International Journal of Educational Studies and Policy (IJESP)

Volume: 5, Issue: 2, November 2024

Impact of Artificial Intelligence on Assessment and Evaluation Approaches in Education

Hande Tanberkan[1] , Mahmut Özer[2] , Selahattin Gelbal[3]

ABSTRACT

Artificial Intelligence (AI) technologies are being applied commonly in all aspects of life. Education is one of the leading areas in this respect. AI applications offer significant opportunities for students, educators and education administrators. Students can benefit from these technologies for individualized education and addressing their deficiencies. A similar situation applies to educators. However, students are the most vulnerable group to the current and long-term risks posed by these technologies. While students fulfill a significant part of their responsibilities through the opportunities provided by AI technologies, they face two options: succeeding through ethical violations or addressing their deficiencies ethically. Students lacking AI ethical literacy often choose the first option, masking their failures and getting involved in ethical violations that will bring heavy burdens in the long run. This study discusses the benefits offered by AI technologies in education and the problems they cause in the context of measurement and evaluation. AI will have an important place in the measurement and evaluation as an auxiliary tool in producing texts, creating questions using the produced texts, scoring open-ended exams, solving problems and creating research reports in accordance with ethical rules. It is highlighted that developing AI technologies in education with a participatory approach involving all educational stakeholders and continuously monitoring potential risks during the implementation phase are crucial for establishing a responsible AI culture in education. Finally, considering the dramatic pace of developments in AI, the importance of dynamically updating the measures against ethical violations at the same pace is emphasized.

Keywords: Artificial intelligence, measurement, evaluation, ethics

DOI: https://doi.org/10.5281/zenodo.14016103

Article Info:

Received: 05.06.2024 **Accepted: 22.10.2024** **Article Type: Review Article**

Cite as: Tanberkan, H. Özer, M. & Gelbal, S. (2024). Impact of artificial intelligence on assessment and evaluation approaches in education. *International Journal of Educational Studies and Policy*, *5*(2), 139-152.

[1]Corresponding Author: Hande Tanberkan, Commission of National Education, Culture, Youth and Sports, TBMM, handetanberkan@gmail.com , ORCID: 0000-0001-7142-5397

[2]Mahmut Özer, Commission of National Education, Culture, Youth and Sports, TBMM, mahmutozer2002@yahoo.com , ORCID: 0000- 0001-8722-8670

[3] Selahattin Gelbal, Hacettepe University, Faculty of Education, sgelbal@gmail.com , ORCID: 0000-0001-5181-7262

Introduction

Artificial intelligence (AI) technologies are among the most significant technological disruptions in history, their capacity and effects have not been fully evaluated yet. Specifically, AI applications, which began to emerge in fields such as education, healthcare, and finance in the 1970s, have gained significant momentum over the past decade, elevating this technological disruption to a new phase compared to previous technological transformations (Acemoglu & Restrepo, 2018; Frank et al., 2019; İlikhan et al., 2024; Ozer, 2024; Perc et al., 2019; Septiandri et al., 2023). At this point, there are no areas untouched by AI technology, and new applications are being introduced in every field daily. Consequently, an AI ecosystem encompassing all areas of life is now being discussed (Ozer et al., 2024a; Stahl, 2023). In other words, there is a significant potential for jobs to fall under the dominance of AI (Ozer & Perc, 2024). Furthermore, the advancement of generative AI such as ChatGPT and Gemini has accelerated these developments (Hosseini et al., 2023; Lo, 2023).

The rapid proliferation of AI technologies is now drawing the attention of all segments of society, revealing significant risks alongside the advantages it provides (Suleyman, 2023). It is known that these technologies reproduce the biases present especially in the training data sets used during the learning process, thereby deepening inequalities (Ozer et al., 2024a). In this context, it has been shown that AI technologies in healthcare services deepen racial and ethnic disparities, further exacerbating the disadvantages of already socioeconomically disadvantaged groups (İlikhan et al., 2024; Obermeyer et al., 2019). Similar biases manifest themselves in various fields from education to security and law (Ozer et al., 2024a; 2024b). In short, awareness of AI is increases globally, discussions emerge on how to create an AI ecosystem that does not exacerbate social inequalities, respects social and ethical values, and particularly does not adversely affect employment (Acemoglu & Restrepo, 2018; Ozer & Perc, 2024; Ozer et al., 2024a; Varma et al., 2023).

The opportunities provided by AI increase the allure of AI utilization in also educational environments; however, discussions continue regarding the reliability of AI in education, whether these opportunities are offered within the framework of ethical principles, and whether they support the improvement of educational quality (Acemoglu, 2024). In other words, by utilizing educational data, AI offers opportunities to transform and enrich the learning and teaching processes on one hand, while on the other hand, it carries the risk of increasing existing inequalities among students and ethical misuse (Aquino, 2023; Silva-Jurado & Silva-Jurado, 2024; Ozer, 2024). Therefore, AI system needs to regulate the learning journey in a healthy manner while supporting student development at the same time. In this context, measurement and evaluation processes play a critical role. The development of AI systems signals the revision requirements of traditional assessment methods (Gardner et al., 2021). AI offers a process that encourages individual development, provides guidance, and supports fair assessment outcomes while also having the potential to reduce the workload of educators (Kamalov et al., 2023).

However, integrating AI into measurement and evaluation systems also brings along a series of risks (Cotton et al., 2024; Sok & Heng, 2023; Surahman & Wang, 2022; Verhoeven et al., 2023). AI's text generation capability leads to the risk of cheating and plagiarism in assessment processes. The text generation capabilities offered by generative AI tools like ChatGPT raise serious concerns about undetectable cheating and plagiarism (Kamalov et al., 2023; Ozer, 2024; Surahman & Wang, 2022). The inability of assessment mechanisms to be able to detect AI-generated content provides an unfair advantage to students who use AI for text generation,

adversely affecting the fundamental quality of fair evaluation in assessments. Another negative impact of this situation is the loss of opportunities for individuals to achieve the targeted learning outcomes (Lancaster, 2023). In this context, unethical use of AI for cheating and plagiarism distorts the true value of assessment results and disrupts academic development processes.

On the other hand, the capacity of generative tools like ChatGPT is limited by the dataset they are trained on (İlikhan et al., 2024; Ozer et al., 2024a). In other words, the outputs generated by ChatGPT are limited by the size, scope, and diversity of its training dataset. Since the datasets are derived from real-life scenarios where biases exist, there is a risk of obtaining biased results from AI. While the use of AI in text-related tasks such as translation, essay writing, and automatic content generation offers opportunities for saving time and managing the learning process, it also carries the risk of reinforcing and perpetuating these biases (Gardner et al., 2020; Rane et al., 2023). For example, when translating from a gender-sensitive language to a gender-insensitive language, biased results are observed. In Turkish, he/she is a nurse is translated as she is a nurse by referring to women, while he/she is a doctor is translated as he is a doctor by referring to men. In other words, AI-assisted translation perpetuates the societal biases, such as gender bias (Akgun & Greenhow, 2021; Johnson, 2020).

In addition, it is known that not all content produced by AI systems is accurate and often exhibits a behaviour called hallucination (Ji et al., 2023). When productive AI systems exhibit hallucinatory behaviour, they may produce context-irrelevant or non-existent content, although it may seem plausible. For example, it has been shown that most of the references provided by ChatGPT for use in scientific papers do not actually exist (Athaluri et al., 2023). More interestingly, it is known that once ChatGPT shows hallucinatory behaviour, it maintains this behaviour to ensure the consistency of the content it produces, thus leading to the snowball effect of hallucination (Zhang et al., 2023).

The impact of AI's capabilities on assessment and evaluation systems should be examined in terms of the opportunities it presents and the risks it entails. While maximizing the benefits of AI usage in assessment processes, risks should be minimized to a maximum as well. Enhancing the accuracy and reliability of assessment while using this technology to achieve maximum efficiency should be an essential responsibility for educators. Therefore, the purpose of this article is to identify the opportunities created by AI in measurement and evaluation processes, along with the ethical risks involved, and to propose solutions to mitigate these risks.

Method

This study employs a descriptive approach to evaluate the risks associated with the use of artificial intelligence in text generation and assessment processes in education and also to offer recommendations for managing these risks. The document analysis method is employed to conduct a detailed examination aimed at mitigating the risks associated with the use of artificial intelligence in text generation and assessment. As a qualitative method, document analysis involves conducting a comprehensive review of relevant documents including articles, books, and reports (Bowen, 2009). Accordingly, the literature on the use of artificial intelligence in text generation and assessment has been reviewed. The study illustrates application of artificial intelligence in text generation and assessment including a detailed examination of potential ethical risks. Additionally, a thorough analysis has been conducted to provide recommendations for addressing the risks associated with the use of artificial intelligence in text generation and assessment.

Text generation

Generative AI tools such as ChatGPT have the capability to generate coherent text tailored to assignments and exam questions using machine learning and natural language processing (Cotton et al., 2024; Salvagno et al., 2023). The biggest challenge encountered with AI in education is its predominant use in creating assignments and projects. By doing so, students put themselves into two problematic situations. Firstly, they engage in behavior where they present knowledge that is not their own as if it were theirs, potentially leading to significant long-term behavioral distortions. Secondly, they may mask their shortcomings by presenting deficiencies as competencies (Kasneci et al., 2023; Sok & Heng, 2023). In this case, the complementary support that artificial intelligence could provide to the measurement and evaluation processes could lead to a reverse effect and could enable students at all levels of education to progress successfully despite all their shortcomings. In other words, students may appear to succeed even though they are failing. This situation could deprive students of remedial training for their shortcomings, ultimately leading to their graduation from education without gaining the expected skills in human capital in the long term.

On the other hand, due to the widespread use of AI text generation among students, concerns have emerged that traditional evaluations must be revised (Khalil & Er, 2023). The high grades achieved by texts generated with ChatGPT validate these concerns (Stokel-Walker, 2022). As a strategy to address these concerns, some universities have developed policies prohibiting the use of ChatGPT (Sullivan et al., 2023). However, instead of banning it, universities should focus on integrating AI into their systems ethically and should provide regular trainings to their students on how to benefit from AI (Ozer, 2024; Sok & Heng, 2023; Yu, 2023).

In education, ethical concerns related to AI are not limited to K-12 levels but also encompass higher education institutions (Huallpa, 2023). The impact of AI technologies in higher education institutions is significantly higher for both the researchers and the students. In other words, AI's capabilities in generating texts extend beyond educational assignments, and offer substantial opportunities in scientific article production (Rane et al., 2023). Recently, there has been even discussion regarding whether ChatGPT should be considered a co-author in scientific publications (Stokel-Walker, 2023; Thorp, 2023). This is because ChatGPT's abilities in text generation, translation, and summarization enhance its potential utility in scientific writing (Verhoeven et al., 2023). Hence, AI is expected to support reseachers in organizing ideas, translating, drafting, providing feedback and proofreading in the process of producing scientific articles (Rane et al., 2024). This support makes the preparation of academic studies faster and easier and provides the advantage of increasing the quantity and quality of publications (Rane et al., 2023; Salvagno et al., 2023). However, in scientific articles, authors bear collective responsibility for the content (Stokel-Walker, 2023; Thorp, 2023). Computers are just tools and cannot assume responsibility for the content (Thorp, 2023). Therefore, it cannot be expected by AI to produce original, creative, and critical ideas like humans in scientific article production (Rane et al., 2023; Salvagno et al., 2023). AI-generated texts have not yet established significant credibility in scientific research due to risks such as containing biased content and hallucinations (Salvagno et al., 2023; Stokel-Walker, 2022; Verhoeven et al., 2023).

In summary, the main risks in creating text with AI are ethical concerns in general and the accuracy of the content produced in particular (Lo, 2023). In this context, using AI to generate entire texts rather than as a guide poses the greatest risk leading to fundamental ethical violations.

Therefore, AI's role in text writing should remain that of a helpful assistant to improve efficiency and performance (Verhoeven et al., 2023).

Automatic item generation

The preparation of measurement tools such as exams, assignments, and presentations to measure knowledge and skills increases the workload of educators due to the time involved in the item generation process. Additionally, questions created through traditional methods are constraint by the capabilities of the item writers. Therefore, developments in AI lead to the expectation of speed and increased quality in automatic item generation (AIG) (Bezirhan & von Davier, 2023; Cotton et al., 2024). As AI applications provide educators with the opportunity to easily create achievement tests for classroom assessments. In this context, it is seen that significant developments have been achieved in AI-supported creation of assessment tools by analyzing the educational content such as multiple-choice questions and open-ended questions. (Owan et al., 2023; Qi et al., 2020; Swiecki et al., 2022). In particular, generative AI tools such as ChatGPT and Gemini have the capacity to generate questions in line with the basic skills expected from students.

AIG with correct sentence structures has been partially achieved with existing AI algorithms; however, the content quality of these questions remains debated (Du et al., 2017, Swiecki et al., 2022). For instance, automatically generated items have been criticized for being difficult to understand and aligned with the objectives to be measured (Mulla& Gharpure 2023, Scialom et al., 2019). However, in the process of text generation with AI, studies are ongoing to produce questions with desired features by intervening in AI. For example, Sayın and Gierl (2024) standardise the item generation process by providing ChatGPT with templates containing limitations such as question types, word counts, and sentence structures during the process of item generation with AI. The quality of the generated items was evaluated by experts in the field and item statistics were also calculated. As a result, it was determined that AI increases efficiency in item development processes (Sayın& Gierl, 2024). On the other hand, Bezirhan and von Davier (2023) found that ChatGPT was effective in generating paragraph-based questions in a large-scale reading assessment (PIRLS), but it needed supervision to ensure the content quality of the questions in the AIG process.

The use of AIG studies by teachers in classroom assessments for getting ideas such as creating scenarios, preparing games, adding distractors to multiple-choice questions and improving assessment will increase the efficiency to be obtained from AI-supported AIG (Sherman et al., 2020). It is recommended that educators use these tools to guide them for the time being against the risks of the relevance, quality and content validity of an automatically generated measurement tool (Al-Worafi et al., 2023). In addition, efficiency will increase if the educators review the generated questions.

In summary, AIG with AI presents significant opportunities for educators. However, the generated items should not be used directly for measurement and evaluation purposes before being checked for accuracy, reliability and purposefulness by educators. In other words, educators should actively participate in the item generation process and assign AI an assistant role in supporting and complementing educators.

Automatic assessment and feedback

The concept of automatic text assessment was first introduced by Page (1966). This work laid the foundation for automatic assessment. As technology advanced, awareness of automatic assessment and feedback grew and led to the more widespread use of automatic assessment

platforms. Automated assessment platforms provide the opportunity to conduct exams and assignments electronically. These platforms are preferred because they allow question presentation with animation, video, and audio content that cannot be presented on paper. Additionally they produce objective results, perform automatic evaluations, and provide instant feedback by identifying students' strengths and weaknesses (Akgun & Greenhow, 2021; Kasneci et al., 2023; Owan et al., 2023; Swiecki et al., 2022; Wang et al., 2023). In these respects, they support student development, save time for the teacher, and enable rapid evaluation and feedback.

Advances in AI, have enhanced the potential to assess student-written texts, to provide feedback, and to reduce educators' grading time and, ultimately, have impacted classroom assessment processes (Huang et al., 2023; Jang, 2014). Classroom assessments are rooted in more diverse and enriched student responses compared to standardized tests. They aim to measure students' creative thinking, whether they have learned a topic in depth, and their written and oral communication skills (Warschauer & Grimes, 2008). For example, in essay assessments, although there is a strong relationship between machine and human scoring, machines focus on technical aspects such as word count and correct use of punctuation, while humans focus on skills such as fluency, completeness and creativity. Therefore, at this point, AI-supported automated assessment can increase the efficiency of human-led processes (Gardner et al., 2021). On the other hand, a teacher who realises that a student makes a mistake in a math problem knows that the source of this mistake may vary according to the student's visual defect, psychological state, and misconception. Therefore, teachers tailor the intervention method in the teaching process according to the student's situation. However, AI usually does not have the data to detect these differences while automating the decision-making process (Cardona et al., 2023). For this reason, it is recommended that AI-supported automated assessments should be used for formative assessment in schools until AI can assess similarly to humans (Cardona et al., 2023; Gardner et al., 2021).

On the other hand, despite the opportunities offered by AI-supported automated assessment software, it has the potential to provide unfair advantage or disadvantage to certain student groups. For example, in an exam canceled in the UK due to the Covid-19 pandemic, it is revealed that an algorithm that determines student results based on the success of schools in previous years provides an advantage to the students studying at private schools. This situation reveals that as a result of the decision making mechanism supported AI may let the unfair situations to continue (Akgun & Greenhow, 2021). For this reason, instead of fully automating the evaluation processes with support from AI, using automatic evaluation in situations where objective results will be obtained, such as evaluating multiple-choice questions and giving feedback, will increase the productivity of educators and students.

Responsible and participatory management

The ethical risks associated with the integrating AI in education extende beyond the issue of bias, cheating,and plagiarism in text creation, which fall under fairness. Data protection, data privacy, and accountability also stand as issues that should be confronted in measurement and evaluation processes (İlikhan et al., 2024; Huang, 2023; Lebovitz et al., 2021). In order to build measurement and evaluation processes that deal with these ethical risks, it is necessary to develop strategies to promote responsible AI practices (Theodorou & Dignum, 2020). Responsible AI discusses the question of who is responsible for the ethical use of AI. In this context, responsible AI encompasses AI developers, users, policy makers, societal norms, and even the system itself

(Ozer & Perc, 2024; Stahl, 2023). Therefore, in this section, the risks in the ethical dimension of measurement and evaluation processes are examined from a holistic perspective.

The ethical responsibility for using AI technologies, which develop as a dynamic ecosystem, is the responsibility of all stakeholders. Therefore, all stakeholders should take part in the integration of AI into measurement and evaluation processes, the functioning mechanism of the system should be open to users, and AI should be a system that is constantly monitored, evaluated and updated (Ozer et al., 2024a; Stahl, 2023). In this context, the ethical use encompasses both the development of AI technologies and the use of them. Therefore, both dimensions should be taken into consideration in order to create an AI ecosystem that is compatible with ethical and social values.

The rapid and profound advancement in AI technologies poses a risk that educators may lose control of their assessment processes related to AI. Since the control process is time-consuming and costly, there is a distance between educators' decisions and the evidence-gathering process on which these decisions should be based (Couldry, 2020; Swiecki et al., 2022). Additionally, the fact that AI algorithms are closed to users (blackbox) and controlled by a few companies raises concerns that AI will use student data through its ability to organize datasets, ultimately exacerbating concerns about trust in society (Khosravi et al., 2022; Sullivan et al., 2023; Stahl, 2023). For this reason, it is of great importance to work in a participatory way in which the development processes of AI systems to be used in education are open to all educational stakeholders and the system is constantly evaluated and updated (Ozer et al., 2024a). In other words, educators and educational administrators should be actively involved in the development processes of AI technologies used in education. This approach will also reduce the risk of educators being misled as it will increase their AI technology knowledge and awareness (Khosravi et al., 2022). In short, the participatory AI model will not only enable the development of more ethical AI practices by involving all stakeholders in the process, but also strengthen educators' immunity to the risks that these practices may pose (Ozer et al., 2024a). This approach is also important to ensure social acceptance of AI systems and to build trust among the society (Blasimme & Vayena, 2020; Huang, 2023; Ozer et al., 2024a).

On the other hand, it is necessary to rethink assessment systems to mitigate ethical risks in AI-influenced measurement and evaluation processes (Halaweh, 2023; Lancaster, 2023; Surahman & Wang, 2022). Existing measurement and evaluation approaches should be continuously updated in parallel with the advancements in AI. New tools to measure students' contribution to their work will increase academic honesty (Ozer, 2024). In this context, heterogeneous measurement and evaluation approaches should be integrated into the education systems to ensure fairness and objectivity (Yu, 2023). For example, measurement and evaluation approaches that reveal the student's contribution, such as oral presentations, laboratory activities, group work, and assignments with limited scope, will increase students' responsibility and significantly reduce the risk of using AI as a cheating tool (Sullivan et al., 2023).

Additionally, structured and clear instructions in measurement and evaluation processes encourage students to express their original thoughts and prevent them from engaging in unethical behaviors (Cotton et al., 2024). To reduce the risk of plagiarism and cheating, it is also recommended to compare current performence with previous ones (Sullivan et al., 2023). Assessors' close monitoring of students during the measurement and evaluation process is also an effective way to determine whether AI is being utilized (Cotton et al., 2024). Progress has been made in plagiarism detection with the development of tools that can detect how much of a text has

been written with AI. However, rapid developments in AI technologies often limit the reliability of these tools as well (Foltýnek et al., 2019; Khaled & Al-Tamimi, 2021). Therefore, such platforms are required to be closely monitored the developments in AI technology and be constantly updated.

On the other hand, the proliferation of AI increases the amount of data accumulated in AI systems every day, making it necessary to take serious steps to ensure data privacy (Yanisky-Ravid & Hallisey, 2018). In this sense, educational institutions and AI developers should determine data protection strategies together to ensure the security of students' data (Huang, 2023; Ozer et al., 2024a). At the same time, users of the AI system should have the freedom to protect their personal data and rights.

Certainly, the key approach to both the development and implementation of AI applications in education is to increase the AI literacy of students, educators and educational administrators (European Commission, 2022; Sok & Heng, 2023; Ozer, 2024). Thus, it will be possible to benefit from the advantages of these technologies with the least risk by increasing awareness of how to use them, their limitations and risks. Teachers' gaining competence in AI will support the more successful use of course contents and assessment and evaluation processes and will let them determine whether there is cheating in the assignments submitted by the students easily.

Discussion and Conclusions

AI technology affects and transforms all sectors in the context of the multifaceted opportunities it offers. Measurement and evaluation is one of the most important application areas of AI in education. The use of AI-supported text generation, individualised learning platforms, automatic item generation, automatic assessment and feedback applications is becoming widespread. Therefore, the effects of AI on students, educators and education administrators would increase swiftly (Ozer, 2024). In this context, the potential to produce content using AI in education raises ethical risks with texts being fully or partially authored by AI. In other words, the fact that AI tools using large language models such as ChatGPT offer the opportunity to perform textual tasks such as answering exam questions and writing essays poses the risk of deceiving assessment systems with masked performances. However, AI tools should play a role in supporting the development of students and teachers and organizing teaching processes (Owan et al., 2023). When assessment and evaluation systems lack the ability to detect ethical violations, AI is seen as a tool that supplies unfair advantage to students. An educational life sustained by AI-generated texts will distort the behavioural patterns of students in the short term and negatively affect the quality of human capital in the labour market in the long run.

On the other hand, AI-supported AIG promises to produce questions with high quantity and quality in a short time to be used in both large-scale exams and classroom assessment practices. Although there are still debates on the content quality of automatically generated items, progress is being made in producing items that are compatible with the skills targeted to be measured thanks to the natural language processing models. Although AIG in classroom assessments saves time for teachers, the items are still in need of teachers' supervision in terms of the appropriateness and accuracy of the content to the grade level. At the point where AI has reached today, the leadership of teachers cannot be given up in the process of AIG.

In addition, AI-supported automatic assessment applications also offer the potential to transfer assessment tasks to machines. Assessment processes that are completely left to machines have the potential to fail to monitor student development and produce unfair results due to the

limitations of AI such as algorithm biases, inability to be as effective as humans in assessing high-level skills (Acemoglu et al., 2023). An assessment process that integrates AI should offer a combination of automated and manual assessment methods to measure students' abilities with the least margin of error and support their learning processes. In this context, in automatic assessment processes, AI should be utilized in such a way that it can contribute to teachers and students such as giving feedback and making objective measurements strengthening indivualized learning.

Additionally, studies investigating the contributions of generative AI systems to employee performance have shown that these technologies most significantly enhance the performance of individuals with medium and low skill levels (Brynjolfsson et al., 2023; Noy& Zhang, 2023; Peng et al., 2023). In other words, the contribution of AI technologies to individual performance varies according to skill level, with the level of contribution decreasing as skill levels rise. These findings open up a significant opportunity for assessment and evaluation practices. First, these systems can significantly help reduce achievement gaps in schools by quickly addressing learning deficiencies, especially in low-skilled students. Secondly, considering that AI systems can rapidly elevate novices to targeted proficiency levels in workplaces (Alam et al., 2024; Korinek, 2023), teachers have the potential to quickly improve their measurement and evaluation skills and achieve a common convergence in measurement and evaluation.

The aspect of AI that supports student development such as generating ideas and providing guidance should be emphasized and the idea of using it as a cheating tool should be suppressed. In this context, the priority in AI systems should be given to educating the users (Khosravi et al., 2022; Sullivan et al., 2023). Because students need to be educated about academic honesty and awareness should be raised about the role of AI systems in this regard (Cotton et al., 2024). In other words, instead of deceiving individuals with short-term gains, the complementary aspects that constantly support them should be emphasized. Because those who can use this system efficiently will be ahead not only in their educational journey but also in the labour market (Ozer, 2024). When individuals make use of AI to organize their learning journey, they will gain a valuable companion in their educational journey (Lancaster, 2023).

Consequently, to deal with these ethical risks, AI should not go beyond being supportive and complementary tool for the individuals. To realize this goal, it is crucial to increase the AI literacy of all stakeholders of education. In order to achieve this goal, we suggest the development and use of processes of AI applications in a participatory way involving experts, students, teachers, administrators, and worker unions. On the other hand, it is critical to develop and continuously update digital platforms that will contribute to prevent ethical violations that may arise despite all precautions.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethics

In this study, ethical approval was not required as a review was conducted using previously published studies in the literature.

References

Acemoglu, D. (2024). *The simple macroeconomics of AI (No. w32487*). National Bureau of Economic Research. https://doi.org/10.3386/w32487

Acemoglu, D., & Restrepo, P. (2018). *Artificial intelligence, automation and work (No. w24196).* National Bureau of Economic Research. https://doi.org/10.3386/w24196

Acemoglu, D., Autor, D., & Johnson, S. (2023). Can we have pro-worker- AI? Choosing a path of machines in service of minds. *CEPR Policy Insight No.123*.

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, *2*(3), 431-440. https://doi.org/10.1007/s43681-021-00096-7

Alam, M. F., Lentsch, A., Yu, N., Barmack, S., Kim, S., Acemoglu, D., Hart, J., Johnson, S., & Ahmed, F. (2024). From automation to augmentation: Redefining engineering design and manufacturing in the age of NextGen-AI. *An MIT Exploration of Generative AI*. https://doi.org/10.21428/e4baedd9.e39b392d

Al-Worafi, Y. M., Hermansyah, A., Goh, K. W., & Ming, L. C. (2023). Artificial intelligence use in university: Should we ban ChatGPT? *Preprints* https://doi.org/10.20944/preprints202302.0400.v1

Aquino, Y.S.J. (2023). Making decisions: bias in artificial intelligence and data-driven diagnostic tools. *Australian Journal of General Practice*. *52* (7), 439-442. https://doi.org/10.31128/AJGP-12-22-6630

Athaluri, S. A., Manthena, S. V., Kesapragada, V. K. M., Yarlagadda, V., Dave, T., & Duddumpudi, R. T. S. (2023). Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus*, *15*(4). https://doi.org/10.7759/cureus.37432

Bezirhan, U., & von Davier, M. (2023). Automated reading passage generation with OpenAI's large language model. *Computers and Education: Artificial Intelligence*, *5*. https://doi.org/10.1016/j.caeai.2023.100161

Blasimme, A., & Vayena, E. (2020). What's next for COVID-19 apps? Governance and oversight. *Science*, *370*(6518), 760–762. https://doi.org/10.1126/science.abd9006

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27-40.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work (No. w31161)*. National Bureau of Economic Research.

Cardona, M. A., Rodríguez, R. J., & Ishmael, K. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations.* Office of Educational Technology, US Department of Education. https://www2.ed.gov/documents/ai-report/ai-report.pdf

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in education and teaching international*, *61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Couldry, N. (2020). Recovering critique in an age of datafication. *New Media & Society*, *22*(7), 1135-1151. https://doi.org/10.1177/1461444820912536

Du, X., Shao, J., & Cardie, C. (2017). Learning to ask: Neural question generation for reading comprehension. *arXiv preprint arXiv:1705.00106.* https://doi.org/10.48550/arXiv.1705.00106

European Commission. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators, *Publications Office of the European Union*. https://data.europa.eu/doi/10.2766/153756

Foltýnek, T., Meuschke, N., & Gipp, B. (2019). Academic plagiarism detection: a systematic literature review. *ACM Computing Surveys (CSUR), 52*(6), 1-42. https://doi.org/10.1145/3345317

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *PNAS, 116*(14), 6531-6539. https://doi.org/10.1073/pnas.190094911

Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment:'Breakthrough? Or buncombe and ballyhoo?'. *Journal of Computer Assisted Learning*, *37*(5), 1207-1216. https://doi.org/10.1111/jcal.12577

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, 15(2), ep421. https://doi.org/10.30935/cedtech/13036

Hosseini, M., Rasmussen, L. M., & Resnik, D. B. (2023). Using AI to write scholarly publications. Accountability in Research, 1-9. https://doi.org/10.1080/08989621.2023.2168535

Huallpa, J. J. (2023). Exploring the ethical considerations of using ChatGPT in university education. *Periodicals of Engineering and Natural Sciences, 11*(4), 105-115. http://dx.doi.org/10.21533/pen.v11i4.3770

Huang, L. (2023). Ethics of artificial intelligence in education: Student privacy and data protection. *Science Insights Education Frontiers, 16*(2), 2577-2587. https://doi.org/10.15354/sief.23.re202

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, *26*(1), 112-131. https://doi.org/10.30191/ETS.202301_26(1).0009

İlikhan, SU., Ozer, M., Tanberkan, H., Bozkurt, V. (2024). How to mitigate the risks of deployment of artificial intelligence in medicine? *Turkish Journal of Medical Sciences, 54(*3), *483-492*. https://doi.org/10.55730/1300-0144.5814

Jang, E. E. (2014). *Focus on Assessment-Oxford Key Concepts for the Language Classroom*. Oxford University Press.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Yj., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1-38.

Johnson, M. (2020, April 22). *A scalable approach to reducing gender bias in Google Translate*. Google Research Blog. https://research.google/blog/a-scalable-approach-to-reducing-gender-bias-in-google-translate/

Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, *15*(16), 12451. https://doi.org/10.3390/su151612451

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. https://doi.org/10.1016/j.lindif.2023.102274

Khaled, F., & Al-Tamimi, M. S. H. (2021). Plagiarism detection methods and tools: An overview. *Iraqi Journal of Science*, 2771-2783. https://doi.org/10.24996/ijs.2021.62.8.30

Khalil, M., & Er, E. (2023, June). Will ChatGPT get you caught? Rethinking of plagiarism detection. In *International Conference on Human-Computer Interaction* (pp. 475-487). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-34411-4_32

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, 100074. https://doi.org/10.1016/j.caeai.2022.100074

Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*, *61*(4), 1281-1317. https://doi.org/10.1257/jel.20231736

Lancaster, T. (2023). Artificial intelligence, text generation tools and ChatGPT – does digital watermarking offer a solution?. *International Journal for Educational Integrity, 19*(10). https://doi.org/10.1007/s40979-023-00131-6

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AIground truth really true? The dangers of training andevaluating ai tools based on experts' know-what. *MISQuarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/misq/2021/16564

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, *13*(410). https://doi.org/10.3390/educsci13040410

Mulla, N., & Gharpure, P. (2023). Automatic question generation: a review of methodologies, datasets, evaluation metrics, and applications. *Progress in Artificial Intelligence*, *12*(1), 1-32.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187-192. https://doi.org/10.1126/science.adh2586

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447-453. https://doi.org/10.1126/science.aax2342

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *EURASIA Journal of Mathematics, Science and Technology Education*, *19*(8), em2307. https://doi.org/10.29333/ejmste/13428

Ozer, M. (2024). Potential Benefits and Risks of Artificial Intelligence in Education. *Bartın University Journal of Faculty of Education*, *13*(2), 232-244. https://doi.org/10.14686/buefad.1416087

Ozer, M., & Perc, M. (2024). Human complementation must aid automation to mitigate unemployment effects due to AI Technologies in the labor market. *Reflektif Journal of Social Sciences,* 5(2), 503-514.

Ozer, M., Perc, M., & Suna, H.E. (2024b). Artificial Intelligence Bias and the Amplification of Inequalities in the Labor Market. *Journal of Economy Culture and Society*, 69, 159-168. https://doi.org/10.26650/JECS2023-1415085

Ozer, M., Perc,M., & Suna, HE. (2024a). Participatory management can help AI ethics adhere to the social consensus. *İstanbul University Journal of Sociology*, 44(1), 221-238. https://doi.org/10.26650/SJ.2024.44.1.0001

Page, E. B. (1966). The imminence of... grading essays by computer. *The Phi Delta Kappan*, *47*(5), 238-243.

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. *arXiv preprint arXiv:2302.06590*. https://doi.org/10.48550/arXiv.2302.06590

Perc, M., Ozer, M., & Hojnik, J. (2019). Social and juristic challenges of artificial intelligence. *Palgrave Communications*, *5*(1). https://doi.org/10.1057/s41599-019-0278-x

Qi, P., Zhang, Y., & Manning, C. D. (2020). Stay hungry, stay focused: Generating informative and specific questions in information-seeking conversations. *arXiv preprint arXiv:2004.14530*. https://doi.org/10.48550/arXiv.2004.14530

Rane, N. L., Choudhary, S. P., Tawde, A., & Rane, J. (2023). ChatGPT is not capable of serving as an author: ethical concerns and challenges of large language models in education. *International Research Journal of Modernization in Engineering Technology and Science*, *5*(10), 851-874. https://www.doi.org/10.56726/IRJMETS45212

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing?. *Critical care*, *27*(1), 75. https://doi.org/10.1186/s13054-023-04380-2

Sayın, A. & Gierl, M. (2024), Using OpenAI GPT to Generate Reading Comprehension Items. Educational Measurement: Issues and Practice, 43, 5-18. https://doi.org/10.1111/emip.12590

Scialom, T., Piwowarski, B., & Staiano, J. (2019, July). Self-attention architectures for answer-agnostic neural question generation. In *Proceedings of the 57th annual meeting of the Association for Computational Linguistics* (pp. 6027-6032), Florence, Italy. https://doi.org/10.18653/v1/P19-1604

Septiandri, A. A., Constantinidies, M., & Quercia, D. (2023). *The impact of AI innovations on US occupations.* Nokia Bell Labs, UK.

Sherman, A. T., Herman, G. L., Oliva, L., Peterson, P. A., Golaszewski, E., Poulsen, S., Scheponik, T., & Gorti, A. (2020). Experiences and lessons learned creating and validating concept

inventories for cybersecurity. In *National Cyber Summit Research Track 2020* (pp. 3-34). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-58703-1_1

Silva-Jurado, R. J., & Silva-Jurado, M. D. (2024). Educational innovation in the 21st century: gamification, artificial intelligence and art as transformative tools. *YUYAY: Educational, Methodologies & Didactics Strategies*, *3*(1), 35-52. https://doi.org/10.59343/yuyay.v3i1.60

Sok, S., & Heng, K. (2023). ChatGPT for education and research: A review of benefits and risks. *Cambodian Journal of Educational Research*, *3*(1), 110-121. https://doi.org/10.2139/ssrn.4378735

Stahl, B. C. (2023). Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems. *Scientific Reports*, *13*(1), 7586. https://doi.org/10.1038/s41598-023-34622-w

Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays - Should professors worry? *Nature*. https://doi.org/10.1038/d41586-022-04397-7

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature, 613*(7945), 620-621. https://doi.org/10.1038/d41586-023-00107-z

Suleyman, M. (2023). *The coming wave: Technology, power and the 21st century's greatest dilemma*. Crown.

Sullivan, M., Kelly, A. & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, *6*(1). https://doi.org/10.37074/jalt.2023.6.1.17

Surahman, E., & Wang, T.H. (2022). Academic dishonesty and trustworthy assessment in online learning: A systematic literature review. *Journal of Computer Assisted Learning, 38*(6), 1535–1553. https://doi.org/10.1111/ jcal.12708

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S.,Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 100075. https://doi.org/10.1016/j.caeai.2022.100075

Theodorou, A., & Dignum, V. (2020). Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence*, *2*(1), 10–12. https://doi.org/10.1038/s42256-019-0136-y

Thorp, HH. (2023). ChatGPT is fun, but not an author. *Science, 379*(6630), 313. https://doi.org/10.1126/science.adg7879

Varma, A., Dawkins, C., & Chaudhuri, K. (2023). Artificial intelligence and people management: A critical assessment through the ethical lens. *Human Resource Management Review*, *33*(1), 100923. https://doi.org/10.1016/j.hrmr.2022.100923

Verhoeven, F., Wendling, D., & Prati, C. (2023). ChatGPT: when artificial intelligence replace the rheumatologist in medical writing. *Annals of the Rheumatic Diseases, 82*(8), 1015-1017. https://doi.org/10.1136/ard-2023-223936

Wang, X., Gong, Z., Wang, G., Jia, J., Xu, Y., Zhao, J., Fan, Q., Wu, S.,Hu, W., & Li, X. (2023). ChatGPT performs on the Chinese national medical licensing examination. *Journal of Medical Systems, 47*(1), 86. https://doi.org/10.1007/s10916-023-01961-0

Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, *3*(1), 22-36. https://doi.org/10.1080/15544800701771580

Yanisky-Ravid, S., & Hallisey, S. (2018). 'Equality and privacy by design': Ensuring artificial intelligence (AI) is properly trained & fed: A new model of AI data transparency & certification as safe harbor procedures. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3278490

Yu, H. (2023). Reflection on whether ChatGPT should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology, 14.* https://doi.org/10.3389/fpsyg.2023.1181712

Zhang, M., Press, O., Merrill, W., Liu, A., & Smith, N. A. (2023). How language model hallucinations can snowball. *arXiv preprint arXiv:2305.13534.* https://doi.org/10.48550/arXiv.2305.13534