out whether students got any help from the ChatGPT or not.

D. Role-playing

One way to combat improper use of ChatGPT is to put yourself in the shoes of a student. Once an assessment is designed, the instructor should play the role of a student and start solving it by taking the help of ChatGPT in the same fashion as students do it. In this, they will be aware of the AI tool’s capacity to complete the assessment tasks. This insight will help him adapt the assessments that are resistant to the AI tools.

E. Interviews

One way to design an authentic writing assessment is to include oral work ensuring that learning takes place and students achieve the learning outcomes. It is suggested that oral presentations and synchronous interviews should be incorporated as one of the grading tools. This interview gives insight into the trade-offs taken into consideration and permits a more thorough examination of the judgments made. This method makes it easier to comprehend the candidate's reasoning and evaluates their comprehension—especially if they used ChatGPT support. Furthermore, it helps to prolong the evaluation and facilitate the candidate's interview experience.

F. Ask for deep insight and detail about a particular topic or domain

ChatGPT suffers from an inability to provide insight into a particular topic and tends to be more verbose. This is more

comprehensive than concise and tends to provide lengthy answers indirectly. This can be used to the instructor's advantage when designing written assignments which discourages the usage of these tools. The instructor should ask about topics that are more specific to a domain, asking for more in-depth information in fewer words. He/she might also use a very specific task that has been completed in class as the foundation for a writing project. AI is unlikely to know about and will not be able to write this assignment for the students. Also, having students summarize the main ideas of the lesson would also meet the requirements.

G. Use specific references

Research assignments are often used to collect information, analyze, and report your findings using artifacts such as presentations or research papers. When asking the students to complete writing assignments, limit the students to a specific set of literature by providing references instead of leaving them open-ended. As of right now, the AI struggles to cite sources consistently or to include sources as references or in-text citations.

H. Criticize or review instead of answering the questions

Students normally use the prompts to answer a particular question in the assignment. If the assignment settings allow, students can be provided with the questions as well as answers and they should be asked to critically review the answers identifying the flaws and reasoning them. This strategy will guarantee that students will deeply engage in the material and promote higher-order thinking. The teacher should be very considerate when designing such assignments. Firstly, it should focus on inculcating higher-level thinking instead of introducing surface-level shortcomings. Guidelines should be provided on how to address the problem systematically and what is expected from them. From this type of assignment, students benefit from all the advantages offered by peer review techniques. Do not write code but ask about the quality of the code. Modern AI generative tools can produce code and advise on programming. This level of help makes the coding assessment less authentic, therefore, redesign is required in a way that makes it less susceptible to AI tools. For example, students may be asked to explain the code instead of writing the code. Explaining the code requires a very deep knowledge of programming and language models are not equipped with such capabilities yet. Similarly, students can be asked to review the code and identify syntax as well as semantic errors and it is not easy for AI tools to perform such tasks which require a hands-on approach.

I. Embedd calculations

Using calculation in assessments has been an effective way to develop higher-level thinking skills among students. Now such calculations can be performed by ChatGPT immediately and students can solve such assessments that are error-free without any understanding and being challenged intellectually. This means that it would not be feasible to provide exercises in the manner that has been done for years. Instead of asking the students to perform calculations, the instructors should use a flipped approach. They can show students an incorrect calculation or proof that was supplied by ChatGPT and ask them to identify the error or suggest a solution. Alternatively, they can provide students with an activity that contains the provided solution and can be solved using the provided technique. Students can then be asked to create an exercise that will lead to the provided solution.

J. ChatGPT lacks information about recent developments

While ChatGPT is a powerful language model, it indeed has limitations when it comes to providing information about very recent developments. As of the last update in early 2022, it may not be aware of events or information that occurred post that time. Users should always cross-verify and supplement their responses with the latest and most reliable sources, especially when seeking information on rapidly evolving topics. Continuous updates and improvements in AI technology may address this limitation in future versions, enhancing the model's ability to provide more up-to-date information. It is essential to use ChatGPT as a supplementary tool and not a sole source for the latest news or developments.

K. Self-reflection on learning in the subject

Developing a self-reflection assessment on learning in a certain subject requires careful planning and consideration. A self-reflection assessment involves defining its purpose and key learning objectives. Instructions should be clear, with reflective prompts aligned to Bloom's Taxonomy for critical thinking [19]. Examples should be included, linking reflections to real-world applications should be encouraged, and a multimodal approach [20] for diverse expression should be employed. A self-assessment component should be integrated and ample time for thoughtful reflection should be allocated. Feedback should be provided and reflections as recurring elements in the curriculum should be emphasized. Diverse perspectives must be respected, and the effectiveness of the assessment should be continuously evaluated and adjusted. The process aims not just to evaluate students but also to foster a deeper understanding of their learning processes and encourage continuous improvement.

L. Use game-based learning

Digital Game-Based Learning (DGBL) **Error! Reference source not found.** can be employed to assess students in a manner that extends beyond ChatGPT's capabilities. Learning objectives can be defined, and an appropriate game format should be selected, aligning with the course content. Educational content can be incorporated seamlessly into the game, mirroring curriculum concepts. Clear instructions should be provided, ensuring students can comprehend the rules and evaluation parameters. Problem-solving skills can be assessed through game scenarios requiring critical thinking. Opportunities for collaboration should be integrated, including multiplayer settings or team-based challenges. Students' progression and achievements can be tracked in real time, facilitating continuous improvement. Immediate feedback mechanisms within the game enhance the learning experience. Assessments are seamlessly integrated into gameplay, contributing to the overall game narrative. Motivation and engagement should be fostered through engaging storylines, rewards, and competition. Reflection and debriefing after the game-based assessment can encourage students to articulate lessons learned and insights gained. Collaboration skills within the game environment should be assessed, evaluating communication and teamwork effectiveness. The game-based assessment should evaluate not only the process but also the learning outcomes, gauging students' internalization, and application of embedded concepts.

M. Responses to the feedback

Creating an assessment based on responses to feedback on a specific subject can enhance students' learning experiences.

To develop such kinds of assessments, clear objectives should be set, and expectations must be communicated. Constructive feedback from earlier assessments can be used as a foundation, and reflective prompts that encourage critical analysis can be created. Action plans and self-assessment components should be incorporated, linking the feedback to real-world applications. Exemplars should be provided to guide students, and time for revision should be allocated, fostering an iterative learning process. Ongoing support can be offered, communication skills should be assessed, and consideration can be given to incorporating peer review for a collaborative environment. Feedback on the assessment process should be collected to refine future implementations, and a growth mindset should be emphasized, positioning feedback as a tool for improvement rather than a measure of success. Also, mandating the students to submit preliminary drafts of their assignments for assessment prior to the final submission will enable faculty to detect potential instances of AI-generated content and provide constructive feedback [20]. This will force the student to comprehend the feedback and engage with the material in a meaningful way to improve their final submission.

N. Context-specific assignments

Developing context-specific assignments involves tailoring tasks to the unique characteristics and objectives of a particular context, such as a course, subject area, or real-world scenario. To develop such assignments, a thorough understanding of the unique context must be sought by aligning tasks with learning objectives. Real-world relevance should be prioritized, and assignments can be adapted to diverse learning styles. Authentic assessment methods reflecting real-world challenges should be employed, and staying updated on current trends within the context must be emphasized [17]. Collaboration with industry professionals should be encouraged for insights and credibility. Clear instructions should be provided, allowing for choice and flexibility while promoting critical thinking. Varied formats such as case studies and simulations should be utilized for engagement. Timely feedback is given, and assessment results should be reflected upon for continuous improvement. Reflective components in assignments should be encouraged to deepen understanding. Finally, assignments must be iteratively and continuously refined based on ongoing assessment, feedback, and changes in the context, ensuring ongoing relevance and effectiveness.

O. Presentations

Creating assignments based on presentations can be an effective way to assess students' communication skills, research abilities, and understanding of a topic. Assignments based on presentations involve the definition of clear objectives, the selection of relevant topics aligned with course content, and the provision of specific guidelines for format and time limits. Thorough research with proper citations, the use of visual elements, and consideration of various presentation styles are encouraged. Q&A sessions are included to promote critical thinking, and peer evaluation components are incorporated for collaborative learning. Real-world applications are related to presentation topics, and constructive feedback is provided on both content and communication skills. Technical proficiency is addressed, and reflection components are included for insights gained and strategies for improvement. Audience engagement strategies are encouraged, and a balance between content and delivery is sought for a comprehensive evaluation of students' presentation skills and knowledge. The authors in [22] suggested that oral assessment methods, such as the viva, are particularly valuable as it is straightforward and highly effective methods for obtaining direct and verifiable evidence of learners' understanding and competence.

P. Use in-class examples

Another way to combat this threat to validity is to design writing assignments by incorporating very topic-specific examples, original case studies, or handout material discussed in the class. This will make the assignment more authentic and realistic, fostering more interest in the topic. Also, it will be hard to feed such prompts to the AI tool and get the desired result. Even if they do so this requires them to go through the process and make sure that writing is in line with the requirements. This engagement with material will enhance their ability to learn resulting in long-term retention of the knowledge. This ensures that the student archives the learning outcome thus reducing the threat to the validity of the assessment.

Q. Use of remote-lab concepts

Leveraging remote labs is an innovative way [23] to assess students, going beyond the limitations of ChatGPT. Initially, learning objectives must be defined, and appropriate lab activities should be selected for hands-on application of theoretical knowledge. Remote lab platforms can be utilized, with clear instructions provided for accessing and performing tasks. Real-world scenarios should be incorporated into lab activities to assess students' understanding comprehensively. Critical thinking can be assessed through tasks that prompt the analysis of results and connections between theory and practice. Collaboration should be encouraged, and remote monitoring tools should be implemented to observe students' interactions. Data analysis components should be integrated, emphasizing the practical application of knowledge. Reflection is facilitated, and timely, constructive feedback is provided on students' performance. Lab reports should document procedures, results, and conclusions for a comprehensive assessment. Lab skills, including technique and precision, must be evaluated, to ensure a holistic understanding. Continuous improvement is emphasized through iterative assessments based on feedback and outcomes.

R. Higher penalty for false information

This is well well-known fact that ChatGPT has the tendency to generate false information because they are trained to predict word sequences that closely align with your input and lack the ability to engage in reasoning or address any factual inconsistencies present in their responses resulting in AI hallucinations. This limitation of the tool can be used to discourage the improper use of ChatGPT. The student should be informed ahead of time that they will have to pay a high price in the form of a deduction of the point, compared to the past, if they present any false information, made-up quotes, incorrect citations, or not factual. This will make them pay more attention to the information generated by AI tools and force them to verify the quotes or citations and ensure that they are real before putting them in the assignment. Students can even be asked to provide the link for citation allowing the teacher to verify that this is true information and not hallucinated one.

V. Challenges and Limitations

Efforts from faculty to integrate preventive measures into assessments to mitigate threats to validity demand significant dedication and investment of time and resources as is the case with any new instruction design technology [25]. Designing such assessments requires faculty members to engage in professional development activities to enhance their technological literacy and pedagogical skills. This adaptation process can be stressful, particularly for those who are not familiar with AI or who may feel overwhelmed by the rapid pace of technological change in education. Effectively preventing AI cheating may require technical knowledge and skills that instructors may not possess, such as implementing anti-cheating algorithms or encryption methods. Furthermore, AI technology is constantly evolving, with new techniques and tools emerging regularly. Keeping up with these advancements and updating assessments accordingly can be challenging for instructors. Also, in the absence of clear ethical guidelines related to the use of AI in academia, faculty may face ethical dilemmas when implementing measures to counter AI cheating, such as invasion of student privacy or unintentional discrimination against certain groups. Balancing the creation of secure assessments that withstand AI manipulation while ensuring accessibility and fairness for all students poses a delicate challenge. In short, given limited institutional support, it is challenging to expect teachers to independently adapt their working approach to effectively navigate the risks posed by new AI technology.

VI. Conclusion

In conclusion, this paper underscores the imperative for a comprehensive reassessment of academic assessment methodologies considering the challenges and limitations associated with integrating ChatGPT. Recognizing the expanding influence of artificial intelligence in educational settings, particularly during assessments, calls for a nuanced understanding of the dynamic relationship between technology and evaluation practices. The identified constraints in ChatGPT's ability to provide comprehensive and contextually accurate responses necessitate a proactive redesign of assessments to mitigate students' potential reliance on the model. Instructors play a pivotal role in this process, and their ability to effectively evaluate student work that might involve ChatGPT is enhanced by adopting alternative assessment approaches such as context-based questions, interviewing, presentations, etc. This shift ensures a fair and robust evaluation process that aligns with academic integrity standards. The findings presented in this paper contribute to the ongoing discourse on the responsible integration of AI in education, emphasizing the importance of proactive measures to preserve the authenticity and efficacy of assessment practices.

References

[1] NCTM. (n.d.). NCTM Publications. https://pubs.nctm.org/

[2] Taylor, Linda J. C., and Jeri A. Nicholas. “Graphing Calculators Aren’t Just for High School Students.” Mathematics Teaching in the Middle School 1, no. 3 (1994): 190–96.

[3] A. D. Thompson and S. L. Sproule, “Deciding when to use calculators,” *Math. Teach. Middle Sch.*, vol. 6, no. 2, pp. 126–129, 2000.

[4] Gilliland, Kay. "Families Ask: Why Not Just Use a Formula?." Mathematics Teaching in the Middle School 7, no. 9 (2002): 510-511.

[5] Introducing ChatGPT. (n.d.). https://openai.com/blog/chatgpt

[6] Mok, A. (2023, January 19). CEO of ChatGPT maker responds to schools’ plagiarism concerns: “We adapted to calculators and changed what we tested in math class.” Business Insider. https://www.businessinsider.com/openai-chatgpt-ceo-sam-altman-responds-school-plagiarism-concerns-bans-2023-1

[7] A.-T. Umar and A. Majeed, “The Impact of Assessment for Learning on Students’ Achievement in English for Specific Purposes: A Case Study of Pre-Medical Students at Khartoum University: Sudan,” *English Lang. Teach.*, vol. 11, no. 2, pp. 15–25, 2018.

[8] N. Dehouche, “Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3),” Ethics Sci. Environ. Polit., vol. 21, pp. 17–23, 2021. https://www.int-res.com/articles/esep2021/21/e021p017.pdf

[9] D. Mhlanga, “Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning,” *Educ. Responsible Ethical Use ChatGPT Towar. Lifelong Learn. (February 11, 2023)*, 2023. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354422

[10] E. Kasneci *et al.*, “ChatGPT for good? On opportunities and challenges of large language models for education,” *Learn. Individ. Differ.*, vol. 103, p. 102274, 2023.

[11] S. Akhtar, M. Nadeem, M. Rashdan, B. Hussain, E. A. Ansari, and M. H. Aslam, “Online Mode of Teaching and Learning Process in Engineering Discipline: Teacher Perspective on Challenges Faced and Recommendations,” *Educ. Sci.*, vol. 13, no. 2, 2023, doi: 10.3390/educsci13020200.

[12] G. Wiggins, *Educative Assessment. Designing Assessments To Inform and Improve Student Performance.* ERIC, 1998.

[13] F. McMartin, A. McKenna, and K. Youssefi, “Scenario assignments as assessment tools for undergraduate engineering education,” *IEEE Trans. Educ.*, vol. 43, no. 2, pp. 111–119, 2000.

[14] I. S. Chaudhry, S. A. M. Sarwary, G. A. El Refae, and H. Chabchoub, “Time to Revisit Existing Student’s Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT—A Case Study,” *Cogent Educ.*, vol. 10, no. 1, p. 2210461, 2023.

[15] OpenAI. (n.d.). https://openai.com/

[16] J. E. Raymond, C. S. E. Homer, R. Smith, and J. E. Gray, “Learning through authentic assessment: An evaluation of a new development in the undergraduate midwifery curriculum,” *Nurse Educ. Pract.*, vol. 13, no. 5, pp. 471–476, 2013.

[17] K. Ashford-Rowe, J. Herrington, and C. Brown, “Establishing the critical elements that determine authentic assessment,” *Assess. Eval. High. Educ.*, vol. 39, no. 2, pp. 205–222, 2014.

[18] M. Beisler and A. Medaille, “How do students get help with research assignments? Using drawings to understand students’ help seeking behavior,” *J. Acad. Librariansh.*, vol. 42, no. 4, pp. 390–400, 2016.

[19] Krathwohl, D.R., 2002. A revision of Bloom's taxonomy: An overview. Theory into practice, 41(4), pp.212-218.

[20] Litonjua, E., 2021. What is multimodal learning? eLearning Industry. https://elearningindustry.com/what-is-multimodal-learning

[21] Muhammad Nadeem, M. Oroszlanyova, and Wael Farag, “Effect of digital game-based learning on student engagement and motivation,” Computers, vol. 12, no. 9, p. 177, 2023. https://doi.org/10.3390/computers12090177

[22] D. R. E. Cotton, P. A. Cotton, and J. R. Shipway, “Chatting and cheating: Ensuring academic integrity in the era of ChatGPT,” *Innov. Educ. Teach. Int.*, pp. 1–12, 2023.

[23] Pearce, Jacob, and Neville Chiavaroli. “Rethinking assessment in response to generative artificial intelligence.” *Medical education* vol. 57,10 (2023): 889-891. doi:10.1111/medu.15092

[24] Wael Farag, “An Innovative Remote-Lab Framework for Educational Experimentation”, International Journal of Online Engineering, Vol. 13(2), pp. 68-86, Feb 2017. https://online-journals.org/index.php/i-joe/article/view/6609.

[25] Nadeem M, Lal M, Cen J, Sharsheer M. AR4FSM: Mobile Augmented Reality Application in Engineering Education for Finite-State Machine Understanding. Education Sciences. 2022; 12(8):555. https://doi.org/10.3390/educsci12080555.