INTERNATIONAL RESEARCH JOURNAL OF HUMANITIES AND INTERDISCIPLINARY STUDIES
(Peer-reviewed, Refereed, Indexed & Open Access Journal)
DOI : 03.2021-11278686 ISSN : 2582-8568 IMPACT FACTOR : 7.560 (SJIF 2024)

UNPACKING EXPLAINABLE AI (XAI) IN EDUCATION: A COMPREHENSIVE REVIEW AND OUTLOOK

Dr. Vidyullata S. Jadhav
Assistant Professor,
V. P. Institute of Management Studies Research,
Sangli (Maharashtra, India)
E-mail: vsjadhav@vpimsr.edu.in

Dr. Tejashree T. Moharekar
Assistant Professor,
Yashwantrao Chavan School of Rural Development, Shivaji University,
Kolhapur (Maharashtra, India)
E-mail: ttm.50649@unishivaji.ac.in

DOI No. 03.2021-11278686 **DOI Link ::** https://doi-ds.org/doilink/01.2024-96855284/IRJHISIC2401018

Abstract:

AI, or artificial intelligence, refers to the process of using computers and other technologies to make them smarter than people and to assist them in solving problems that they encounter on a daily basis. Some examples of artificial intelligence technologies are machine learning and deep learning. These technologies make use of algorithms to improve their ability to predict future events without the assistance of a human being. Explainable Artificial Intelligence, often known as XAI, is a sort of artificial intelligence that is capable of explaining to humans the reasoning behind a decision or a prediction that it has made. When it comes to crucial duties like security, healthcare, or money, the objective of XAI is to make artificial intelligence systems more open, trustworthy, and accountable. This is especially important when these systems are employed. Within the scope of this essay, an orderly evaluation of the literature on XAI methodologies that can be applied in a variety of contexts, particularly within the education sector, is presented. The purpose of this study is to initiate a comprehensive investigation of XAI in the field of education, with the goal of examining its significance, potential applications. This will be accomplished by examining the current situation as well as possible future advances of XAI in education. We believe that our systematic review contributes to the existing body of research on XAI by pointing the way for additional research to be conducted in this field.

Keywords: *Artificial Intelligence, Machine learning, Deep learning, Explanation, Explainable Artificial Intelligence, student performance*

INTRODUCTION:

Applying artificial intelligence (AI) in a way that humans can understand the findings is what XAI is all about. The "black box" idea of machine learning stands in stark contrast, as not even the creators of the system can deduce the reasoning behind an AI's decision-making process. It's possible that XAI is a way to put the social right to explanation into practice. There may not be a legal or regulatory need for XAI, but it can still be useful; for instance, it might increase end users' faith in an AI's decision-making abilities, which in turn improves the product or service's usability.

By doing so, XAI hopes to shed light on the past, present, and future of an activity while also revealing the data that formed the basis for it. All three of these features allow us to verify, question, and develop new ideas about the world.

Education is only one area where artificial intelligence (AI) has quickly spread and become ubiquitous. The potential for it to transform learning, individualised instruction, and administrative work is enormous if it finds its way into school systems. But in educational contexts, where openness is key, the intrinsic complexity and opaqueness of AI algorithms create enormous obstacles to comprehending the decision-making process.

To help teachers, students, and lawmakers understand, trust, and improve AI-powered educational systems, Explainable AI (XAI) is emerging as a crucial path to decipher the 'black box' aspect of AI algorithms. To guarantee the ethical, fair, and accountable use of AI in education, the pursuit of transparency in these systems is essential, not just a technological problem.

RELATED WORK:

The convergence of Artificial Intelligence (AI) and education has attracted significant interest in recent years, as academics investigate the capabilities of AI-powered systems to individualise learning experiences and enhance educational results. Nevertheless, the lack of transparency in AI algorithms used in educational environments has generated apprehension regarding their comprehensibility, reliability, and ethical ramifications. This review summarises the main topics and discoveries from the current body of work on Explainable Artificial Intelligence (XAI) and its significance in educational settings.

Methods of machine learning are helpful in addressing concerns raised by students and parents over the most favourable career path to pursue. The extraction of knowledge from previous contacts is accomplished through the use of various methods and algorithms in data mining. For the purpose of providing students with well-informed recommendations for career guidance, the machine learning-based system makes use of advanced technologies such as artificial intelligence, deep learning, neural networks, and natural language processing (Guleria, P & Sood, M. , 2023).

It is common for parents to seek career counselling in order to address their concerns about the future of their children. When it comes to assisting parents in making well-informed judgements regarding the educational path that is most suitable for their children, a guiding cell is absolutely necessary. Within the scope of this discussion, machine learning will be employed in order to carry out the computational intelligence duties in an intelligent and efficient manner

(Yang, C, Huan, S, & Yang, Y. , 2020).

Explainable artificial intelligence (XAI), machine learning (ML), and the internet of behaviour (IoB) are examples of technologies that are having an effect on the development of smart

education in this era of smart cities. The use of these technologies makes it possible to personalise and customise educational experiences. The purpose of this study is to offer a framework for intelligent education that incorporates technologies such as Internet of Behaviour (IoB) and Explainable Artificial Intelligence (XAI). For the purpose of determining the extent to which the existing educational systems are able to respond appropriately to the requirements of students, the study makes use of data that was gathered on the behaviours of students. In spite of the fact that there are advanced education systems, they have not yet reached the degree of development that is necessary to personalise instruction in accordance with the cognitive needs of students and to provide support in situations where there is no face to-face instruction. A total of forty-one students participated in the study, and data was collected on their actions in relation to academic activities. Additionally, the study explored whether or not operating systems were able to accurately capture and respond to academic behaviours. The research employed evaluation methods that established a direct connection between the adoption of Internet of Things and Internet of Business monitoring and the academic improvement of students, hence providing a proactive support system for the success of the students (Embarak, 2022).

The application of Artificial Intelligence (XAI) that is capable of ensuring a higher level of ethical conduct, transparency, and the capacity to be scrutinised in the utilisation of educational data. A dataset consisting of students from the Federal Institute of Rio Grande do Norte (IFRN), a technical university located in Brazil, was used in this study to forecast the probability of students dropping out of school. Explainable Artificial Intelligence (XAI) approaches were utilised and reviewed in order to make this prediction. In order to achieve this goal, a comprehensive review of the existing body of literature was carried out in order to develop a checklist of evaluation measures. An innovative explainability score was developed as a result of this research in order to evaluate XAI frameworks. Authors hope that by doing this, they will be able to make it simpler to employ XAI models to enhance our comprehension of data connected to schools, which will in turn reinforce the significant research efforts that are being made in this field (Melo E, Silva I, Costa DG, Viegas CMD, & Barros TM, 2022).

A thorough investigation into the application domains and workloads was carried out in order to acquire an understanding of the educational domains that have benefited from the use of XAI. After conducting a comprehensive review of a number of papers from a variety of fields, the author presented their findings. There were around fifty percent of the publications that did not deal solely to any particular subject among the papers that were chosen. The healthcare industry was responsible for publishing fifty percent of the remaining total publications. Other areas of interest for researchers working in XAI include the business world, the transportation sector, the legal system, the

entertainment industry, academic institutions, and a variety of other departments (Islam MR, Ahmed MU, Barua S, & Begum S, 2022).

Explainable AI (XAI), a relatively new subfield of artificial intelligence, has the potential to be implemented in educational environments, namely Virtual Learning Environments (VLEs), in order to supply teachers with statistics that are simple to comprehend regarding the academic performance of a number of students who are enrolled in online classes. When compared to traditional machine learning (ML) and deep learning (DL) methods, explainable artificial intelligence (XAI) is distinguished by the fact that it enables instructors to gain a comprehensive comprehension of the specific capabilities and limitations of each individual student. Education professionals are able to provide individualised feedback and guidance in a timely manner as a result of this. Through the utilisation of demographic, clickstream, and assessment information, a number of different classical and ensemble machine learning algorithms were trained in order to ascertain the strategy that results in the most ideal performance outcome. Ultimately, the most effective machine learning technique was chosen and utilised as an input for the XAI model in order to evaluate the study behaviour of students at various phases of the course, both on a local and global basis. The researchers have employed a variety of Explainable Artificial Intelligence (XAI) methodologies in order to provide instructors with reports that are both clear and understandable regarding the performance of students at various points throughout the course they are teaching. By utilising intermediate data analysis and performance reports, educators and other essential stakeholders will be able to make decisions that are well-informed and provide support to students who are engaged in online learning in an efficient manner (M. Adnan, et al., 2022).

According to the authors, significant measures of learning were developed from the interaction data that was collected in order to evaluate the level of involvement exhibited by the students. After taking into consideration these characteristics, the TrAdaBoost-based transfer learning model was made a suggestion. The training of the model through the use of past course interaction data was followed by its application to the current study semester. This was done in order to evaluate the model's capacity to generalise and predict the engagement behaviour of new students. Despite the fact that the most recent data were not sufficient for training the model, the results of the experiments demonstrated that the model was capable of achieving an impressive degree of pinpoint accuracy and overall performance. Furthermore, the technique was beneficial in assisting teachers in providing support to academically vulnerable students who were typically at risk of dropping out of school or performing poorly in their academic pursuits (H. Wan, K. Liu, Q. Yu, & X. Gao, 2019). Students need to model continuously daily, hourly, or even minute-to-minute—in order to anticipate dropout rates, which is a time-series challenge, according to the findings of a study. The suggested

model was able to incorporate the regularisation component. This was accomplished through the use of a logistic regression model. When it came to identifying children who were at danger of dropping out of school, the Input-Output Hidden Markov Model (IOHMM) performed better than the baseline machine learning and deep learning models. It was possible to reach an accuracy rate of 84% (A. A. Mubarak, H. Cao, & W. Zhang, 2020).

An intriguing study was conducted in which the facial expressions of students, which include body language, eye movements, and expressions, were analysed to determine the correlation between their internet activity and their facial expressions. We employed a wide range of criteria to determine how the students behaved while watching instructional films, and we documented the students' shifting levels of involvement throughout the process. Following that, a variety of output labels were utilised in order to annotate the components of student behaviour that were connected with the level of student involvement. A new multiple-instance learning framework was implemented so that it would be possible to accurately determine the degrees of engagement that the online students were experiencing at various points throughout the video. Massive open online courses (MOOCs) and virtual learning environments (VLEs) are two examples of organisations that can use this platform to create educational videos (A. Kaur, A. Mustafa, L. Mehta, & A. Dhall, 2018).

The Learning Management System, sometimes known as LMS, is a widely used education management platform that is utilised by schools in order to monitor the actions of their students. Deep learning neural networks, also known as LSTM networks, were utilised by the authors of a study in order to conduct an analysis of the temporal study activity of online students. A student's performance could be broken down into minute-by-minute or second-by-minute chunks, which is an example of a temporal study behaviour. Such challenges are sometimes referred to as problems with time series, which is another name for them. An analysis of the results revealed that LSTM networks performed significantly better than typical machine learning models when it came to identifying the time-series behaviour of students. With the help of time series data, such as clickstreams from students, long-term short-term memory (LSTM) networks were able to detect students who were at risk of dropping out of school or who were on the verge of dropping out. In addition, when compared to more conventional approaches to machine learning, deep learning models exhibit superior performance on tasks that are related to time series and have improved generalizability (F. Chen & Y. Cui, 2020).

It was suggested by the authors that an early warning system be implemented. Through the use of the dashboard visualisation, both the students and the instructors were able to gain a comprehensive understanding of the progress that the students had made. In light of this, a feedback

prediction system was designed to provide educators with assistance in the delivery of customised treatments, with the ultimate goal of lowering the likelihood of students leaving school before their time. According to the findings of the evaluation, the early warning system was able to precisely identify students who were at danger of failing as well as the most common reasons for students to drop out of school (D. Baneres, M. E. Rodriguez, & M. Serra, 2019).

The authors provided a comprehensive presentation of the current explainability status of the collaborative-filtering recommender system (RS) that is utilised by the "WhoTeach" educational platform. This system's primary objective is to develop new educational programmes and courses on the basis of the results of an experiment that was carried out on a specific group of users. This is accomplished through the use of a cooperative evaluation technique. A substantial experimental finding suggests that it is of the utmost importance to significantly improve the explainability of WhoTeach features. It goes on to claim that attentional mechanisms must to be improved in order to make it possible for instructional materials to be suggested in accordance with the user's individualised profile and requirements. Through the utilisation of attention weights that have already been subjected to statistical testing, these models make it possible to understand the reasoning behind the model's selected proposals (Luca Marconi, et al., 2020).

FUTURE TRAJECTORIES AND RECOMMENDATIONS:

It is without a doubt that the function of XAI in the development of accountable artificial intelligence systems will grow more prominent as educational technology continues to advance. It is recommended that future research concentrate on reducing the impact of algorithmic biases, developing user-friendly XAI approaches that are suited to the needs of educational stakeholders, and building rigorous frameworks for the ethical deployment of artificial intelligence in educational settings.

THE IMPERATIVE OF XAI IN EDUCATION:

The confluence of artificial intelligence and education has the potential to change various elements of education, including teaching, learning, and administrative procedures. To recap, this convergence has the ability to reform more than one component of education. The adoption of explainable artificial intelligence is not merely a technical necessity when it comes to shaping the future of education; rather, it is a fundamental condition that must be reached in order to cultivate trust, promote transparency, and guarantee the ethical and equitable use of AI. Understanding how AI can be explained is a fundamental condition that must be satisfied. When stakeholders push for the incorporation of XAI into educational frameworks, they have the opportunity to take use of the revolutionary power of artificial intelligence while still adhering to ethical standards. When everything is said and done, this will ultimately lead to the

establishment of a learning environment that is more open, accountable, and welcoming to people of all backgrounds, including classroom teachers and students.

CONCLUSION:

The implementation of artificial intelligence (AI) into educational systems holds a great deal of promise in terms of enhancing the quality of learning experiences and achieving the highest possible educational outcomes. The opaqueness that is inherent in AI algorithms, on the other hand, has given rise to important concerns surrounding the interpretability, trustworthiness, and ethical ramifications that they may have within educational contexts. These issues have been brought about by the fact that AI algorithms are inherently opaque. This research was carried out with the intention of conducting a comprehensive analysis of the significance of XAI in the field of education, as well as investigating the methodology and prospective uses of this technology. This study will investigate the current landscape of artificial intelligence (AI) in education as well as the future prospects of XAI in education in order to throw light on how XAI may improve educational systems, generate enhanced learning outcomes, and successfully integrate AI in educational settings.

REFERENCES:

[1] A. A. Mubarak, H. Cao, & W. Zhang. (2020). Prediction of students' early dropout based on their interaction logs in online learning environment. Interact. Learn. Environments, 1- 20.

[2] A. Kaur, A. Mustafa, L. Mehta, & A. Dhall. (2018). Prediction and localization of student engagement in the wild. Digital Image Computing: Techniques and Applications (DICTA), 1-8.

[3] D. Baneres, M. E. Rodriguez, & M. Serra. (2019). An early feedback prediction system for learners at-risk within a first-year higher education course. IEEE Trans. Learn. Technology, 249–263.

[4] Embarak, O. (2022). An adaptive paradigm for smart education systems in smart cities using the internet of behaviour (IoB) and explainable artificial intelligence (XAI). 8th International Conference on Information Technology Trends (ITT), Dubai, United Arab Emirates (pp. 74-79). Dubai, United Arab Emirates: IEEE.

[5] F. Chen, & Y. Cui. (2020). Utilizing student time series behaviour in learning management systems for early prediction of course performance. J. Learn. Anal, 1-17.

[6] Guleria, P, & Sood, M. . (2023). Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling. Springer Educ Inf Technol, 1081–1116.

[7] H. Wan, K. Liu, Q. Yu, & X. Gao. (2019). Pedagogical intervention practices: Improving learning engagement based on early prediction. IEEE, 278-289.

[8] Islam MR, Ahmed MU, Barua S, & Begum S. (2022). A Systematic Review of Explainable

Artificial Intelligence in Terms of Different Application Domains and Tasks. Applied Sciences, 1353.

[9] Luca Marconi, Ricardo Anibal Matamoros Aragon, Serena Fossati, Italo Zoppis, Rossana Actis Grosso, Sara Manzoni, . . . Francesco Epifania. (2020). Towards Explainable AI for Personalized Teaching:results on experimental activities on the "WhoTeach" educational platform. Informatica per la Didattica DIDAMATiCA 2020.

[10] M. Adnan, M. I. Uddin, E. Khan, F. S. Alharithi, S. Amin, & A. A. Alzahrani. (2022). Earliest Possible Global and Local Interpretation of Students' Performance in Virtual Learning Environment by Leveraging Explainable AI. IEEE Access, 129843-129864.

[11] Melo E, Silva I, Costa DG, Viegas CMD, & Barros TM. (2022). On the Use of eXplainable Artificial Intelligence to Evaluate School Dropout. Education Sciences. Education Sciences, 845.

[12] Yang, C, Huan, S, & Yang, Y. (2020). A practical teaching mode for colleges supported by artificial intelligence. International Journal of Emerging Technologies in Learning (IJET), 195–206.