Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 8, No. 3, June, 2024

Framing Assessment Questions in the Age of Artificial Intelligence: Evidence from ChatGPT 3.5

Mohammad Owais Farooqui [1*], Mohd Imran Siddiquei [2], Shashank Kathpal [3]

[1] *Department of Public Law, College of Law, University of Sharjah, Sharjah, United Arab Emirates.*

[2] *Department of Business Administration, University of the People, Pasadena, California, United States.*

[3] *Dr. Vishwanath Karad MIT World Peace University, Pune, India.*

Abstract

With the rise of artificial intelligence (AI), higher education faces a significant challenge in learning assessment. The emergence of tools like ChatGPT raises concerns regarding the potential for cheating and the reliability of assessment outcomes. This paper aims to address these concerns by proposing a methodology for framing questions that effectively measures learning outcomes while reducing the risk of AI-enabled cheating. To achieve this objective, we employ a methodological approach that involves getting responses from ChatGPT 3.5 to various question prompts across different domains. These responses are then evaluated by faculty members specializing in management education. Through this process, we aim to identify question-framing strategies that effectively assess learning outcomes while minimizing susceptibility to AI Cheating. Our analysis reveals several key findings. Certain question Types (Decision Making, Recent Events, and Experiential Learning) demonstrate greater resilience against AI-generated responses, indicating their potential effectiveness in assessing student learning. This study offers original insights into the challenges and opportunities associated with learning assessment in the context of AI integration. The paper tries to provide valuable guidance for Policymakers, educators & students seeking to enhance the integrity and reliability of their assessment practices.

Keywords:

Human-Computer Interface;
Improving Classroom Teaching;
Teaching/Learning Strategies.

Article History:

Received:	09	February	2024
Revised:	29	April	2024
Accepted:	07	May	2024
Published:	01	June	2024

1- Introduction

ChatGPT has received a lot of press coverage recently for its capacity to answer questions [1], offer advice on practically any issue in fluent, well-written English, create computer code, perform mathematical calculations, and do a variety of other things. ChatGPT is formulated using the OpenAI language model and undergoes training with an extensive collection of human conversations. This equips it to execute intricate tasks and generate responses that closely resemble human interaction [2]. The chatbot, which was introduced in November 2022 [3], has already been tested using several questions from tests, including those from law, medical, and business schools [4]. Some of these tests were passed comfortably by ChatGPT [5].

The implementation of ChatGPT in education has raised discussions regarding its possible educational impacts. While promoters of ChatGPT applaud its capacity to support education, such as by providing flexible and customized environments [6], many research scholars have concerns about ChatGPT's ethical concerns [7] and its possible adverse impact on assessment practices [8]. Universities' reactions to ChatGPT and other comparable AI technologies have been diverse [9], falling into three categories: preventive, prohibiting, and welcoming [10]. To avoid the use of AI

* **CONTACT**: mfarooqui@sharjah.ac.ae

DOI: http://dx.doi.org/10.28991/ESJ-2024-08-03-09

technologies, several colleges and universities may revert to traditional in-person exams using pen and paper. However, examinations and tests have never been the perfect means of assessment. They do not show an individual's capacity to work well in groups or to communicate and express material verbally, and they disfavor people who suffer from crippling test anxiety. Furthermore, prohibiting the use of ChatGPT could only be effective if all of the course assessments were for in-person work. To ensure that no learners could utilize ChatGPT, the proportion of marks for old-school tests and examinations would have to be increased, which would be a step backward. There are many AI tools available on the market for detecting ChatGPT-written text [11]. These AI detection technologies are not perfect, and they can be circumvented [12].

The problem with this approach is that students will spend more time trying to fool the system rather than understanding and learning the material. All these problems are even troublesome for universities offering pure online programs where in-person examinations are not possible. Cheating in online tests is more widespread than in traditional face-to-face exams [13]. According to Arnold [14], educators believe that academic dishonesty is on the rise and that online evaluation is especially favorable to cheating. Assessment methodologies and institutional regulations in schools and universities should be updated immediately [3]. Framing questions for the assessment of learning in the times of ChatGPT is really challenging for teachers [15]. The study aims to fill the research gap in creating assessment methodologies as identified by the following studies (Table 1):

Table 1. Research Gap in Assessment Methodologies

	Research Gap
Moqbel et al. (2023) [16]	Alternative assessments are crucial in the AI age for foreign language learning, as traditional methods may not capture learners' true abilities influenced by AI chatbots like ChatGPT.
Ifelebuegu (2023) [17]	Assessments need to focus on higher-order cognitive skills to maintain authenticity and prepare students for the evolving challenges of the 21st century.
Izzo et al. (2023) [18]	Assessments in the age of AI face challenges like design complexity, limited insights, lack of adaptability, and focus on outdated skills.
Goralski & Tan (2023) [19]	Assessments in the age of AI should focus on human capabilities like personal epistemology and evaluative judgment, enhancing graduates' skills to navigate an AI-enabled world effectively.

It is clear from various studies identified in the research gap that assessment in the age of AI is a significant challenge; therefore, this paper will try to understand different types of questions with evidence from ChatGPT 3.5 that has the least chance of cheating. These responses will be evaluated by faculty members in the management domain, and then key findings and implications will be drawn.

2- Objectives & Methodology

The study aims to understand which types of questions can be framed that can't be answered by ChatGPT 3.5 in order to reduce the chances of cheating among students and whether there is a difference in the ability of ChatGPT 3.5 to answer the questions if the prompts are different. The study uses ChatGPT 3.5, as it is free and popular among students. Due to the exploratory nature of the study, a systematic flow chart has been drawn in order to achieve the objectives of the paper (see Figure 1).

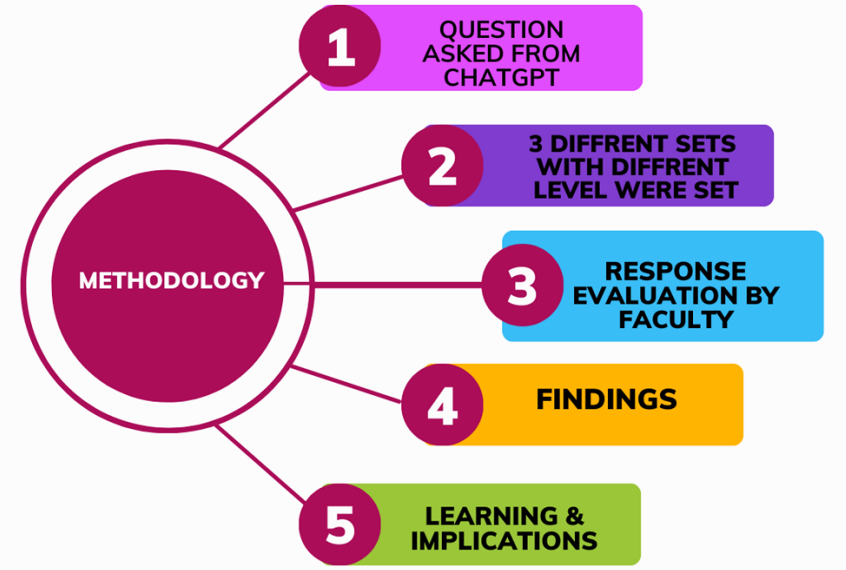

Figure 1. Flowchart of the Methodology

In order to frame assessment questions, the first prompt was run on ChatGPT and focused on what types of questions could not be answered by ChatGPT.

Prompt: Which type of question can't be answered?

Answer by ChatGPT[*]*:*

- *As an AI language model, there are some types of questions that I may not be able to answer or may provide inaccurate answers for. These include:*
- *Questions that require subjective opinions or personal experiences that I don't have as I am not a sentient being.*
- *Questions that involve information that is not available in my training data or knowledge cutoff.*
- *Questions that require physical perception or sensory input, as I do not have access to the physical world.*
- *Questions that are ambiguous, incomplete, or poorly formed.*

On the basis of the above information, different sets with different levels of prompts (Table 1) were designed from the perspective of the core areas of management:

Set 1 Questions were formed to check the ability of ChatGPT 3.5 to answer questions based on Subjective opinions or personal experiences (Table 2).

Set 2 Questions were formed to check the ability of ChatGPT 3.5 to answer questions based on the Latest Information (Table 2).

Set 3 Questions were formed to check the ability of ChatGPT 3.5 to answer questions based on Physical Perception or sensory input (Table 2).

Table 2. ChatGPT 3.5 Evaluation Questions

Table 1

Subjects	Set	Factors	Questions
Marketing	Set 1	Subjective opinions or personal experiences	Explain the Marketing perspectives from the Following article* in 100 words with Bullet points. (*Article Taken from Websites)
	Set 2	Latest Information	Share the latest Advertisement spending of the Top 3 Indian companies (According to Market Cap) in 150 words with Bullet points.
	Set 3	Physical Perception	Visit your nearby market and social media page and observe marketing campaigns (Reliance trends) in 200 words with Bullet points.
Human Resource	Set 1	Subjective opinions or personal experiences	Explain the HR perspectives from the Following article* in 100 words with Bullet points. (*Article Taken from Websites)
	Set 2	Latest Information	Share the latest Retrenchment practices of the Top 3 Indian companies (According to Market Cap) in 150 words with Bullet points.
	Set 3	Physical Perception	Visit the Ambition box and review the employee reviews of the Top 3 Indian companies (According to Market Cap) in 200 words with Bullet points.
Finance	Set 1	Subjective opinions or personal experiences	Explain the Financial perspectives from the Following article* in 100 words with Bullet points. (*Article Taken from Websites)
	Set 2	Latest Information	Share the latest Financial Ratios of the Top 3 Indian companies (According to Market Cap) in 150 words with Bullet points.
	Set 3	Physical Perception	Visit Moneycontrol.com and take a call (Buy/sell) of the Top 3 Indian companies (According to Market Cap) with technical analysis in 200 words with Bullet points.
Operations	Set 1	Subjective opinions or personal experiences	Explain the Operational perspectives from the Following article* in 100 words with Bullet points. (*Article Taken from Websites)
	Set 2	Latest Information	Share the latest Delivery practices of the top 3 Indian e-commerce companies (According to Market Share) in 150 words with Bullet points.
	Set 3	Physical Perception	Visit the Amazon warehouse and suggest Empirical solutions Based on PERT & CPM in 200 words with Bullet points.

To evaluate questions, we asked 13 faculty members teaching in MBA colleges that are AICTE/UGC affiliated. We asked them to evaluate responses on four criteria, i.e., content coverage [20], cognitive complexity [21], meaningfulness [22], and representativeness [23], to evaluate the authenticity of the answers (Baartman et al., 2007) [24].

[*] *https://chat.openai.com/*

3- Findings

3-1- Qualifying Score

The assessment of ChatGPT's performance across different question sets offers a clear explanation of how AI can support learning, but also where it struggles to answer the questions. Set 1's decent average score of 6.92 indicates that ChatGPT can offer thoughtful responses based on personal experiences and opinions, which is impressive. It seems to imitate human-like interaction quite well, which is promising for AI's role in higher education but also a greater challenge due to the probability of higher cheating chances. However, the scores take a hit for sets 2 and 3, averaging 5.37 and 4.07, respectively. These lower scores indicate that ChatGPT isn't quite able to answer and fails to achieve passing standards for students set by regulatory agencies like the UGC (University Grant Commission). It seems to struggle, especially in providing accurate and up-to-date information or describing things based on physical experiences. This discrepancy clearly explains that while AI tools like ChatGPT can be great in some areas, they're not foolproof. It works great in creativity and casual conversation but struggles with tasks requiring real-time data interpretation or sensory understanding.

Figure 2, with its box plot, reflects a clear picture of these differences, visually showing how ChatGPT performs across the various question sets. These findings emphasize the importance of considering both the strengths and limitations of AI in educational assessments. We need a balanced approach that combines AI's capabilities with human expertise to ensure assessments effectively measure learning outcomes with integrity and reliability.

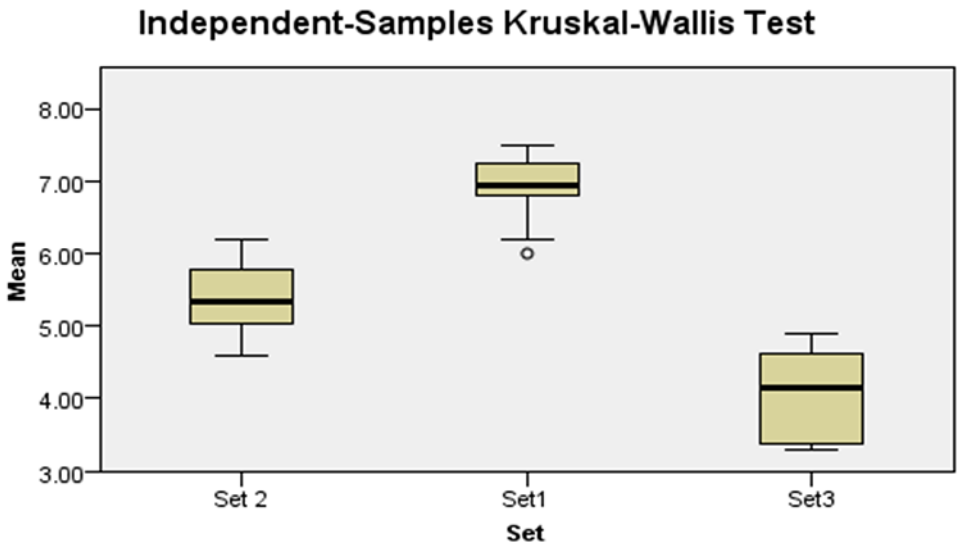

Figure 2. Box Plot Comparison of ChatGPT Performance Across Question Sets

3-2- Kruskal-Wallis Test

The Kruskal-Wallis H test, often referred to as the "one-way ANOVA on ranks," serves as a robust statistical tool for assessing differences among multiple groups concerning any independent variable. In our study, we utilized this test to understand the variance between scores across different question sets, providing valuable insights into the effectiveness of each set-in measuring learning outcomes. Figure 2 reflects the output of the Kruskal-Wallis Test, that set3-set2 has a significance value of 0.012, set3-set1 has a significance value of 0.000, and set1-set2 has a significance value of 0.002. It reflects that the distribution value of each set (set1, set2, set3) at the $p<0.05$ level of questions is significantly different from the other set. Specifically, it indicates the difference between sets 3 and 2 is significant with a p-value of 0.012, with a test statistic of 14.116 indicating that the scores of set 3 scores are substantially higher than the scores of set 2. Further, it indicates the difference between sets 3 and 1 is significant with a p-value of 0.000, with a test statistic of 30.241 indicating that the scores of set 3 scores are substantially higher than the scores of set 1. It indicates the difference between sets 1 and 2 is significant with a p-value of 0.002, with a test statistic of -16.125 indicating that the scores of set 1 scores are substantially lower than the scores of set 2 (Figure 3). These findings carry significant implications for educational assessment practices. The significant difference between the scores of set 3 and the other sets suggests that this set may be particularly effective in measuring learning outcomes that can be helpful in creating strong assessment strategies.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Set3-Set 2	14.116	4.907	2.877	.004	.012
Set3-Set1	30.241	4.907	6.163	.000	.000
Set 2-Set1	-16.125	4.741	-3.401	.001	.002

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure 3. Pairwise Comparisons

4- Learnings

4-1- Questions based on Experiential Learning

Experiential learning is the process of acquiring knowledge and abilities via personal experience rather than just reading or lecturing [25]. This method has grown in popularity in recent years because it is more successful at helping students remember information and build practical skills that can be utilized in real-world situations. One of the most significant benefits of experiential learning is that it helps students learn from their own experiences rather than depending exclusively on theoretical information [26]. This is especially vital in fields such as business and marketing, where success depends on the capacity to comprehend and respond to complicated and ever-changing consumer needs, competition, target markets, government policies, etc.

While AI models have made significant advances in analyzing and predicting data in recent years, they remain constrained in their capacity to duplicate the insights that come from first-hand observation. This is due to the fact that experience learning entails a subjectivity and complexity that is difficult to reproduce using automated techniques. In marketing research, for example, interviews and surveys are frequently used to gather insights into customer behavior and preferences. The value of these tools, however, rests not only in the data that they provide but also in the knowledge that may be garnered from the interactions that occur between surveyors and respondents.

Similarly, developing a real-money portfolio is a great experiential learning activity since it allows students to obtain actual experience in investing and financial management. This includes not only the process of purchasing and selling stocks but also the behavioral and psychological components of investing, such as tolerance for risk and decision-making under high pressure. Overall, the importance of experiential learning stems from its potential to provide individuals with a more in-depth grasp of a subject through first-hand knowledge and application in real life. While AI models have many advantages in terms of analyzing information and forecasting, they are unable to precisely replicate the insights and lessons learned through direct experience and exposure to the financial markets.

As a result, questions based on experiential learning may be a useful tool in assisting individuals in developing the practical skills and knowledge required for success in a variety of professions, including marketing and business. These questions, by allowing students to learn from their own experiences, can assist in bridging the gap between theory and practice, ultimately leading to improved outcomes in real-world contexts.

4-2- Questions based on Recent Events

In today's fast-paced world, keeping up with the latest news and developments is more vital than ever in business education. This is especially true in industries such as business and marketing, where success depends on the capacity to adapt rapidly and effectively to business environment changes. AI models have developed as a strong tool for processing and analyzing massive volumes of data, and they have grown in popularity in recent years for answering queries about current events. While AI models have many advantages in terms of analyzing information and prediction, they are not perfect, and there have been numerous instances of factual and data errors [27].

One of the primary reasons for this is that AI models are only as good as the information upon which they are built [3]. The AI model generates inaccurate conclusions if the data is inadequate, biased, or wrong. Furthermore, AI models can make mistakes due to unanticipated changes in the data or errors in the algorithms used to process it. It is crucial to

highlight, however, that these questions must be updated frequently, as AI models are very swift to update themselves with what is going on in the real world. This implies that evaluation questions based on current events must be revised frequently to remain pertinent and correct.

4-3- Questions based on Decision-Making

Making decisions is an important ability in both personal and professional situations, but it may be difficult to teach and analyze. While AI models' ability to analyze and interpret data related to decision-making is improving, making decisions on moral judgment is still a challenge for Artificial Intelligence [28]. There continues to be value in asking queries requiring reasoning and subjective assessment. While chatbots such as ChatGPT can provide general answers to decision-making queries, they may not always be as particular, or objective as needed. This is due to the fact that decision-making is a complicated process that includes a range of aspects such as individual principles, opinions, perceptions, and experiences. It is challenging for AI models to consider these subjective elements when making judgments or answering decision-making questions. As a result, decision-making problems are an excellent alternative for beating AI models because they demand users to use critical thinking and subjective analysis. These sorts of questions can assist learners with a better understanding of the concepts and principles that underlie decision-making as well as in developing their own decision-making skills and talents. Examples of decision-making questions include circumstances in which individuals must make a difficult ethical decision, as well as problems in which persons must analyze evidence and offer a suggestion based on that analysis.

Educators and trainers may help students enhance their critical thinking abilities and become better decision-makers by asking decision-making-based questions.

5- Implications

5-1- Policymakers

Artificial intelligence (AI) is rapidly revolutionizing several industries, including education. The rise of artificial intelligence has prompted calls for an overhaul of the education system, including universities and accreditation organizations. It is becoming increasingly clear that allowing students to pass without putting in the necessary effort to develop excellent knowledge and skills would result in poor learning outcomes and meaningless degrees. The resultant availability of labor in the market is useless if it lacks quality learning, as it will have implications for future productivity in the industry.

The traditional approach to education has become obsolete in the face of artificial intelligence. AI is a constantly upgrading area, and today's courses may become obsolete in a matter of months. As a result, universities must be proactive in upgrading their courses [29]. to ensure that students have the most up-to-date information and abilities. Education policymakers must prioritize the development of AI-proof assessments [30] and courses that cannot be easily replicated by AI-based machines, such as creativity, critical thinking, and problem-solving. Universities must re-evaluate and refresh their programs and incorporate courses with updated changes that focus on building these skills in students. Universities can ensure that their graduates stay competitive in the labor market by providing them with capabilities that artificial intelligence cannot simply imitate.

5-2- Teaching Faculties

The introduction of artificial intelligence (AI) into education has necessitated considerable advancements in teaching methodology, question formulation, and assessment systems. The role of teaching faculty in preparing students for the challenges and possibilities of the AI era has grown in significance. Faculty must assist students in gaining the required skills and information to thrive in a fast-changing job market by focusing their attention on experiential learning, decision-making, and updated course material. It is critical that faculties must accept these changes and attempt to adapt to the new reality of artificial intelligence in education, as tools like ChatGPT can be great assistants in effective teaching [30].

5-3- Students

AI in education provides several potential benefits for students. It can assist them in better managing their time, generating initial ideas [31, 32], simplifying complex subjects, and providing personalized learning experiences [33]. However, it is critical to note that ethical concerns must be followed consistently all the time to ensure that AI is utilized to benefit students in learning rather than to cheat or plagiarize. AI should be utilized as additional help rather than replacing human teachers.

6- Conclusion

The arrival of artificial intelligence (AI) has posed a significant challenge to higher education. There is a growing concern regarding the usage of AI tools, such as ChatGPT, in answering queries across diverse academic domains, potentially giving rise to instances of academic misconduct. To tackle this challenge, this study adopts a methodology that leverages ChatGPT prompts for generating assessment questions and subsequently analyzes the outcomes. The research aims to explore effective strategies for crafting assessment questions that accurately measure student learning while upholding the integrity of the evaluation process and maintaining learning standards.

Experiential learning, learning based on current events, and decision-making scenarios emerge as pivotal components of student learning assessment. These aspects can significantly help in the acquisition of practical skills, up-to-date knowledge, and enhancement of critical thinking and decision-making abilities among students by minimizing academic misconduct and resulting in better learning outcomes. Given the urgency of the need for educational reform, policymakers should advocate for meaningful learning and assessment strategies that would provide students with competencies and discourage them from using AI for cheating purposes. In addition to policymakers, faculty members have a significant role in this, considering that they prepare students for the future. All instruction methods, question formulation, and framework for evaluation should be adapted to tackle the challenges of the AI era. At the same time, they should encourage the ethical use of AI tools among students to make AI a catalyst for the enhancement of students learning and development rather than a tool for cheating and dishonesty. It is also the student's responsibility to follow ethical guidelines while using AI tools. They must understand that the objective of higher education is a better learning outcome rather than only getting passing marks and achieving the degree. In summary, the study broadens the domain of learning assessment by providing empirical evidence, learning, and implications to discourage and minimize the use of AI among students for academic misconduct. This investigation further helps to expand the discussion about the need for change and innovative approaches in education and the consequences of the increased use of AI tools in higher education.

6-1- Limitation & Research Implications

The study has employed ChatGPT 3.5 for framing assessment questions; however, the better version, at 20 dollars per subscription, is available on the internet, which can potentially be used in future research studies. The prompts were designed at three different levels, while more complex levels can be further understood. The findings of the study can't be generalized, as different prompts have different types of answers. Answers were analyzed by a small sample of management faculty members; future studies can utilize a higher number of faculty members with different domain knowledge.

7- Declarations

7-1- Author Contributions

Conceptualization, M.O.F. and M.I.S.; methodology, M.I.S.; software, M.O.F.; validation, M.O.F., M.I.S., and S.K.; formal analysis, M.O.F.; investigation, M.I.S.; resources, S.K.; data curation, M.O.F.; writing—original draft preparation, M.O.F.; writing—review and editing, M.I.S.; visualization, S.K.; supervision, M.O.F.; project administration, M.I.S.; funding acquisition, M.O.F. All authors have read and agreed to the published version of the manuscript.

7-2- Data Availability Statement

Data sharing is not applicable to this article.

7-3- Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

7-4- Acknowledgements

We extend our sincere gratitude to the University of Sharjah for providing the necessary resources and support for this research.

7-5- Institutional Review Board Statement

Not applicable.

7-6- Informed Consent Statement

Not applicable.

7-7- Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the authors used [ChatGPT 3.5] to [Understand the ability to answer the questions].

7-8- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

8- References

[1] Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). ChatGPT is not all you need. A State-of-the-Art Review of large Generative AI models. arXiv preprint, arXiv:2301.04655. doi:10.48550/arXiv.2301.04655.

[2] Susnjak, T. (2022). ChatGPT: The end of online exam integrity? arXiv preprint, arXiv:2212.09292. doi:10.48550/arXiv.2212.09292.

[3] Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. Education Sciences, 13(4), 410. doi:10.3390/educsci13040410.

[4] Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. Education and Information Technologies, 27(6), 7893–7925. doi:10.1007/s10639-022-10925-9.

[5] Chen, L. E., Cheng, S. Y., & Heh, J.-S. (2021). Chatbot: A Question Answering System for Student. 2021 International Conference on Advanced Learning Technologies (ICALT), Tartu, Estonia. doi:10.1109/icalt52272.2021.00110.

[6] Qadir, J. (2023). Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education. 2023 IEEE Global Engineering Education Conference (EDUCON), Kuwait, Kuwait. doi:10.1109/educon54358.2023.10125121.

[7] Mhlanga, D. (2023). Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning. SSRN Electronic Journal, 1-199. doi:10.2139/ssrn.4354422.

[8] Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching, 6(1), 342–363. doi:10.37074/jalt.2023.6.1.9.

[9] Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. Journalism & Mass Communication Educator, 78(1), 84–93. doi:10.1177/10776958221149577.

[10] Sims, A. (2023, March 9). ChatGPT and the future of University Assessment. Times Higher Education, Washington, United States. Available online: https://www.timeshighereducation.com/campus/chatgpt-and-future-university-assessment (accessed on June 2024).

[11] Loh, E. (2023). ChatGPT and generative AI chatbots: challenges and opportunities for science, medicine and medical leaders. BMJ Leader, 8(1), 51–54. doi:10.1136/leader-2023-000797.

[12] Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection. http://arxiv.org/abs/2301.07597.

[13] Noorbehbahani, F., Mohammadi, A., & Aminazadeh, M. (2022). A systematic review of research on cheating in online exams from 2010 to 2021. Education and Information Technologies, 27(6), 8413–8460. doi:10.1007/s10639-022-10927-7.

[14] Arnold, I. J. M. (2016). Cheating at online formative tests: Does it pay off? Internet and Higher Education, 29, 98–106. doi:10.1016/j.iheduc.2016.02.001.

[15] Joyner, D. A. (2023). ChatGPT in Education: Partner or Pariah? XRDS: Crossroads, The ACM Magazine for Students, 29(3), 48–51. doi:10.1145/3589651.

[16] Moqbel, M. S. S., & Al-Kadi, A. M. T. (2023). Foreign language learning assessment in the age of ChatGPT: A theoretical account. Journal of English Studies in Arabia Felix, 2(1), 71-84. doi:10.56540/jesaf.v2i1.62.

[17] Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. Journal of Applied Learning and Teaching, 6(2), 1-8.

[18] Izzo, Z., Liu, R., & Zou, J. (2023, July). Data-driven subgroup identification for linear regression. In International Conference on Machine Learning. Proceedings of the 40th International Conference on Machine Learning, PMLR 202:14531-14552, 2023.

[19] Goralski, M. A., & Tan, T. K. (2020). Artificial intelligence and sustainable development. The International Journal of Management Education, 18(1), 100330. doi:10.1016/j.ijme.2019.100330.

[20] Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, Performance-Based Assessment: Expectations and Validation Criteria. Educational Researcher, 20(8), 15–21. doi:10.3102/0013189X020008015.

[21] Natriello, G. (2013). The Impact of Evaluation Processes on Students. School and Classroom Organization, 227–246, Routledge, Milton Park, United Kingdom. doi:10.4324/9780203056950-9.

[22] Birenbaum, M. (2007). Evaluating the Assessment: Sources of Evidence for Quality Assurance. Studies in Educational Evaluation, 33(1), 29–49. doi:10.1016/j.stueduc.2007.01.004.

[23] Maclellan, E. (2004). How convincing is alternative assessment for use in higher education? Assessment and Evaluation in Higher Education, 29(3), 311–321. doi:10.1080/0260293042000188267.

[24] Baartman, L. K. J., Prins, F. J., Kirschner, P. A., & van der Vleuten, C. P. M. (2007). Determining the Quality of Competence Assessment Programs: a Self-Evaluation Procedure. Studies in Educational Evaluation, 33(3–4), 258–281. doi:10.1016/j.stueduc.2007.07.004.

[25] Roberts, T.G. (2011). What Is Experiential Learning. The Agricultural education magazine, 83, 4.

[26] Wurdinger, S., & Allison, P. (2017). Faculty perceptions and use of experiential learning in higher education. Journal of e-learning and Knowledge Society, 13(1), 1-12.

[27] Osoba, O., & Welser, W. (2017). An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence, RAND Corporation, Santa Monica, United States. doi:10.7249/rr1744.

[28] Banerjee, D. N., & Chanda, S. S. (2020). AI failures: A review of underlying issues. arXiv preprint, arXiv:2008.04073. doi:10.48550/arXiv.2008.04073.

[29] Perkins, M., & Roe, J. (2023). Decoding Academic Integrity Policies: A Corpus Linguistics Investigation of AI and Other Technological Threats. Higher Education Policy. doi:10.1057/s41307-023-00323-2.

[30] Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and Opportunities of Generative AI for Higher Education as Explained by ChatGPT. Education Sciences, 13(9), 856. doi:10.3390/educsci13090856.

[31] Alafnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). ChatGPT as an Educational Tool: Opportunities, Challenges, and Recommendations for Communication, Business Writing, and Composition Courses. Journal of Artificial Intelligence and Technology, 3(2), 60–68. doi:10.37965/jait.2023.0184.

[32] Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. doi:10.1016/j.lindif.2023.102274.

[33] Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 61(3), 460–474. doi:10.1080/14703297.2023.2195846.