



Transforming Educational Assessment: Insights Into the Use of ChatGPT and Large Language Models in Grading

Chokri Kooli & Nadia Yusuf

To cite this article: Chokri Kooli & Nadia Yusuf (2025) Transforming Educational Assessment: Insights Into the Use of ChatGPT and Large Language Models in Grading, International Journal of Human-Computer Interaction, 41:5, 3388-3399, DOI: [10.1080/10447318.2024.2338330](https://doi.org/10.1080/10447318.2024.2338330)

To link to this article: <https://doi.org/10.1080/10447318.2024.2338330>



Published online: 15 Apr 2024.



Submit your article to this journal [↗](#)



Article views: 5062



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 45 View citing articles [↗](#)



Transforming Educational Assessment: Insights Into the Use of ChatGPT and Large Language Models in Grading

Chokri Kooli^a  and Nadia Yusuf^b 

^aFaculty of Social Sciences, University of Ottawa, Ottawa, Canada; ^bDepartment of Economics, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

The integration of Artificial Intelligence (AI) technologies in the field of education has prompted significant advancements, particularly in the domain of assessment and grading. This research delves into the potential of large language models, specifically OpenAI's ChatGPT, in simulating human-like interactions and accurately grading student assessments. To accomplish its objectives, the study compares the grading performance of ChatGPT with that of human correctors in a sample of second-year university students. The research findings indicate only a moderate correlation between the grades assigned by ChatGPT and those of human correctors, suggesting nuanced capabilities in providing comprehensive feedback and streamlining the grading process. While the study highlights the benefits of AI integration in education, it also addresses potential risks, including the exacerbation of educational inequalities and the limitations associated with AI's automated nature. This research contributes to the ongoing discourse surrounding AI's role in education, emphasizing the importance of striking a balance between AI and human instruction for optimal educational outcomes.

KEYWORDS

Artificial intelligence;
education assessment;
grading; ChatGPT;
automated systems

1. Introduction

The advent of Artificial Intelligence (AI) has been a topic of discussion for several decades, ever since being initially articulated as a distant idea in the 1960s. Throughout the first decades of the twenty-first century, the concept gained in popularity with the pervasive adoption of computer-assisted solutions in various sectors, including education. In particular, the role of AI-based services in grading and assessing student assignments has become increasingly relevant as can be observed in the evolving landscape of social sciences studies. Nevertheless, AI is an emerging technology that holds transformative potential for the future of humanity, while also prompting significant ethical concerns (Kooli, 2023a). By investigating the capability of AI to replicate human-like interactions in education and adequately grade student examination papers, this study seeks to uncover the potential of ChatGPT for automating and improving the grading process.

While artificial Intelligence has introduced innovative capabilities across various sectors, its impacts on the education sphere deserve a focused examination. By enabling machines to perform grading roles usually reserved for human professionals, advanced AI models introduce both opportunities and concerns (Barua et al., 2022; Kooli, 2023b). The evidence preceding the release of ChatGPT shows that AI-based integrations helped to facilitate a spectrum of educational technologies, from automated tutoring

systems to interactive early childhood education platforms. Furthermore, AI assists with a multitude of tasks by analyzing student performance data to promote personalized learning, generating tailored educational resources, and streamlining administrative tasks (Ahmad et al., 2022; Chen et al., 2023; Rastrollo-Guerrero et al., 2020). As exemplified by ChatGPT, the recent advancements in AI technologies have introduced a new era in AI-driven education (AIED) by promising enhancements in automated assessments and real-time feedback mechanisms (Hao, 2023; van Lent & Schmorow, 2022). These developments point to a pivotal shift in the educational paradigm, underscoring the importance of understanding and harnessing AI's potential to advance the teaching and learning experience.

Whereas AI's integration into educational settings proved to be comprehensive over the past decades, the core of this study pertains to the practical implementation of large language models (LLM) in grading. The development of Natural Language Processing (NLP) technologies transformed this landscape but, until recently, it has been limited to applying computer-based grading systems for questions with fixed responses (Alawida et al., 2023). Developed by OpenAI, ChatGPT models emerged at the forefront of educational developments as this LLM embodies key advancements in NLP by enabling machines to generate human-like text in response to the received inputs (Abdullah et al., 2022). The described capability makes ChatGPT an ideal candidate for assisting social sciences professors as well as

educators in other fields in grading exams. In addition to its potential for streamlining the grading process, the new technology can help to maintain the integrity and fairness that are vital in academic assessments.

For these reasons, the research question central to this study is: “How can advanced AI, like ChatGPT, be effectively utilized to assist professors in grading exams in social sciences while ensuring the integrity and fairness of the grading process?” This question reflects the pertinent necessity for evaluating and harnessing AI’s potential in academic settings, particularly in grading and assessment. By focusing on the application of ChatGPT in social sciences education, the current study seeks to contribute valuable insights into the development of auto-grading technology that could mark a significant step toward the integration of AI in education. Following the consequent Literature Review chapter focusing on extant research in the field, the study continues with the Research Methods and Results chapters that represent the core of this research. Finally, the study concludes with comprehensive Discussions and Conclusion chapters that encapsulate the ultimate contributions of this work as well as potential future implications.

2. Literature review

2.1. Defining AI in education

As the first step, it is essential to offer a comprehensive definition of Artificial Intelligence as a concept. As defined by Baker and Smith (2019, p. 10), AI encompasses computers performing cognitive tasks typically associated with those tied to human minds, such as learning and problem-solving. This definition encompasses a broad spectrum of technologies, including algorithms, applications, machine learning, neural networks, data mining, and natural language processing (Baker & Smith, 2019; Zawacki-Richter et al., 2019). Within the educational context, AI’s potential benefits are increasingly promising as they span from personalization and improved learning outcomes to increased access to education and efficient resource utilization (Yuskovych-Zhukovska et al., 2022; Zawacki-Richter et al., 2019; Zhang & Ma, 2023). The landscape of educational possibilities has broadened in recent years due to the adoption of ChatGPT and similar LLMs.

2.2. AIED innovations and enhancements

As discussed previously, AI introduced unique enhancements to educational systems that were previously deemed impossible or unfeasible. As highlighted by Chen et al. (2020) and Patel et al. (2022), the immediate outcomes can be seen in AI’s role in ensuring the overall quality of e-learning platforms while enabling personalized learning experiences. Additionally, AI appears to have a sizable impact on decision-making processes in the managerial departments of educational institutions by providing statistics and other supplementary data (Wang et al., 2023). The most striking features of AIED emerge from its role in

augmenting teaching, learning, and professional development with the help of intelligent tools (Al-Zyoud, 2020; Bindra et al., 2021). Unlike conventional educational software, these tools are capable of producing evidence-based strategies and predictive analytics that would ultimately enhance teaching methodologies.

Recent advancements in AIED and its capacity to advance educational practices can be appreciated in studies focusing on specific fields and domains. For example, Benítez et al. (2024) demonstrated the application of large language models in medical education. According to the authors, it is possible to enhance clinical skills development through adaptive and personalized learning processes facilitated by AI solutions. From another perspective, Mu and Zhao (2024) explored AI’s impact on entrepreneurship education as well as the possibility of augmenting one’s practical skills via intelligent simulations. In the research by Sain and Huri (2024), the authors focused on the so-called soft skills, such as critical thinking and problem-solving. The obtained results suggest that AI-based software is well-suited for providing a dynamic and individualized learning environment. Whereas many educators prioritize the application of AI in already established use cases, others emphasize AI’s significance in promoting creative thinking during micro-lessons by incorporating interactive and learner-centered approaches (Lüy et al., 2024). The AIED advancements in different fields provide a snapshot of AI’s current contributions when seeking to enhance conventional educational software and practices. For a more detailed exploration of the diverse applications of AI in education, refer to Table 1, which presents a variety of AI tools along with their use cases in different educational settings.

2.3. AI’s role in grading and assessment

When focusing specifically on the sphere of grading and assessment, AI’s applications hold transformative potential. For example, numerous studies have shown that AI can serve as personal tutors, provide intelligent support for collaborative learning, facilitate intelligent virtual reality experiences, and contribute to profiling and prediction tasks. One of the key aspects of AIED pertains to assessment and evaluation, a domain where AI tools can generate learning contexts and activities based on student characteristics and abilities. With the help of their advanced capabilities in personalized and adaptive learning, ChatGPT and other AI or AI-like tools can be instrumental in creating tailored educational experiences (Diwan et al., 2023; Luckin et al., 2016; Zhang & Ma, 2023). As a result, both educators and learners are expected to enjoy the benefits of an improved educational process due to the integration of ChatGPT and ChatGPT-based solutions.

One-to-one tutoring and collaborative learning stand out as key applications of AI in education. For the majority of disciplines, these two distinct approaches are capable of addressing the individualized needs of students and facilitating assessment tasks (Conati, 2009; García-Ros et al., 2018; Zawacki-Richter et al., 2019). For example, Intelligent

Table 1. Overview of AI applications in education.

AI application	Purpose in education	Impact/use case
AI chatbots and content generators (e.g., ChatGPT, Jasper)	Automated grading and student feedback; course content generation	Improved student engagement and personalized support (Labadze et al., 2023)
AI assessment tools (e.g., Gradescope, Teachflow.AI)	Automated grading for diverse assignments	Streamlined grading process, reduced workload for educators (Belda-Medina & Kokošková, 2023)
Writing tools (e.g., Grammarly, QuillBot)	Writing assistance and refinement	Enhanced student writing skills, improved assignment quality (Mageira et al., 2022)
Transcription tools (e.g., Otter.ai, Rev)	Transcription and note-taking	Accessible lecture content, aid for study and revision (Annuš, 2023)
Voice assistants (e.g., Amazon Alexa, Apple Siri)	Interactive learning and assistance	Enhanced learning through interactive study sessions (Al Shamsi et al., 2022)
Administrative tools (e.g., Gradescope, Querium)	Streamlining administrative tasks	Efficient resource management, time-saving on administrative tasks (Chen et al., 2020)
Project management tools (e.g., ClickUp, Asana)	Enhancing productivity in education settings	Improved project management and collaboration among students and educators (Okonkwo & Ade-Ibijola, 2021)

Tutoring Systems (ITS) can serve the following practical functions: enhancing individualized learning; promoting student engagement; and inducing socialization. Other authors highlight AI's inclusivity potential as many of the described solutions can assist students and teachers with special needs while also providing immersive educational experiences to those living in remote areas (Havrdá & Kloczek, 2023; Ladd, 2023; van Lent & Schmorow, 2022). From another standpoint, the growing interest in so-called flipped lessons, as described by Chace (2020), signifies a milestone in education due to their innovative approaches to performance assessment. Capable of integrating virtual reality, generating personalized content, and providing automated feedback, AI solutions have already started redefining traditional educational models.

2.4. Comparative analysis of AI-grading across disciplines

Few recent studies highlighted the practical feasibility of utilizing AI when evaluating and grading student work in educational settings. Although the novelty of the subject predicates the lack of relevant research, certain advancements in the scholarly field can already be established, particularly in regard to programming and data science courses. In their research emphasizing the AI's capacity for providing clear grading explanations, Tomić et al. (2023) highlighted the benefits of using AI techniques, including fuzzy rules and machine learning methods, for grading soft skills in programming courses. On the other hand, Swiecki et al. (2022) studied the automation of assessment tasks by AI and noted a strong shift in assessment responsibilities when comparing humans to machines. The authors coined the term "black-boxing of accountability" to describe a phenomenon of educators and students preferring AI assessments over conventional methods. Some of the reasons identified by the authors include impartiality and reduced grading disputes. Similar to the study by Tomić et al. (2023), Vittorini et al. (2021) demonstrated the effectiveness of an automated grading system for data science courses. However, this research also reveals significant correlations between AI-computed and traditional human grading, with the system's main benefits emerging from its efficiency and error

reduction capacity. As González-Calatayud et al. (2021) conducted a systematic review focusing on the role of AI in student assessment, their results reveal that in addition to providing feedback, AI tools are increasingly suited to automate the grading of student assignments. Nevertheless, only research by Grivokostopoulou et al. (2017) supported the immediate adoption of AI in grading exams by showcasing an educational system's capability to automate grading and provide immediate feedback. Other automated educational systems are yet to be fully developed as they don't exhibit similarly strong agreement levels with human tutors. The discussed findings collectively highlight AI's transformative potential in fundamentally changing grading processes in the near future. Many of the proposed AI-based grading and scoring solutions already show promising results and would be useful for enhancing learning experiences across various disciplines.

In most fields, scholars tend to agree that AI's integration into educational assessment and evaluation significantly enhances efficiency by automating tasks like grading and feedback provision. In turn, educators can use the freed-up resources to focus on other pedagogical roles (van Lent & Schmorow, 2022; Zawacki-Richter et al., 2019). While AI proves to be capable of assuming routine tasks, human instructors remain critically important due to their ability to forge social connections, respond to non-standard questions, and offer personalized support. This collaboration between AI and humans allows for a synergistic approach to education, with AI seamlessly handling repetitive grading tasks and human educators focusing on more complex, interactive aspects (van Lent & Schmorow, 2022). As clarified in the research of AI autograders by Li et al. (2023), the inclusion of advanced LLMs like OpenAI's ChatGPT and Google's BARD is expected to further the transformative potential of AI by offering the possibility of providing human-like interactions and detailed feedbacks. As such, the development of autograding technology remains a highly promising sphere in various fields and disciplines.

2.5. Challenges and limitations of AI in grading

While several authors focused on the capacity of AI in grading written exams, none were able to adequately explore the

extent to which modern LLMs can replicate human-like interactions and grade written assessments of social sciences students. Similarly, the feasibility of employing ChatGPT to accurately grade student assessments remains largely understudied and presents a critical research opportunity. Understanding the precise role and potential of LLMs in enhancing the quality of grading assessments and their capacity to simulate human interactions are essential areas for the current research. For these reasons, the following hypothesis emerges from the viewed literature: the utilization of LLMs in student assessment and grading will exhibit a significant correlation with human assessments, supporting the former's potential to provide comprehensive feedback and emulate human interactions. By evaluating the accuracy and effectiveness of LLMs in grading student assessments, the study aims to address knowledge gaps and provide insights into ChatGPT's role in optimizing the efficiency and quality of the grading process, thus contributing to the advancement of auto-grading technology in the field of social sciences.

3. Materials and methods

3.1. Subjects and research design

The study setting encompassed a mid-term exam in a second-year bachelor's degree program focused on key approaches and interventions in health, social protection, and education in developing countries. This course emphasizes developing critical thinking in the context of globalization and the privatization of public goods. Consisting of four development questions and 10 Multiple-Choice Questions, the exam required a total of 80 min for completion. The study's focus was further narrowed down to the writing question requiring reflection and critical thinking, with an average length of responses being 250 words. The specific question addressed in the experiment was related to the Canadian government's diversification of programs and measures concerning sustainable development goals. It was worth 7 points and had two parts:

3.1.1. Mention the two types of diversification

1. Diversification by nature, which includes legal, administrative, and operational measures/programs.
2. Demographic diversification, where measures/programs cover all segments of society, such as youth, elderly, immigrants, women, blacks, Inuit, Aboriginals, etc.

Each correct mention and explanation of these diversification types earned one point, with an additional point for explaining through examples.

3.1.2. Mention and explain three reasons justifying the usefulness of this diversification

1. Inclusiveness: Ensures all members of society are included, avoiding exclusion or marginalization.
2. Effectiveness: Makes measures more effective.

3. Realization of objectives: Enables the achievement of target objectives.

Each of these reasons earned one point for mentioning and explaining, totaling three points for this part of the question.

This question was chosen due to its significance in understanding how governmental policies can be inclusively designed to address a broad spectrum of societal needs, thereby reflecting the course's objectives to develop critical analytical skills among students. It was deemed particularly suitable for evaluation through AI grading, given its open-ended nature that demands reflective and analytical responses. A rubric, prepared in advance and designed to accommodate for potential ambiguities, guided the ensuing evaluation process. The rubric was made increasingly detailed for the purposes of outlining unambiguous criteria for assessing a student's depth of understanding, analytical capabilities, and the clarity of the arguments (Table 2).

In terms of grading, both ChatGPT and the human grader were presented with the same prompts, answers, and rubric presented above. The study was conducted using the 3.5 iteration of ChatGPT for AI grading due to the model's advanced natural language processing capabilities as compared to other alternatives. Moreover, this choice was further justified by the model's relevance to the academic context of the exams, as well as its universal accessibility and ever-increasing coverage in academic literature. The human grader was a PhD student with previous exposure to different types of grading systems, having evaluated several hundred papers in the same field of study. Being in their third year of PhD studies, the grader was deemed to have sufficient knowledge of the taught course in addition to previous grading process experience.

Although the inclusion of only one human grader and one set of AI-based grading for 25 respondents can be considered a limitation, this choice was purposeful to ensure a focused and controlled study. When comparing two highly distinct grading methodologies, the introduction of additional human graders and extra rounds of AI grading would potentially lead to inconsistencies, human error, and biases. For these reasons, the AI-grading process involved a single round, with the first round's results retained. This approach provided an additional opportunity to test the practical reliability and feasibility of AI-based grading in real-world academic scenarios. Despite students' right to request a review of their marks, no such requests were made, reflecting the general acceptance of the grading outcomes.

3.2. Data collection and analysis

For the purposes of the current research, a random sampling approach was used by selecting second-year university students from the social sciences discipline for analysis. Out of the original sample, including nearly 100 students, the first 25 submitted exam papers were analyzed in alphabetical order. Considering our research objectives, the chosen approach was deemed sufficient to obtain a representative dataset reflecting the grading performance and accounting for statistical significance in the consequent comparative

Table 2. Rubric used for the evaluation question.

Question	Potential answers	Rubric
Q1. In our course, we have seen that the Canadian government is diversifying its programs/measures related to sustainable development goals. (7 points) A. Mention the two types of diversification. (4 points)	1/ Diversification by nature: Measures/programs are varied (legal in nature, such as laws and regulations. Administrative, such as policies, conventions, and treaties. Operational, such as diversified programs) 2/ Demographic diversification: The measures/programs cover all spheres of society (youth, elderly, immigrants, women, blacks, Inuit, Aboriginals...)	One point for the right diversification and one point for the explanation through example. (1 + 1 + 1 + 1)
B. Mention and explain 3 reasons that justify the usefulness of this diversification. (3 points)	This diversification is essential to cover all the inclusive needs of society's members. It avoids the exclusion or marginalization of all components of society. It makes the measures taken more effective and enables the target objectives to be achieved.	One point for mentioning and explaining each one of these elements: Inclusiveness Exclusion or marginalization Effectiveness and realization of objectives

analysis. A larger dataset was not feasible at the time of research due to resource constraints. The graded sections were extracted from each student paper to be graded according to the study design.

During the following stage, each exam copy underwent initial human evaluation followed by AI-based grading using ChatGPT 3.5. Transparency and consistency were maintained by providing both the human corrector and ChatGPT with identical key answers and detailed mark distribution. A sample assessment input by the researchers and output by ChatGPT can be seen in [Figures 1](#) and [2](#), respectively. Throughout the data analysis stage, we utilized a systematic approach by comparing mean and standard deviations between the human corrector and ChatGPT 3.5. As these metrics are not illustrative out of context, the alignment between the two grading systems was further assessed using correlation coefficients, including the Pearson correlation coefficient (r) and Spearman's rank correlation coefficient, for understanding the linear and rank-order relationship between the two marking systems, respectively.

4. Results

4.1. Descriptive statistics

The comparison of mean scores reveals a mean of 2.92 out of 7 for the human corrector and a lower mean of 2.54 out of 7 for ChatGPT, indicating that, on average, ChatGPT's grading is slightly more conservative than that of the human corrector ([Table 3](#)). The standard deviations for the human corrector and ChatGPT are 0.702 and 0.978, respectively. This suggests that while both the human corrector and ChatGPT demonstrate consistency in grading, ChatGPT exhibits a higher variability in its grading pattern.

4.2. Correlation analysis

The Pearson's correlation coefficient ([Table 4](#)) between human corrector scores and ChatGPT scores is 0.46, indicating a moderate positive correlation. This signifies that, although there is some alignment between the two scoring

entities, it's not as strong as previously assumed. Despite the moderate strength of the correlation, it is important to add that its statistical significance emphasizes a reliable alignment between the two marking systems. The Spearman's rank correlation coefficient of 0.518, with a p -value of 0.008, indicates a statistically significant moderate positive correlation. This suggests that the ranking order of scores assigned by the human corrector and ChatGPT are generally similar but not perfectly aligned. As justified by the low p -value, the significance of this correlation underscores that the observed relationship is unlikely to be due to chance. The result reinforces the validity of the correlation observed between the two variables.

Further visual analysis of the grading differences between the human corrector and ChatGPT reveals a distinct vertical clustering of grades at key points, indicating a pattern of discrepancy in several grading areas ([Figure 3](#)). Specifically, the best-fit line demonstrates a moderate positive correlation as it starts from the point of agreement between both grading sources and then slightly diverges, suggesting an alignment with the increasing disparity in higher grades. The divergence is accentuated by the AI's use of non-integer points and a broader grading spectrum (0–5), compared to the human evaluator's scores (2, 3, or 4 points). These observations can potentially be attributed to different interpretations of the rubric. Moreover, the absence of the highest possible scores (6 or 7 points) from both graders could indicate a rigorous standard of assessment, high expectations set by the rubric, the nature of the received responses, or other factors that will be explored in the Discussion.

The Bland-Altman plot, with a mean difference of ~ 0.4 , highlights a general agreement between the two grading methods, with the limits of agreement framing the expected range of difference ([Figure 4](#)). The data's clustering and the presence of an outlier at a higher grade level accentuate the necessity for nuanced interpretation. Considering that the human corrector's grades have a limited range, consequent assessments of the consistency of the grading methods across the spectrum of student performance are recommended. In combination with Pearson's correlation coefficient, these visual representations underscore the potential

You

This is the question asked during the exam:

Q1. In our course, we have seen that the Canadian government is diversifying its programs/measures related to sustainable development goals.

A. Mention the two types of diversification. (4 points)

B. Mention and explain 3 reasons that justify the usefulness of this diversification. (3 points)

This is the correction scale:

A. Mention the two types of diversification. (4 points)

1/ Diversification by nature:
Measures/programs are varied (legal in nature, such as laws and regulations. Administrative, such as policies, conventions, treaties. Operational, such as diversified programs)

2/ Demographic diversification:
The measures/programs cover all spheres of society (youth, elderly, immigrants, women, blacks, Inuit, Aboriginals...)

Give One point for each right diversification and one point for each explanation through example. (1+1+1+1)

B. Mention and explain 3 reasons that justify the usefulness of this diversification. (3 points)

This diversification is essential to cover all the inclusive needs of society's members.
It avoids the exclusion or marginalization of all components of society.
It makes the measures taken more effective and enables the target objectives to be achieved.
Give one point for mentioning and explaining each one of these elements:
Inclusiveness
exclusion or marginalization
Effectiveness and realization of objectives

This is the answer of the student :

In two ways, the Canadian government is diversifying its programmes and measures related to the Sustainable Development Goals. Diversification of the programme, This entails developing and implementing a number of programmes that address various aspects of the SDGs. For example, the government may have one programme aimed at reducing poverty and inequality (SDG 1), another at promoting sustainable economic growth (SDG 8), and yet another at protecting biodiversity and natural resources (SDG 13). (SDG 15). Diversification of stakeholder groups, This entails involving a wide range of stakeholders in the development, implementation, and monitoring of SDG-related programmes and measures. Representatives from civil society organisations, the private sector, academia, and indigenous communities, among others, may be included. It encourages stakeholder participation and buy-in. The government can build partnerships and coalitions necessary for achieving SDGs by involving a diverse range of stakeholders in SDG-related initiatives. It allows for greater adaptability and flexibility. The government can respond to changing circumstances and emerging challenges more effectively and timely by implementing a variety of programmes and measures. Overall, programme and stakeholder diversification can assist the Canadian government in achieving its long-term development goals in a more comprehensive, inclusive, and adaptive manner.

↓

What is the student's final grade? Be flexible in grading.

Message ChatGPT...

Figure 1. A sample instruction input provided to ChatGPT 3.5.

of AI in grading while also emphasizing the importance of understanding its limitations and variances compared to human assessments.

In summary, the analysis indicates that while there is a moderate positive correlation between ChatGPT's grading and that of the human corrector, the strength of the alignment is not as robust as previously thought. The statistically significant correlations, both Pearson and Spearman, suggest a consistent but not particularly strong relationship. This result is promising for the development of automated grading systems but warrants additional experimentation. For this reason, further examination focusing on specific discrepancies and outliers can enhance the alignment and effectiveness of ChatGPT in grading scenarios.

5. Discussion

The actual research aimed to explore the extent to which large language models, such as OpenAI's ChatGPT 3.5, can effectively replicate human-like interactions and accurately grade written assessments of students, thereby contributing to the development of autograding technology in the field of education.

5.1. Comparison of grading outcomes

The first stage of the empirical research consisted of comparing the mean scores between the human corrector and ChatGPT. The mean score for the human corrector was 2.92 out of 7, higher than ChatGPT's mean score of 2.54 out of



Figure 2. A sample assessment output provided by ChatGPT 3.5.

Table 3. Descriptive statistics.

Grader	Mean score	Standard deviation
Human corrector	2.92	0.702
ChatGPT	2.54	0.978

Table 4. Correlation coefficients for human corrector vs. ChatGPT.

Method	p-Value	Coefficient
Pearson's correlation coefficient	0.02	0.45
Spearman's rank correlation coefficient	0.008	0.518

7. While this difference is not substantial, it indicates that ChatGPT's grading is comparable to that of the human corrector on average. Then, the research examined the standard deviations for both the human corrector and ChatGPT. The human corrector demonstrated a standard deviation of ~ 0.702 , whereas ChatGPT exhibited a slightly higher standard deviation of ~ 0.978 . These results suggest that the grading by both the human corrector and ChatGPT is consistent, albeit with slightly higher variability for ChatGPT. Furthermore, the correlation analysis revealed a Pearson's correlation coefficient of ~ 0.46 , indicating a moderate

positive correlation between the scores assigned by the human corrector and ChatGPT. Despite not being exceptionally strong, this correlation is still statistically significant and implies a reliable alignment between the two grading systems. As for Spearman's method, the rank correlation coefficient of 0.518 with a p -value of 0.008 reinforces the abovementioned finding and indicates a moderate yet significant alignment in the rank ordering of the scores by both the human corrector and ChatGPT.

Based on the observed grading patterns between the AI and human graders, several conclusions can be drawn. First, the absence of scores of 6 or 7 from either grader suggests a consistency in the assessment criteria or standards applied by both the AI and human grader. This consistency indicates that both assessors likely evaluated the submissions using similar metrics or guidelines, resulting in a convergence of scores within a certain range. Second, the disparity in the grading spectrum employed by the AI, ranging from 0 to 5, compared to the human grader, who only assigned scores of 2, 3, or 4, may indicate differences in their approaches to evaluation. While the AI grader adhered

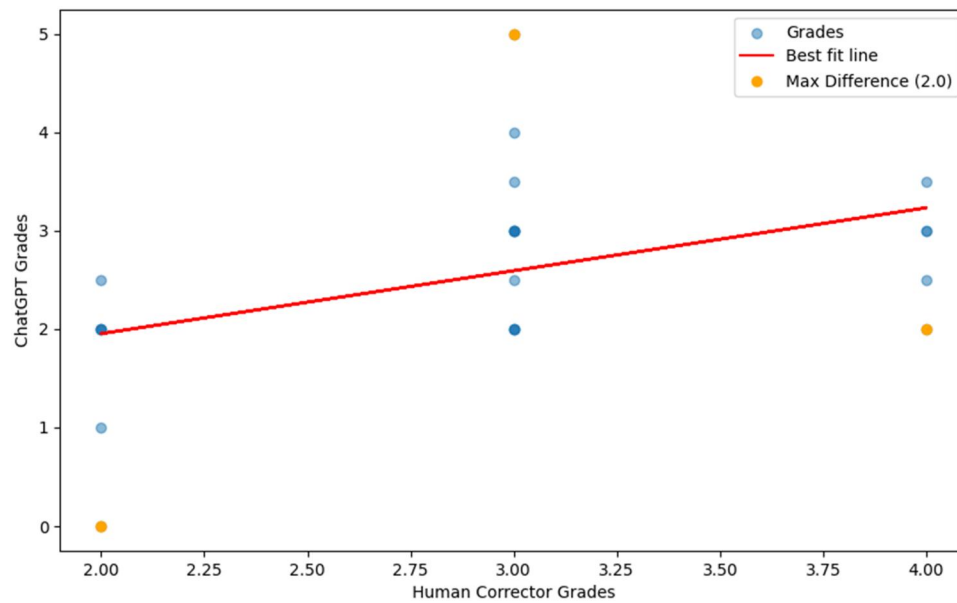


Figure 3. The correlation between human corrector and ChatGPT grades (based on the author's data).

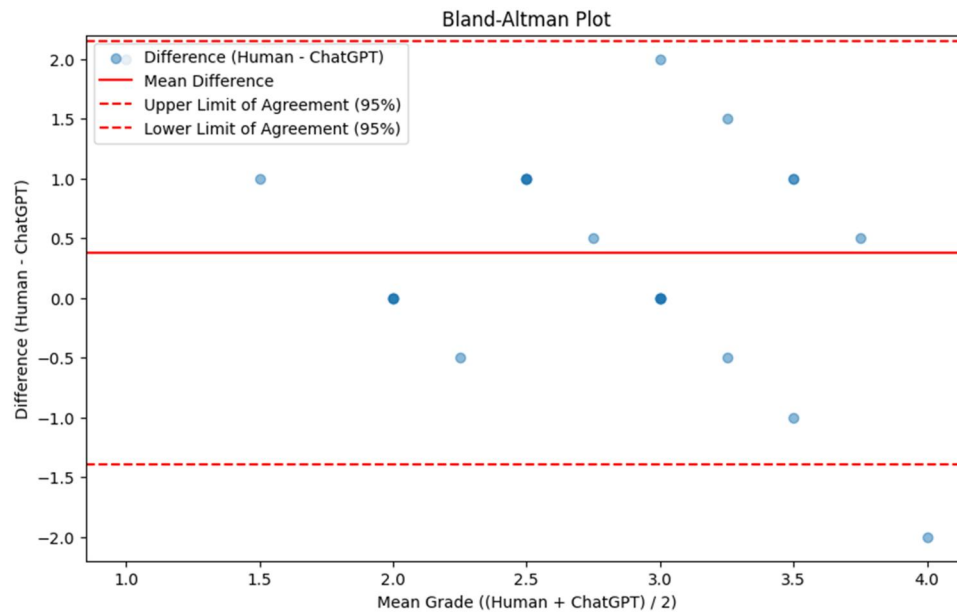


Figure 4. The Bland-Altman plot (based on the author's data).

strictly to the defined scoring spectrum, potentially reflecting a more objective assessment process, the human grader may have exercised subjectivity or discretion in assigning scores, possibly influenced by personal biases or sensitivities.

The findings additionally suggest that ChatGPT has the potential to simulate human-like interactions and provide comprehensive feedback, indicating promising prospects for the development of autograding technology in the field of education. Several researchers have acknowledged that LLMs can provide comprehensive feedback to students' post-assessments (Kuo et al., 2023). The empirical research was able to demonstrate the practical capabilities of LLMs in providing comprehensive feedback following written assessments. Concerning other aspects of examining the grading outcomes, notable patterns emerged in regard to the vertical clustering of grades at specific points, reflecting potential

discrepancies between human and ChatGPT assessments. As the trajectory of the best-fit line elucidates a moderate yet diverging alignment, it justifies a further discussion on its implications for grading consistency. Alongside clustered differences and a notable outlier, the Bland-Altman plot underscores the nuanced interconnection between human and AI grading systems. Although these insights are integral for understanding the grading dynamics, they also prompt a deeper exploration of the underlying factors contributing to these patterns. Overall, the research partially supports the hypothesis that the utilization of large language models like ChatGPT in student assessment and grading can demonstrate a degree of correlation with human assessments, signifying their potential to simulate human-like interactions and contribute to the advancement of auto-grading technology in the educational field.

5.2. Advantages of automated grading with LLMs

The research suggests that ChatGPT's grading has the capacity to enhance and complement assessments by human correctors, indicating the potential for improved efficiency in the grading process. This has the benefit of saving time and effort for educators (Vittorini et al., 2021), leading to a more streamlined assessment procedure. Thanks to more objective assessments, automated grading solutions can help mitigate biases and ensure uniformity across all students, an outcome positively characterized by its fairness. The ability of ChatGPT to generate immediate feedback and detailed performance reports can alleviate pressure on students and support personalized learning improvements. In cases of grading disputes, the impartiality of AI offers a transparent mechanism for dispute resolution (Figure 5).

When considering the advancements of AI in educational settings, it is critical to acknowledge the machine-learning capabilities of such tools. ChatGPT's ability to process data and generate human-like text using advanced algorithms and training data is a unique advantage that can be leveraged in education. Specifically, machine learning algorithms

can be actively fine-tuned in response to educator interactions and student feedback. This dynamic engagement with the AI tool would enable a more nuanced comprehension of student responses. Often characterized by its unprecedented consistency and scalability, ChatGPT can further address the inherent shortcomings associated with human graders, such as subjectivity and fatigue. As a result, the overall potential for automation and self-adjustment can be considered as an invaluable asset within educational settings.

The growing administrative responsibilities and grading burdens faced by professors represent a suitable foundation for the adoption of many AI-based services in different capacities. To prevent distractions from educators' primary roles, technological solutions like AI-driven tools can be deployed for the purposes of automated grading (Ahmad et al., 2022; Pavlova, 2008). Nevertheless, the scholars also highlight multiple disadvantages, such as the risks of misinformation, reliability issues, and intellectual property infringements (Castaldo, 2023; Hao, 2023). For these reasons, the decision to integrate or restrict AI technologies like ChatGPT in educational settings should be based on a well-informed scholarly consensus that would consider

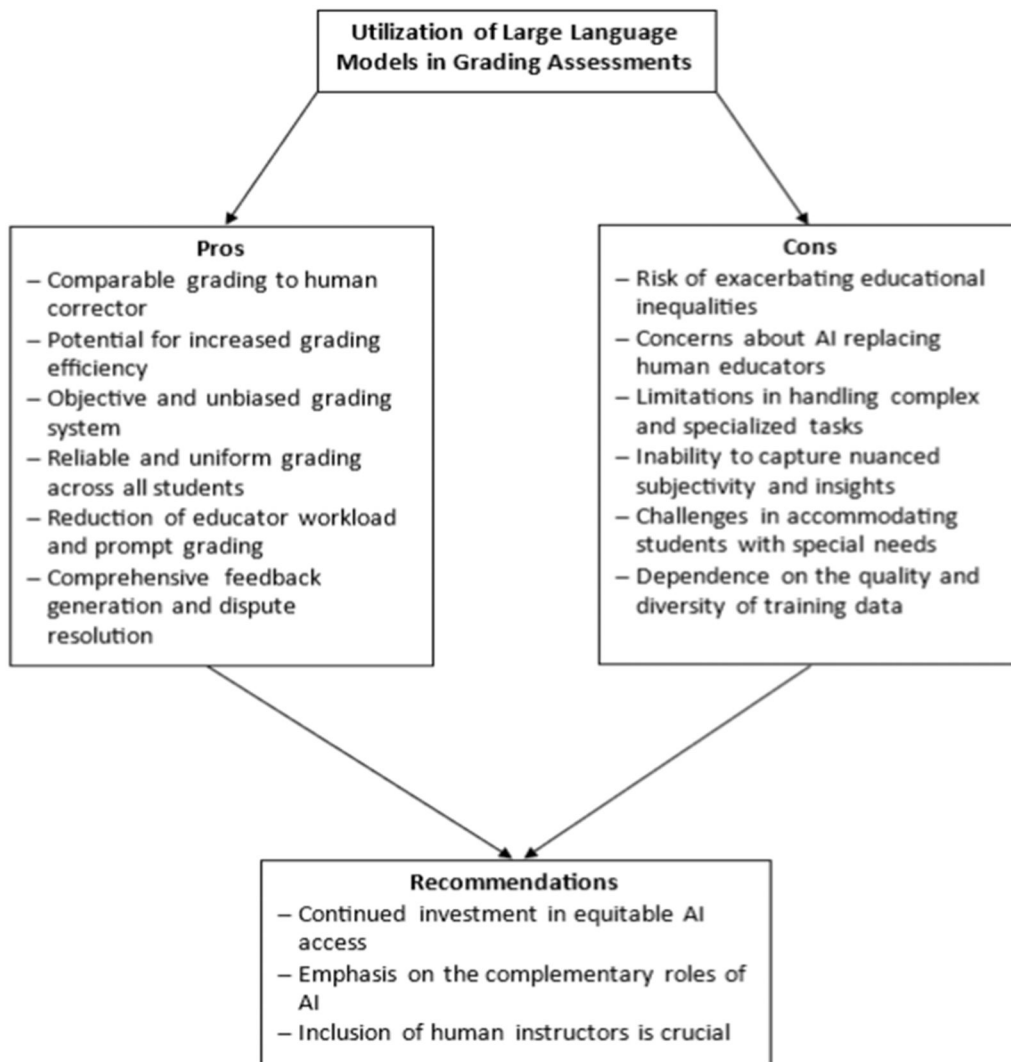


Figure 5. Utilization of large language models in assessments (pros and cons).

evidence bases from various fields (Kooli, 2023b; Meyer et al., 2023). As the role of AI in education is expected to increase, Ladd (2023) suggests educators to shift their focus to fostering key competencies like creativity and critical thinking to enhance learning experiences further. In this context, both students and educators are expected to actively engage in AI's multifaceted integration into educational paradigms.

5.3. Ethical and practical implications

The integration of AI in grading could pose significant risks (Figure 5) of exacerbating educational inequalities due to disparities in AI access and operational skills between developed and developing countries. Notably, existing AI accessibility challenges are likely to perpetuate educational disparities and create new ones (Diwan et al., 2023; Zawacki-Richter et al., 2019). The exploration of AI's different applications in education suggests the possibility of students interacting with LLMs like ChatGPT to enhance their learning experience. Although AI-based grading solutions are designed to be consistent and objective, a theoretical option is available for students to leverage these systems unilaterally. Through regular interactions with AI as a part of the learning process, students would be able to familiarize themselves with the grading criteria. Due to their improved understanding of the academic criteria and expectations, the validity of the consequent assessments could become compromised. The potential for exploiting AI grading systems should be carefully explored to ensure educational integrity and promote a holistic learning environment.

From another perspective, the balance between AI assistance and human instruction is crucial for ensuring that AI complements rather than replaces human educators. Completely substituting human exam graders with AI could result in the loss of millions of jobs, primarily held by university students. Alternatively, the cost savings from automating grading systems could be reallocated to students as scholarships. While these are real possibilities, the fear that AI could replace human educators understates the irreplaceable value of human interaction in fostering creativity and critical thinking (Zawacki-Richter et al., 2019). Finally, Marra (2023) emphasized the significance of adopting a data-driven, case-based, and evidence-informed policy-making approach, pointing to the use of evaluation information in the planning and implementation of policies. The adoption of such an approach is a prerequisite for integrating AI tools in education, whereas preserving the focus on equitable access and effective utilization across all regions. To mitigate risks and maximize benefits, future policies should emphasize the importance of maintaining educational integrity and promoting equitable learning opportunities.

5.4. Limitations and scope for future research

Although the research demonstrates promising results, the study's focus on relatively straightforward grading tasks may not fully represent the complexities of all types of student

assessments. In other words, ChatGPT's effectiveness might vary when faced with more intricate or specialized subject matter. It also appears that there may have been a lack of alignment in the grading criteria or a misunderstanding of how to apply them consistently. To improve alignment and consistency in the future, it may be beneficial to refine the grading criteria, provide more detailed prompts, and offer training or calibration sessions for both AI and human graders. This research was conducted on a limited sample of only 25 exam papers. Future studies should aim to expand the volume of data to comprehensively assess the potential of LLMs in simulating human-like interactions and contributing to the advancement of autograding technology in the educational field. The automated nature of ChatGPT's grading may also fail to capture the nuanced subjectivity and context-specific insights that human graders can provide. Creative expression, originality, and the specific needs of certain students may not be adequately recognized by an automated system unless explicitly programmed for these considerations. By addressing these limitations, we can harness the potential of AI while maintaining the integrity, fairness, and inclusivity of the educational experience.

6. Conclusion

The research demonstrates that the use of LLMs like OpenAI's ChatGPT in grading student assessments has the potential to assist in the grading process. While the results are promising when compared with those of human graders, it is important to recognize the current limitations in achieving perfect alignment. The correlation coefficient of 0.45 is indicative of a moderate positive relationship, which underscores the necessity for further adjustments of AI-based grading systems to improve their accuracy and consistency. This study highlights the innovative possibilities for incorporating AI into the process of grading educational assessments by demonstrating that with further development, LLMs could deliver reliable and consistent grading. In addition to enabling instructors to allocate more time to develop course content and reduce their grading workload, the swift grading process can alleviate student stress and provide constructive guidance for future improvements. In cases of disputes over human gradings, the automated system can serve as a reliable arbiter, providing more transparency and objectivity to the grading process and its outcomes.

6.1. Implications for educators and policy makers

As we are still exploring the potential of AI, it is vital to balance AI-driven tools with human expertise to ensure evaluations are both efficient and deeply contextual. From this standpoint, policymakers and educators should recognize the possibility of AI perpetuating biases and incorporate bias-mitigation strategies in the deployment of AI evaluation systems. Moreover, the development of training programs for educators and policymakers is critical for the effective use of AI tools. Such programs should be designed to focus

on fostering an understanding of AI capabilities and addressing integration challenges. To maintain the integrity and fairness of AI applications in education, the results point toward the need for establishing frameworks ensuring continuous monitoring, evaluation, and transparency of AI systems. The latter mainly involves regular assessments to address emerging issues as well as open disclosure of methodologies and data sources to ensure the integrity and credibility of the policymaking process.

6.2. Directions for future research

Future advancements in educational technology systems should prioritize the integration of AI-based grading technology. Incorporating LLMs, such as OpenAI's ChatGPT into prominent global educational technology solutions like Blackboard, Moodle, or Brightspace can present educators with a comprehensive array of automated grading solutions and student guidance services. These integrations have the potential to provide instructors with valuable insights and feedback regarding student performance and growth prospects. Additionally, the seamless evaluation and alignment of learning outcomes and objectives will become a more feasible and streamlined task with the aid of these integrated AI technologies. Continued research and development are necessary to enhance AI systems' capabilities, focusing on addressing complexities and accommodating diverse student needs.

6.3. Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used ChatGPT to improve readability and language. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Chokri Kooli  <http://orcid.org/0000-0002-8211-8621>

Nadia Yusuf  <http://orcid.org/0000-0003-3959-5120>

References

- Abdullah, M., Madain, A., & Jararweh, Y. (2022, November). ChatGPT: Fundamentals, applications and social impacts. In *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 1–8). IEEE. <https://doi.org/10.1109/SNAMS58071.2022.10062688>
- Ahmad, S. F., Alam, M. M., Rahmat, M. K., Mubarik, M. S., & Hyder, S. I. (2022). Academic and administrative role of artificial intelligence in education. *Sustainability*, 14(3), 1101. <https://doi.org/10.3390/su14031101>
- Alawida, M., Mejri, S., Mehmood, A., Chikhaoui, B., & Isaac Abiodun, O. (2023). A comprehensive study of ChatGPT: Advancements, limitations, and ethical considerations in natural language processing and cybersecurity. *Information*, 14(8), 462. <https://doi.org/10.3390/info14080462>
- Al Shamsi, J. H., Al-Emran, M., & Shaalan, K. (2022). Understanding key drivers affecting students' use of artificial intelligence-based voice assistants. *Education and Information Technologies*, 27(6), 8071–8091. <https://doi.org/10.1007/s10639-022-10947-3>
- Al-Zyoud, H. M. M. (2020). The role of artificial intelligence in teacher professional development. *Universal Journal of Educational Research*, 8(11B), 6263–6272. <https://doi.org/10.13189/ujer.2020.082265>
- Annuš, N. (2023). *Weigh the pros and cons of using artificial intelligence in education* (Online report). Nesta.
- Baker, T., & Smith, L. (2019). Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges. Retrieved from Nesta Foundation website https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf
- Barua, P. D., Vicnesh, J., Gururajan, R., Oh, S. L., Palmer, E., Azizan, M. M., Kadri, N. A., & Acharya, U. R. (2022). Artificial intelligence enabled personalised assistive tools to enhance education of children with neurodevelopmental disorders—A review. *International Journal of Environmental Research and Public Health*, 19(3), 1192. <https://doi.org/10.3390/ijerph19031192>
- Belda-Medina, J., & Kokošková, V. (2023). Integrating chatbots in education: Insights from the Chatbot-Human Interaction Satisfaction Model (CHISM). *International Journal of Educational Technology in Higher Education*, 20(1), 62. <https://doi.org/10.1186/s41239-023-00432-3>
- Benítez, T. M., Xu, Y., Boudreau, J. D., Kow, A. W. C., Bello, F., Van Phuoc, L., Wang, X., Sun, X., Leung, G. K.-K., Lan, Y., Wang, Y., Cheng, D., Tham, Y.-C., Wong, T. Y., & Chung, K. C. (2024). Harnessing the potential of large language models in medical education: Promise and pitfalls. *Journal of the American Medical Informatics Association*, 31(3), 776–783. <https://doi.org/10.1093/jamia/ocad252>
- Bindra, P., Kshirsagar, M., Ryan, C., Vaidya, G., Gupta, K. K., & Kshirsagar, V. (2021). Insights into the advancements of artificial intelligence and machine learning, the present state of art, and future prospects: Seven decades of digital revolution. In *Smart Computing Techniques and Applications: Proceedings of the Fourth International Conference on Smart Computing and Informatics* (Vol. 1, pp. 609–621). Springer Singapore.
- Castaldo, J. (2023, September 16). THE ABCs OF AI; AI depends on a global army of poorly paid online gig workers to teach the technology what it should and shouldn't do. *Globe & Mail*, B6. <https://link.gale.com/apps/doc/A765477089/HRCA?u=anon~3dbd05db&sid=googleScholar&xid=de6afc32>
- Chace, C. (2020, October 29). The impact of artificial intelligence on education. *Forbes*. <https://www.forbes.com/sites/calumchace/2020/10/29/the-impact-of-artificial-intelligence-on-education/?sh=21d2366d50df>
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
- Chen, A., Liu, H., Li, K. C., & Jia, J. (2023). For educational inclusiveness: Design and implementation of an intelligent tutoring system for student-athletes based on self-determination theory. *Sustainability*, 15(20), 14709. <https://doi.org/10.3390/su152014709>
- Conati, C. (2009). June Intelligent tutoring systems: New challenges and directions. In *Twenty-First International Joint Conference on Artificial Intelligence*.
- Diwan, C., Srinivasa, S., Suri, G., Agarwal, S., & Ram, P. (2023). AI-based learning content generation and learning pathway augmentation to increase learner engagement. *Computers and Education: Artificial Intelligence*, 4, 100110. <https://doi.org/10.1016/j.caeai.2022.100110>
- García-Ros, R., Pérez-González, F., Cavas-Martínez, F., & Tomás, J. M. (2018). Social interaction learning strategies, motivation, first-year students' experiences and permanence in university studies. *Educational Psychology*, 38(4), 451–469. <https://doi.org/10.1080/01443410.2017.1394448>

- González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences*, 11(12), 5467. <https://doi.org/10.3390/app11125467>
- Grivokostopoulou, F., Perikos, I., & Hatzilygeroudis, I. (2017). An educational system for learning search algorithms and automatically assessing student performance. *International Journal of Artificial Intelligence in Education*, 27(1), 207–240. <https://doi.org/10.1007/s40593-016-0116-x>
- Hao, K. (2023). What is ChatGPT? What to know about the AI Chatbot that will power Microsoft Bing. *The Wall Street Journal*. Retrieved October 20, 2023, from <https://wsj.com/articles/chatgpt-ai-chatbot-app-explained-11675865177>
- Havrdá, M., & Kloczek, A. (2023). Well-being impact assessment of artificial intelligence – A search for causality and proposal for an open platform for well-being impact assessment of AI systems. *Evaluation and Program Planning*, 99, 102294. <https://doi.org/10.1016/j.evalprogplan.2023.102294>
- Kooli, C. (2023a). Artificial intelligence dissociative identity disorder (AIDIS): The dark side of ChatGPT. *QScience Connect*, 2023(2), 2. <https://doi.org/10.5339/connect.2023.2>
- Kooli, C. (2023b). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability*, 15(7), 5614. <https://doi.org/10.3390/su15075614>
- Kuo, B.-C., Chang, F. T. Y., & Bai, Z.-E. (2023). Leveraging LLMs for adaptive testing and learning in Taiwan Adaptive Learning Platform (TALP) [Paper Presentation]. In Empowering Education with LLMs – the Next-Gen Interface and Content Generation, July 7, 2023, Tokyo, Japan (pp. 101–110). <https://ceur-ws.org/Vol-3487/paper6.pdf>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Ladd, T. (2023, January 21). Teaching in the age of artificial intelligence and ChatGPT. *Forbes*. <https://www.forbes.com/sites/tedladd/2023/01/21/teaching-in-the-age-of-artificial-intelligence-and-chatgpt/?sh=1469c027788e>
- Li, T. W., Hsu, S., Fowler, M., Zhang, Z., Zilles, C., & Karahalios, K. (2023, August). Am I wrong, or is the autograder wrong? Effects of AI grading mistakes on learning. In *Proceedings of the 2023 ACM Conference on International Computing Education Research* (Vol. 1, pp. 159–176). <https://doi.org/10.1145/3568813.3600124>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson. <https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/innovation/Intelligence-Unleashed-Publication.pdf>
- Marra, M. (2023). Editorial. *Evaluation and Program Planning*, 100, 102352. <https://doi.org/10.1016/j.evalprogplan.2023.102352>
- Lüy, Z., Bakirci, F., Artsın, M., Karataş, S., & Çakmak, E. K. (2024). Innovative technologies to trigger creative thinking in micro-lessons. In *Optimizing education through micro-lessons: Engaging and adaptive learning strategies* (pp. 118–140). IGI Global.
- Magiera, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI chatbots for content and language integrated learning. *Applied Sciences*, 12(7), 3239. <https://doi.org/10.3390/app12073239>
- Meyer, J. G., Urbanowicz, R. J., Martin, P. C. N., O'Connor, K., Li, R., Peng, P.-C., Bright, T. J., Tatonetti, N., Won, K. J., Gonzalez-Hernandez, G., & Moore, J. H. (2023). ChatGPT and large language models in academia: Opportunities and challenges. *BioData Mining*, 16(1), 20. <https://doi.org/10.1186/s13040-023-00339-9>
- Mu, Q., & Zhao, Y. (2024). Transforming entrepreneurship education in the age of artificial intelligence. *Resources Data Journal*, 3, 2–20. https://doi.org/10.50908/rdj.3.0_2
- Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033. <https://doi.org/10.1016/j.caeai.2021.100033>
- Pavlova, M. (2008). *Technology and vocational education for sustainable development: Empowering individuals for the future* (Vol. 10). Springer Science & Business Media.
- Patel, D., Shah, D., Modi, P., & Roy, M. (2022). Artificial intelligence powered material search engine. *Materials Today: Proceedings*, 57(1), 11–13. <https://doi.org/10.1016/j.matpr.2022.01.120>
- Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. *Applied Sciences*, 10(3), 1042. <https://doi.org/10.3390/app10031042>
- Sain, Z. H., & Huri, B. M. (2024). Fostering 21st century skills: AI integration for innovative education enhancement. *Socrates Journal of Interdisciplinary Social Studies*, 10(37), 53–67. <https://doi.org/10.5281/zenodo.10499018>
- 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>
- Tomić, B. B., Kijevčanin, A. D., Ševrač, Z. V., & Jovanović, J. M. (2023). An AI-based approach for grading students' collaboration. *IEEE Transactions on Learning Technologies*, 16(3), 292–305. <https://doi.org/10.1109/TLT.2022.3225432>
- van Lent, M., & Schmorow, D. (2022). The applications of artificial intelligence to education and training. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 19(2), 127–128. <https://doi.org/10.1177/15485129221088717>
- Vittorini, P., Menini, S., & Tonelli, S. (2021). An AI-based system for formative and summative assessment in data science courses. *International Journal of Artificial Intelligence in Education*, 31(2), 159–185. <https://doi.org/10.1007/s40593-020-00230-2>
- Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., Anandkumar, A., Bergen, K., Gomes, C. P., Ho, S., Kohli, P., Lasenby, J., Leskovec, J., Liu, T.-Y., Manrai, A., ... Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972), 47–60. <https://doi.org/10.1038/s41586-023-06221-2>
- Yuskovych-Zhukovska, V., Poplavska, T., Diachenko, O., Mishenina, T., Topolnyk, Y., & Gurevych, R. (2022). Application of artificial intelligence in education. Problems and opportunities for sustainable development. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 13(1Sup1), 339–356. <https://doi.org/10.18662/brain/13.1Sup1/322>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1–27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhang, L., & Ma, Y. (2023). A study of the impact of project-based learning on student learning effects: A meta-analysis study. *Frontiers in Psychology*, 14, 1202728. <https://doi.org/10.3389/fpsyg.2023.1202728>

About the authors

Chokri Kooli holds a PhD in Economics from the School for Advanced Studies in the Social Sciences (Paris). He has more than 17 years of teaching and research experience. Actually, he works as part time professor at the University of Ottawa. Kooli serves also as UNESCO Inclusive Policy Lab Expert.

Nadia Yusuf has a doctorate in economics, and her specialization is microeconomics. She is presently working as a professor in the Economics Department at King Abdulaziz University, Jeddah, Saudi Arabia. Her research interests are economics, finance, banking, sustainable development, education, socioeconomics, international relations, and entertainment.