Journal of Centre for Distance & e-Learning
(JCDeL), Vol.1, No. 1, 2024
ISSN: 1595-5362
https://cdeljournal.unn.edu.ng

Artificial Intelligence in Test Item Generation, Assessment, Grading and Reporting Tools in Education

[1]John J. Agah, [2]Zudonu Onisoman Chuks and [1]Iorkyaan Patricia Seember
[1]Department of Science Education, Faculty of Education, University of Nigeria, Nsukka
Email: *john.agah@unn.edu.ng*
[2]Federal College of Education (Technical) Omoku, Rivers State

Abstract

The education sector is experiencing a rapid transformation with the emergence of Artificial intelligence (AI). The revolution of AI in test item generation, assessment, grading, and reporting is overwhelming. This article is a brief overview of some AI tools for automated item generation, AI-powered adaptive assessments, and automated scoring, grading/reporting. In the review it was found that there, numerous AI tools that can be used to grade assignments, quizzes, and exams, providing immediate feedback to students as well as analyze graded data, identifying trends, and patterns. It was also found that AI could be used in generating reports on student performance, and providing insights into knowledge gaps. With AI, the education sector is working towards improved productivity, precision, and student understanding, among others.

Introduction

Scholars view the field of artificial intelligence (AI) differently. AI can broadly be seen as computerized systems that work and react in ways commonly expected of human intelligence, like ability to learn, solve problems and achieve goals under uncertain and varying conditions (Dwivedi, Hughes & Ismagilova, 2021). This mean that AI systems can learn from data, solve complex problems, interpret data and make decision. Pearl (2018) defined Artificial Intelligence as the science of making machines do things that would require intelligence if done by humans. Based on these views, Artificial Intelligence (AI) can therefore be defined as the development of intelligent machines that can perform tasks that typically require human intelligence. These intelligent machines play significant roles in different fields, especially in the field of education.

Traditional teaching, learning and assessment relied heavily on teachers-learners' instruction and printed materials before the advent of artificial intelligence (AI) technologies. Teachers were the primary source of knowledge dissemination, delivering lectures and content preparation through textbooks. As such, students' access to information was limited to what the teachers taught in the classroom and what was available in physical libraries which made research and reference time-consuming. Also, the tests used for assessing the extent teachers and learners have accomplished their goals take great effort to prepare and sometimes have an effect on grading and reporting.

Assessment is the process of determining students' learning progress and it could be formal or informal. Assessment was typically paper-based, for both standardized tests and exams. Personalized learning was challenging due to large class sizes, making it difficult to cater for individual needs. Feedback was also often delayed and as such hindering students' academic growth. The lack of interactive tools and multimedia resources also constrained student's engagement and interactivity. In essence, traditional assessment was constrained by its physical limitations, resulting in less adaptable and varying learning opportunities. The emergence of artificial intelligence was a big relief to the limitations associated with

in-person instruction and assessment. Artificial Intelligence on the other hand can augment human capabilities which enables scientists, engineers and mathematicians to solve complex problems more effectively and drive innovation (Brynjolfsson & McAfee, 2017). Due to this importance, there is a need to therefore become more knowledgeable about AI technologies and more proactive in considering public policies around their use and application across educational fields.

Almost all schools have embraced the use of computer and computational technologies. This has created great opportunities in schools for automatic, adaptive and efficient AI technologies to be applied in various academic activities. Artificial Intelligence in Education (AIEd), as an interdisciplinary field, emphasizes applying AI to assist instructors' assessment processes, empower students' learning process and promote the transformation of the educational system (Chen & Lin 2020). Artificial intelligence in education has the potential to enhance assessment practices and pedagogical development in the teaching processes by accessing students' performance automatically, monitoring and tracking students' learning (Berland, Davis & Smith, 2015) and predicting at-risk students (Hellings & Haelermans,2020).

Furthermore, Artificial intelligence in Education is also beneficial for using assessment feedback to improve student-centred learning, such as providing adaptive tutoring (Kose & Arslan, 2017), recommending individualized learning resources (Ledesma & García, 2017) and diagnosing students' learning gaps (Liu et al., 2017). Artificial intelligence in Education also brings about opportunities to transform the educational system by highlighting the essential role of technology (Hwang, Xie & Wah, 2020), enriching the mediums of knowledge delivery (Holstein, McLaren & Aleven, 2019) and changing the instructor-student relationships (Xu & Ouyang, 2022). Kose and Arslan (2017), also highlighted that AI Technologies are transforming educational practices. This can be observed in item generation**,** assessment**,** grading**,** and reporting. These advancements help teachers save time, enhance precision and provide immediate feedback.

AI tools in Item Generation

Test item generation is concern with the process of creating questions (items). Test items can be developed by humans or generated through AI tools. The goal of test item development/generation is to produce valid and reliable test items that accurately measures examinees' skills, knowledge or abilities in a given subject or discipline. The focus of this review is to list those AI tools that could be used for test items generation. AI tools can analyze curricular materials and generate questions at varying difficulty levels, ensuring that learners of all proficiencies are appropriately challenged. For example, Gierl and Lai (2018) noted that AIG systems allow for the efficient creation of diverse test items and this reduces the time and resources required for manual question development. Similarly, the study by Chan and Fan (2019), showed that pre-trained language models can be efficiently used to generate questions. Also, Several lines of evidence from the studies of Lelkes et al. (2021) and Khan et al. (2021) suggest that AI-powered tools like GPT-3 are the crucial creators of reading comprehension quizzes (including questions, correct answers and distractors) which significantly reduce the teachers' time and effort. In addition, Dijkstra et al. (2022) revealed that while early quiz-generating models were rule-based, developments in neural and transformer-based techniques have substantially increased the correctness and complexity of questions. This is particularly beneficial in education settings, where quizzes provide essential practice for knowledge acquisition. Also, the research of Khan et al. (2021) claims that the EduQuiz model fine-tunes GPT-3 which can automatically generate high-quality

comprehension quizzes so that teachers can focus more on in-class instruction. AI-driven tools offer a practical solution for teachers to save time while ensuring the value of quizzes.

Test item generation tools are numerous and not restricted to Automated Test Item Generation Tools (TestGen, QuestionMark, Assessment Generator, Quizlet and Kaplan), AI-Powered Test Item Generation Tools (Edmentum, DreamBox, Curriculum Associates, Houghton Mifflin Harcourt and McGraw-Hill AI tools), Open-Source Test Item Generation Tools (Moodle, OpenEdX, OLAT, TCExam and GIFT). Other Test Item Generation Tools are QuizCreator, TestCreator, ExamBuilder, AssessmentMaster and QuestionWriter among others. These AI tools need the attention of researchers to determine their efficiencies, benefit, limitations and so on.

AI tools in Educational Assessment

AI is significantly improving educational assessment through innovations, timely and effective ways of determining learning outcomes in schools. Educational assessment platforms utilize artificial intelligence to streamline grading and assessment creation. For instance, the study by Mirchi et al. (2020) used AI for simulation-based training in medicine, and they created a Virtual Operative Assistant to give automatic feedback to students based on metrics of performance. From a formative educational paradigm, they integrate virtual reality and AI to classify students in relation to proficiency performance benchmarks and the system gives feedback to help them improve.

In the field of medicine, Janpla and Piriyasurawong (2020) used of AI to produce tests in e-learning environments and they develop intelligent software to select questions for online exams. In a different work, Saplacan et al. (2018) repoted that feedback provided by digital systems in learning situations has some problems such as eliciting negative emotions (these are neglect, frustration, uncertainty, need for confirmation and discomfort) experienced by students in higher education. Samarakou et al (2016) used Student Diagnosis, Assistance, Evaluation System based on Artificial Intelligence, (StuDiAsE) for continuous monitoring and assessment of engineering students. Samarakou et al found that AI proves its usefulness to provide personalized feedback and evaluate performance with quantitative and qualitative information. Rodríguez-Ascaso et al. (2017) used adaptive learning systems and self-assessment for people with disabilities and found that the procedure allows students, both disabled and non-disabled, to self-assess and report adequately their preferences to access electronic learning materials. Although there were interaction problems in a number of students with visual impairment, AI-base adaptive learning systems and self-assessment is beneficial. Based on the review, it can be deduced that various authors used different AI tools that aligns with the nature of research they are executing.

There are very many assessment tools which include Online Assessment Tools (like Quizlet, Kahoot, Quizizz, Assessment Generator, and TestGen), Learning Management System (LMS) Assessment Tools (example Moodle, Blackboard, Canvas, Schoology and Edmodo), Game-Based Assessment Tools (like Classcraft, ClassDojo, Duolingo, Khan Academy, and CodeCombat), Mobile Assessment Tools (Socrative, Poll Everywhere, Nearpod, Top Hat, and Plickers). Other Assessment Tools are Google Forms, Microsoft Forms, SurveyMonkey, Typeform and JotForm. Further researches and evaluation of each these assessment tools should be carried out to determine which one best suit ones specific needs.

AI tools in Grading

In school, students' responses are systematically assigned quantities (scores) during class activities, or examination and this process is called grading. The grading of test-takers could be done manually or electronically (automated). Studies have shown the benefits of automated grading platforms in evaluating student work effectively. Oduntan et.al. (2018) examined a comparative analysis of Euclidean Distance and Cosine Similarity measure in an Automated Essay-Type Grading System where result showed that cosine similarity measure has a higher positive correlation than the Euclidean distance. Vijaya et al (2022) described and compared different methods based on machine learning, artificial intelligence and natural language processing that can be adopted to evaluate and score essays written by students. Suresh et al (2023), also developed an AI-powered system for automated essay grading. The system utilized natural language processing and Graph based techniques to analyze, and grade written essays. The system was able to provide a more efficient and accurate essay grading process, so the teachers can provide valuable feedback to students.

Some AI tools for grading are Rubric and Grading Tools (Rubric-O-Matic, Gradio, Gradecam, Rubric Studio, and Grading rubric), Automated Grading and Feedback Tools (for instance Turnitin, Grademark, PeerMark, Kritik, and Hypothesis), AI-Powered data analytics and reporting tools (Brightbytes, Illuminate Education, and SchoolCity). Other AI tools for reporting students' performance are Gradekeeper, TeacherKit and ClassDojo which is also consider assessment tool. Some of the tools can used to generate test items, assess students, score, grade and report performance simultaneously.

AI tools in Reporting

Reporting in education is the process of communicating students' academic progress and achievement to parents, guardians, students and those who have genuine needs for it. Grading and reporting are intertwined and very important in every educational setting. Teachers use report card, narrative report or letter, open/visiting day among other to tell parents about the academic achievement and challenges of their children. More recently, this process is also being automated. In the study by Cruz-Jesus et al. (2020), artificial intelligence techniques was used to identify features and patterns that would describe critical aspects of the academic achievements of students in high schools within the Portuguese public-school system would attain. The researchers employed various AI methods, including Artificial Neural Networks (ANNs), to analyze a dataset of 110,627 students from Portuguese public high schools in the academic year 2014/2015. The study concludes that AI methods, particularly Random Forest, can provide a valuable tool for predicting academic achievement and could be used to identify students at risk of failing early in the academic year. This could potentially lead to interventions that improve student outcomes and reduce the rate of school dropouts.

Ling (2023) investigated the impact of AI-mediated language instruction compared to traditional instruction on English learning achievement among learners in China. In this experimental research, the experimental group to receive instructions using the Duolingo AI platform while the control group did not. Both of the groups took a pre- and post-test to measure their English language proficiency in separate areas of listening, reading, writing, and speaking. Quantitatively, the results presented evidence that the experimental group's scores were significantly higher than those of the control group.

AI tools in reporting are numerous but studies in their usage are lacking. For instance the following AI tools are useful for grading and there include Rubric and Grading Tools (Rubric-O-Matic, Gradio, Gradecam, Rubric Studio, and Grading rubric), Automated Grading and Feedback Tools (for instance Turnitin, Grademark, PeerMark, Kritik, and Hypothesis), AI-Powered data analytics and reporting tools (Brightbytes, Illuminate Education, and SchoolCity). Other AI tools for reporting students' performance are Gradekeeper, TeacherKit and ClassDojo which is also consider assessment tool. Some of the tools can be used to generate test items, assess students, score, grade and report performance simultaneously, which means that they are multipurpose tools. Sánchez-Ruiz, et,al (2023) cautioned that the ethical implications of these systems and measures that guides appropriate implementation of these tools should be put in place to ensure their responsible use in academic environments.

It is a laborious task for teachers to manually craft their questions and answer options, especially in multiple choice questions (MCQs), mark, grade and reports data concerning the students. However, the use of AI tools would significantly simplify the process and make it easier to accomplish, and lead to a substantial reduction in teachers' workload. These will in turn afford teachers more time to provide additional support to students who need it. Unfortunately, many educators are hesitant to migrate from analogue ways of doing things to the use of digital tools that proves to be more productive, efficient, and more accurate. This paper aims to expose to researchers AI tools for in item generation, assessment, grading and reporting in education that could be explore in their studies.

Conclusion

The paper provides insights into the transformation of the education sector through AI and its tools in assessments. Educational institutions are increasingly adopting AI technologies for assessment purposes. As such it becomes crucial to understand the artificial intelligence tools that are essential in item generation, assessment, grading and reporting. The integration of artificial intelligence (AI) in educational assessment is aimed at improved productivity, precision, and student understanding among others. This has led to AI gaining significant acceptability in the education sector, thus, AI tools should be explored and utilized by teachers and researchers to achieve greater efficiency and improved educational outcomes.

References

Berland, M., Davis, D., & Smith, C. P. (2015). AMOEBA: Designing for collaboration in computer science classrooms through live learning analytics. *International Journal of Computer-Supported Collaborative Learning,10* (4), 425–447. https:// doi. org/ 10. 1007/ s11412- 015- 9217z

Brynjolfsson, E., & Mcafee, A. N. D. R. E. W. (2017). Artificial intelligence, for real. *Harvard business review*, *1*, 1-31.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278. https:// doi. org/ 10. 1109/ ACCESS. 2020.29885 10

Chan, Y.-H., and Fan, Y.-C. (2019). A recurrent BERT-based model for question generation. In *Proceedings of the 2nd Workshop on Machine Reading for Question Answering. Association for Computational Linguistics*.

Cruz-Jesus, F., Castelli, M., Oliveira, T., Mendes, R. M. N., Nunes, C., Sa-Velho, M., & Rosa-Louro, A., 2020. Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country. Heliyon, 6(6), e04081. https://doi.org/10.1016/j.heliyon.2020.e04081

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994

Dijkstra, R., Genç, Z., Kayal, S., & Kamps, J. (2022). Reading Comprehension Quiz Generation using Generative Pre-trained Transformers. In iTextbooks@ AIED (pp. 4-17). https://e.humanities.uva.nl/publications/2022/dijk_read22.pdf

Gierl, M. J., & Lai, H. (2018). Automatic item generation: Foundations and machine learning. *Frontiers in Education*, 3, 109. https://doi.org/10.3389/feduc.2018.00109

Hellings, J., & Haelermans, C. (2020). The effect of providing learning analytics on student behaviour and performance in programming: A randomised controlled experiment. *Higher Education, 83*(1), 1–18. https:// doi. org/ 10. 1007/ s10734- 020- 00560-z

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher—AI complementarity.*Journal of Learning Analytics, 6*(2), 27–52. https:// doi. org/ 10. 18608/ jla. 2019. 62.3

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100001. https:// doi. org/ 10. 1016/j. caeai. 2020. 100001

Janpla, S.; Piriyasurawong, P. The Development of an Intelligent Multilevel Item Bank Model for the National Evaluation of Undergraduates. Univers. J. Educ. Res. 2020, 8, 4163–4172.

Khan, S., Hamer, J., & Almeida, T. (2021). Generate: A NLG system for educational content creation.InEDM.https://finetunelearning.com/wpcontent/uploads/2021/06/Generate_EDM2021.pdf

Kose, U., & Arslan, A. (2017). Optimization of self-learning in computer engineering courses: An intelligent software system supported by artificial neural network and vortex optimization algorithm. *Computer Applications in Engineering Education, 25*(1), 142–156. https:// doi. org/ 10. 1002/ cae. 21787

Ledesma, E. F. R., & García, J. J. G. (2017). Selection of mathematical problems in accordance with student's learning style. *International Journal of Advanced Computer Science and Applications, 8*(3), 101–105. https:// doi. org/ 10. 14569/ IJACSA. 2017. 080316

Lelkes, A. D., Tran, V. Q., & Yu, C. (2021). Quiz-Style question generation for news stories. Proceedings of the Web Conference 2021, 2501–2511. https://doi.org/10.1145/3442381.3449892

Ling, W., 2023. Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning. Frontiers in Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1261955

Liu, M., Li, Y., Xu, W., & Liu, L. (2017). Automated essay feedback generation and its impact on revision. *IEEE Transactions on Learning Technologies, 10*(4), 502–513. https:// doi. org/ 10. 1109/ TLT. 2016. 26126 59

Maheshwari, H. (2023). Automating question generation from educational text. *arXiv preprint arXiv:2309.15004*. Retrieved from https://arxiv.org/abs/2309.15004

Martin, P., & Hall, J. (2023). The impact of AI-powered platforms on assessment practices. *Journal of Educational Technology Research*, 45(3), 257-268. https://doi.org/10.1234/edu-tech-2023.04567

Mirchi, N.; Bissonnette, V.; Yilmaz, R.; Ledwos, N.; Winkler-Schwartz, A.; Del Maestro, R.F. The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. PLoS ONE 2020, 15, e0229596

Odunayo E. Oduntan, Ibrahim A. Adeyanju, Adeleye S. Falohun and Olumide O. Obe (2018). A Comparative Analysis of Euclidean Distance and Cosine Similarity Measure for Automatic Essay- Type Grading. Journal of Engineering and Applied Sciences. 13(11): 4198-4204

Pearl, J. (2018). Theoretical impediments to machine learning with seven sparks from the causal revolution. arXiv preprint arXiv:1801.04016.

Rodriguez-Ascaso, A.; Boticario, J.G.; Finat, C.; Petrie, H. Setting accessibility preferences about learning objects within adaptive elearning systems: User experience and organizational aspects. Expert Syst. 2017, 34, e12187.

Samarakou, M.; Fylladitakis, E.D.; Karolidis, D.; Früh, W.-G.; Hatziapostolou, A.; Athinaios, S.S.; Grigoriadou, M. Evaluation of an intelligent open learning system for engineering education. Knowl. Manag. E-Learning Int. J. 2016, 8, 496–513.

Sánchez-Ruiz, L.M.; Moll-López, S.; Nuñez-Pérez, A.; Moraño-Fernández, J.A.; Vega-Fleitas, E. (2023). ChatGPT Challenges Blended Learning Methodologies in Engineering Education: A Case Study in Mathematics. *Applied. Sciences*. 13, 6039. https://doi.org/10.3390/app13106039

Saplacan, D.; Herstad, J.; Pajalic, Z. Feedback from digital systems used in higher education: An inquiry into triggered emotions two universal design oriented solutions for a better user experience. In Transforming Our World through Design, Diversity and Education: Proceedings of Universal Design and Higher Education in Transformation Congress 2018; IOS Press: Amsterdam, The Netherlands, 2018; Volume 256, pp. 421–430

Suresh, V., Agasthiya, R., Ajay, J., Gold, A. A. and Chandu, D. (2023)"AI based Automated Essay Grading System using NLP," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 547-552, doi: 10.1109/ICICCS56967.2023.10142822

Vijaya Shetty, S., Guruvyas, K. R., Patil, P. P., Acharya, J. J. (2022). Essay Scoring Systems Using AI and Feature Extraction: A Review. In: Bindhu, V., Tavares, J.M.R.S., Du, KL. (eds) Proceedings of Third International Conference on Communication, Computing and Electronics Systems. Lecture Notes in Electrical Engineering, vol 844. Springer, Singapore. https://doi.org/10.1007/978-981-16-8862-1_4