# Machine Learning and Explainable Artificial Intelligence in Education and Training - Status and Trends

Dimitris Pantazatos[1], Athanasios Trilivas[2], Kalliopi Meli[3], Dimitrios Kotsifakos[2(✉)], and Christos Douligeris[2]

[1] National Technical University of Athens, Athens, Greece
dpantazatos@netmode.ntua.gr
[2] University of Piraeus, Piraeus, Greece
{a.trilivas,kotsifakos,cdoulig}@unipi.gr
[3] University of Patras, Patras, Greece
kmeli@upatras.gr

**Abstract.** Nowadays, the need to explain the decisions or predictions made by Artificial Intelligence (AI) is emerging more than ever as AI applications are more complex. The research field of eXplainable Artificial Intelligence (XAI) tries to fulfill this need. XAI provides a way to help humans understand how an AI's predictions and decisions come. The scope of this work is to examine the role of XAI in the field of Education, especially in Educational Data Mining in Vocational Education and Training.

**Keywords:** Artificial Intelligence · Machine Learning · Explainable AI · Educational Data Mining · VET

## 1 Introduction

In the past few decades, there has been a significant shift in machine learning practitioners' focus towards enhancing their models' predictive capabilities. This shift has gradually replaced simpler, inherently interpretable machine learning models with more complex and performant networks. However, the complexity of these modern "black box" models presents challenges regarding their interpretability, a significant concern for developers, users, and legal entities who are responsible for their deployment in production environments.

For developers, the complexity of these models poses difficulties in understanding, debugging, and asserting the post-deployment of the models' intended functions. On the other hand, users often need help to receive justifications for the decisions made by these models based on their data. Furthermore, legal regulators face hurdles in ensuring that such models, especially when deployed within critical infrastructures, comply with regulatory mandates such as GDPR, which necessitates model transparency and the right to explanation.


L. A. Maglaras and C. Douligeris (Eds.): WiCON 2023, LNICST 527, pp. 110–122, 2024.
https://doi.org/10.1007/978-3-031-58053-6_8

To mitigate these challenges, eXplainable Artificial Intelligence (XAI) has emerged. XAI proposes techniques that provide insights into the machine's inner workings and deep learning models. Among the spectrum of XAI techniques, the most versatile and universally applicable are those categorized as post-hoc and model-agnostic [1]. These algorithms can be applied after the model's training phase (post-hoc) and are not dependent on the specific type of model used (model-agnostic), making them broadly applicable across various models. XAI methods can furnish explanations on two levels: global explanations that shed light on the model's behavior using the training data as a whole and local explanations that detail how models arrive at specific decisions for individual input examples.

This work describes how XAI facilitates a relatively straightforward understanding for education experts of predictive algorithms and decision-making processes employed by educational platforms. Typically, these platforms provide decision-making outputs without accompanying explanations, leaving educators in the dark about the reasoning behind these outputs. XAI can bridge this gap by elucidating the factors and mechanics behind algorithmic decisions, fostering a transparent and trustful adoption of AI in educational settings.

This work is structured as follows: Sect. 2 provides the reader with the basic terminology related to AI and Machine Learning (ML) algorithms. Section 3 describes what XAI means and how it is used in various fields, while Sect. 4 explains how XAI is currently used in education, especially in applications that require data mining. Section 5 presents a use case in Vocational Education and Training (VET) while Sect. 6 concludes the paper and provides ideas for future work.

## 2 Artificial Intelligence and Machine Learning

This section will provide a brief introduction to Artificial Intelligence (AI) and Machine Learning (ML) and their use in educational settings.

### 2.1 Artificial Intelligence

Several definitions have been proposed for what AI means. John McCarthy [2] proposed that AI can be considered as the science and engineering of making intelligent machines and is related to using computers to understand human intelligence. Nevertheless, as John McCarthy points out, we must consider that AI does not have to confine itself to biologically observable methods.

However, this definition is one of many. AI is used in many senses (even outside of Computer Science), so the definition varies occasionally. According to Pei Wang [3], many people do not consider it a big problem, as many scientific concepts get sufficient definitions only after the research has grown. In addition, he also thinks that there is no correct working definition of AI, as each has theoretical and practical values, but some definitions can be considered better than others. In the scope of this work, we will stick to the definition that John McCarthy provides, as it describes in quite a few sentences the meaning of AI in our period.

### 2.2 Machine Learning

ML can be considered a field of AI that uses data from various sources and algorithms (e.g., Linear/Logistic Regression) to imitate how humans learn [4]. This procedure aims to help computers to understand and gradually improve their accuracy. According to Olladipupo [5], ML algorithms are mainly categorized based on the desired outcome of the algorithm. The most known categories of ML algorithms are:

- Supervised learning: The ML algorithm generates a function that maps inputs to desired outputs (labeled examples are available)
- Unsupervised learning: ML is based on a specific set of inputs (not labeled examples).
- Semi-supervised learning: Supervised and Unsupervised techniques are used in this case.
- Reinforcement learning: In this case, the ML algorithm interacts with the environment. As a result, the algorithm uses a policy of how to act given an observation of the world. Every action impacts the environment, and the environment provides feedback that guides the learning algorithm (a case of a grid with a reward system).

Other types of ML techniques are Transduction and Learning to Learn. Transduction is an inference that makes predictions about new data points using specific training examples without abstracting a general rule first [6]. Learning to Learn (or Meta-learning) follows a different approach as a model is trained to learn from other learning tasks [7]. However, these models are outside the scope of this work as they are not widely used in education-related cases.

### 2.3 Machine Learning in Education

ML has been used in education, as in many cases there is a need to thoroughly analyze the student's performance and to predict their academic development. As a result, there is a need to extract valuable information from data related to education.

Wu [8] presents a use case based on a relatively simple dataset containing 1000 and 8 different variables such as gender, race, parental level of education, lunch, test preparation course, math score, reading score, and writing score. The variables related to scores are ordinal, while the others are nominal. The paper examined correlations based on these variables using linear regression (single-wise and stepwise). Another application of this dataset was to make a simple classifier based on K-nearest Neighbors (KNN) and Logistic Regression (LR). Using these simple supervised methods, she proved that personal study habits are more important than family influence.

Another work that shows the potential of ML in educational data mining is that from Ya Zhou et al. [9], which examines the effectiveness of ML in education Big Data. In this work, the authors stress the need for an intelligent education solution. They present four main practical methods in educational data mining, namely Clustering, Prediction, Relationship Mining, and Model Construction. These methods can be implemented using relatively simple and known ML methods like Multivariate Regression, and K-means.

Kishan Das Menon et al. [10], present other methods for educational data mining, such as Naive Bayes (which is a supervised learning algorithm) and IDE3 (an algorithm used for decision tree generation). They used these algorithms alongside Linear/Logistic

Regression and KNN to provide a mechanism for predicting students' performance and to counsel students for college enrolment more effectively. To test their proposed mechanism, they used a dataset that consisted of details of 132 students of all majors in their university. The dataset included the students' marks/grades from Class XI/Pre-University to the marks of the latest available semester.

Hilbert S., Coors S. et al. [11] conclude that education, especially after the COVID-19 pandemic, will eliminate the one-size-fits-all approach as learning is required to suit the needs of the individual to be more effective. As a result, data collected from different resources must be analyzed more efficiently and accurately. Nevertheless, as they also point out, this requires not only ML models suitable for this cause but also researchers in the educational sciences to be aware of these techniques.

## 3 Definition of XAI

The previous section examined how ML can be used in educational data mining. The view from the previous section is that researchers and educational experts have many ways to analyze their data. Nevertheless, how do these experts know that the results are accurate and suitable for their purpose? XAI aims to answer this question by providing the tools and methodology to explain how this output was provided. In this section, essential aspects of XAI will be presented to better understand how XAI can be used in the case of educational data mining.

### 3.1 Why Do We Need XAI?

During the last two decades, ML algorithms have become increasingly popular as they provide a way to create accurate decision support systems. These systems are based on relatively complex techniques such as Deep Neural Networks (DNNs). DNNs' complexity is based on their sizeable parametric space, utilizing many layers and parameters. These aspects of DNNs make them considered black-box models [12].

As these kinds of black-box models are used on most prediction systems, the demand for transparency is constantly increasing among various stakeholders [13].

This demand can also be based on the ethical aspects of the usage of AI. It makes sense that humans tend not to trust solutions they cannot oversee. In addition, if the user understands how a model works and how the output has been created, they can check if the model is accurate. So, the need for implementing ML models that are transparent is more than necessary. According to Alejandro Barredo Arrieta et al. [14], when an ML model is based on transparency and interpretability, it can improve its implementation ability for three reasons:

- Integrity in decision-making.
- Confidence that only useful variables will be used in the model.
- Accuracy of the prediction.

From the points stressed above, we can conclude that XAI should be considered a necessary part of a successful ML application in terms of user experience and trust.

### 3.2 Basic Aspects of XAI

As shown in the previous paragraph, the nature and complexity of most decision systems based on AI created the need to explain these applications' output. According to Philips et al. [15], four main principles are needed for a successful XAI solution. These principles are mentioned below:

- Explanation: Systems deliver supporting evidence or reason(s) for all outputs.
- Meaningful: Systems provide explanations that are understandable to individual users.
- Explanation Accuracy: The explanation correctly reflects the system's output-generating process.
- Knowledge Limits: The system only operates under the conditions it was designed for or when it reaches sufficient confidence in its output.

Regarding the implementation of an XAI solution, there are a lot of possible solutions. Among various XAI techniques, we will mainly focus on Shapley Additive explanations (SHAP), which can be described in more detail in the work of Lundeberg et al. [16].

SHAP is one of the most prominent model-agnostic XAI methods. It originates in game theory and determines feature importance by considering how much classification decisions vary when specific features are used and when they are removed.

The SHAP values deliver insights into a model's decision-making on two levels: globally and locally. Global explanations using SHAP values highlight which features are pivotal across all predictions, offering a broad understanding of the model's behavior. On the other hand, local explanations focus on individual predictions, detailing how specific features affect each decision made by the model [17]. This dual capacity of SHAP values, complemented by intuitive visualizations, provides developers with a clear window into their model's interpretative framework. Model-agnostic SHAP methods leverage a Kernel Explainer module, which approximates feature importance via a weighted linear regression trained on an input sample. Apart from Kernel Explainer, other explainers are also available, but they only apply to specific models.

The SHAP solutions are the standard for implementing XAI on ML algorithms, as they provide an extensive Python library. Apart from SHAP, according to Angelov et al. [18], other methods are Class Activation Maps (CAMs), Global Attribution Mappings (GAMs), Concept Activation Vectors (CAVs), Local interpretable model-agnostic explanations (LIME), Layer-wise relevance propagation (LRP) and others. The central aspect of these solutions is that they are relatively new and are under active development, as all solutions are related to XAI.

### 3.3 XAI Use Cases

As it has emerged from the previous sections, XAI can be implemented in various ways and contexts. As a result, there are a lot of applications of XAI that are worth mentioning. Some of the fields that XAI currently implements are cases related to Justice, Natural Language Processing (NLP), Financing, Anomaly detection, and others. Other subjects and impact areas that are mentioned by L. Longo et al. [19] are:

- Threat Detection and Triage.
- Explainable Object Detection.

- Protection Against Adversarial ML.
- Open-Source Intelligence (OSINT).
- Trustworthy (autonomous) Medical Agents.
- Autonomous Vehicles.

As mentioned in the same work, several challenges (technical, legal, and practical) are almost identical for all cases. The following section will present some topics related to XAI in education.

## 4 XAI in Education

In the previous section, the fundamental aspects of XAI were presented concisely. This section will examine how XAI is implemented in educational data mining. Even though XAI is implemented in the same way in ML applications in education, there are also unique needs emerging from a pedagogical point of view.

H. Khoravi et al. [20] introduced the ED-XAI framework to present this distinction. This framework consists of six dimensions. These six dimensions can be considered questions that must be answered to build a successful XAI solution. These questions are:

- Who are the main stakeholders (e.g., teachers)?
- What are the main benefits?
- What potential pitfalls need to be considered?
- What approaches are used for presenting explanations?
- What are AI models commonly used?
- How can educational AI tools be effectively designed?

These questions have more than one answer. The answers depend on the context in which this application will be used. Four different applications are presented based on ED-XAI are also presented in [20]. In both cases, trust can be considered the common XAI benefit, while the main stakeholders are educators and students. Regarding the models that were used, they were mostly related to NLP and Classification. Finally, incomplete explanations and inaccurate models were the main drawbacks regarding the pitfalls of needlessly complex models.

On the work of Clancey and Hoffman [21], the XAI systems have much in common with Intelligent Tutoring Systems (ITS). ITS was first created during the 1960s and 1980s. Although XAI and ITS have some shared values and issues (especially trust should be considered a common issue in both technologies), ITS is a more generic way of implementing solutions that would teach learners to solve problems by themselves.

Conati et al. [22] also examine the role of ITS in the development of XAI. However, they seem to compare XAI more to Open Learner Models (OLMs), which, according to Bull and Pain [23], are models that allow learners to access their material with different levels of interactivity. OLMs can be scrutable, cooperative, negotiable, or editable.

Nowadays, there are ITS systems that provide OLM functionalities. OLMs could be a form of XAI for an ITS but suited to user needs. One interesting work regarding adaptive to user needs XAI can be considered the work by Embark [24]. In this work, an intelligent education system was examined. The scope of this educational system was

to provide more personalized learning material to the learners based on data collected by Internet of Things (IoT) and Internet of Bodies (IoB) devices. In this case, XAI was used to filter/select features that students can choose to take a complete view of their performance and to seek help when needed.

Finally, XAI can use a few resources or is intended for experienced users. According to Alonso [24], even kids under 12 could use an XAI application implemented in Scratch, a visual programming language commonly used in primary education that helps students get familiar with programming. In this case, they implemented XAI in NLP applications. The following section will examine how XAI can be integrated into a Vocational Education and Training (VET) case related to the Network Management course.

## 5 Enhancing Network Management Training with Explainable AI

XAI is more significant in VET than in general education due to its tailored and practical application. In VET, students are focused on acquiring specific skills and knowledge directly relevant to their chosen professions. XAI's ability to provide transparent, comprehensible insights into AI decision-making aligns seamlessly with this objective. By understanding the 'why' behind AI's recommendations and errors, VET students can hone their skills more effectively, address their weaknesses, and develop a deeper understanding of the subject matter. Moreover, in vocational fields where safety, ethics, and real-world applicability are paramount, XAI ensures students are well-prepared to navigate AI-driven workplaces, make informed decisions, and uphold ethical standards. XAI's capacity to enhance learning outcomes, facilitate skill development, and prepare students for practical, career-focused challenges makes it uniquely valuable in Vocational Education and Training. In the following sections, we will examine a scenario of xAI in a VET course.

### 5.1 The Network Management Course Use Case

The following use case exemplifies how xAI techniques can be harnessed to enhance the learning experience, improve transparency, and empower instructors and students in critical fields like IT infrastructure management. By examining the practical application of xAI within this context, we can appreciate its transformative potential in Vocational Education and Training (VET).

The use of AI applications in Learning Management Systems (LMS) like Moodle is something that has been introduced previously. Manhica et al. highlight the use of AI for student performance assessment and the most used AI algorithms in LMS like Moodle [25].

The trainer in the proposed case can use the Moodle prediction plugin for this course [26]. This plugin harnesses the power of machine learning algorithms to predict student performance based on an array of data inputs, including quiz scores, forum participation, assignment submissions, and more. While the plugin offers valuable insights, it tends to need more transparency, leaving instructors and students needing clarification about the factors driving its predictions. One relevant work based on this case is from Ogata et al.,

they propose a system where both students and the AI system explain their decision-making processes, enhancing metacognitive skills and providing insights into learners' challenges [27]. The proposed scenario can also utilize a mechanism based on the same principles.

To bridge this transparency gap and elevate the overall learning experience, the adoption of explainable AI (xAI) techniques becomes imperative. Several xAI techniques prove to be highly effective in this context:

- Feature Importance Analysis: A xAI technique that analyses feature importance within the predictive model. In this context, it can identify which factors exert the most substantial influence on student performance predictions. This revelation empowers instructors with a clear understanding of the elements driving the projections.
- Decision Tree Visualization: Another approach related to xAI entails the creation of decision trees that visually represent the logic behind the prediction model's conclusions. These decision trees can serve as enlightening tools for instructors and students, shedding light on the intricate factors contributing to performance assessments.
- Local Explanations: xAI techniques extend their power by providing localized explanations. These explanations are tailored to individual students, offering insights specific to their strengths and weaknesses. For instance, if a student grapples with subnetting concepts, the xAI system can pinpoint this issue and deliver customized resources or guidance focused solely on subnetting.
- Model Agnostic Techniques: Model agnostic techniques such as Local Interpretable Model-agnostic Explanations (LIME) or Shapley Additive explanations (SHAP) enhance model interpretability. These techniques work seamlessly with a broad spectrum of predictive models and provide intricate insights into their predictions, rendering them more transparent and understandable.

Regarding the last two methods, one interesting work is by Adnan [28]. In this work, the author proposes an xAI model that provides an early global and local interpretation of students' performance at various course stages. The model offers interpretations at 20%, 40%, 60%, 80%, and 100% of course length, providing insights into student performance in a human-understandable way and aiding instructors in offering timely personalized feedback and guidance. Various traditional and ensemble ML algorithms (e.g., Logistic Regression) were trained on demographic, clickstream, and assessment features to determine the best-performing algorithm, which was then provided to the xAI model to interpret students' study behavior at various percentages of course length. Another work worth mentioning is the work of Shamy et al., where they explained black-box machine learning models in the context of student performance prediction in MOOCs [29]. They discovered that all the evaluated methods could (partially) detect the prerequisite relationship between weeks while relying on only behavioral features.

The advantages of incorporating xAI in the VET setting are far-reaching:

- Transparency and Accountability: xAI techniques introduce transparency into the prediction process. Instructors gain a precise understanding of why particular predictions are made, establishing accountability within the educational ecosystem [30]. This transparency empowers instructors to make informed decisions regarding interventions and support for students who may be struggling.

- Customized Support: Instructors can deliver targeted support by providing individualized student explanations. When students comprehend their shortcomings, they can receive more effective assistance. For example, xAI analysis can reveal a student grappling with specific networking protocols. In that case, instructors can recommend additional study materials or one-on-one tutoring tailored to that concern.
- Course Design Enhancement: xAI insights are pivotal in shaping course design adjustments. If certain course materials or assessment methodologies consistently yield poor performance predictions, instructors can recalibrate them to better align with student needs. This iterative approach to course improvement invariably enhances the overall learning experience. According to V. Shamy, university-level educators are interested in having concrete explanations as they could transform these insights into actionable course design decisions. Still, more has to be done to make explanations more user-friendly [31].
- Enhanced Student Engagement: Students are more likely to become deeply engaged in their studies When they understand the direct correlation between their actions and performance predictions. xAI provides students with a clear roadmap to success, empowering them to take ownership of their education and strive for continuous improvement.
- Data-Driven Decision-Making: Instructors and educational institutions can make data-driven decisions regarding resource allocation, interventions, and curriculum development. This data-driven approach leads to more efficient utilization of resources and a higher quality of education, ultimately benefiting both students and educators.

The proposed mechanism can be implemented based on these aspects:

- Integration of XAI Techniques: XAI techniques like SHAP (Shapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) can be utilized to interpret the model's predictions. For global explanations, SHAP values can be aggregated across all predictions to understand overall feature importance. For local explanations, generated SHAP values for individual predictions can be used.
- Development of an Interface: An interface within Moodle that presents the predictions and their explanations in a user-friendly manner has to be developed. The explanations must be clear and understandable for non-technical users, potentially using visualizations like bar charts or decision plots.
- Ethical and Privacy Considerations: The ethical implications of predicting student performance, including the potential impact on students and the risk of bias in the model, must be considered. In addition, data privacy is maintained in compliance with regulations such as GDPR, and students have consented to use their data for these purposes.
- Testing and Iteration: Test the integrated system within a live Moodle environment with a pilot group before rolling it out widely. User feedback and iteration on the model and interface to improve accuracy and usability are also necessary.
- Documentation and Training: The system has to provide documentation and training for educators on how to interpret and act on the predictions and explanations.

Incorporating xAI techniques into VET, especially in courses like network management, can significantly enhance the educational experience. XAI promotes transparency,

personalization, and data-driven decision-making, ultimately better-preparing students for successful careers in IT infrastructure management. The following section will offer some extensions of the presented proposal.

### 5.2 Proposed Extensions

Some extensions that would be considered as added value for the proposed scenario can be these:

- Sequential Pattern Explanation Framework: An XAI algorithm tailored to decipher the sequence of student interactions could be constructive in network management courses. This framework could elucidate how the chronological progression through course materials influences student performance, providing insights into optimal learning paths. For instance, it could reveal if studying specific network protocols before practicing with management tools could lead to a deeper understanding and better practical exam scores.
- Demographic Impact Analyzer: An analyzer that leverages XAI to assess the influence of demographic factors on course outcomes would enable educators to tailor the network management curriculum to accommodate diverse educational backgrounds, learning paces, and prior technical experience, fostering a more personalized educational approach.
- Bias Detection and Explanation Engine: An XAI algorithm designed to identify and explain biases within predictive models would ensure equitable education, highlighting if specific model predictions inadvertently favor or disadvantage student segments and clarifying the underlying reason. For example, the work of Arias-Duar et al. seems promising, especially for examining biases in predictive and classification problems, even with noisy data [32].
- Predictive Model Feedback Loop: A predictive feedback loop would enable a dynamic interaction where educators can input their expertise into the XAI system. Disagreements with the model's explanations can be addressed, and this feedback will be instrumental in refining the model, ensuring that the course evolves in line with educator insights and industry relevance.
- Curriculum Alignment Insights: Lastly, an XAI application that could align the course's learning analytics with established network management competencies and industry standards could be highly innovative. This alignment would ensure that the course content remains current and relevant and that educators understand how AI-derived insights correspond with educational benchmarks.

By incorporating these proposed extensions, the corresponding course presented in the previous section could significantly benefit from enhanced interpretability and relevance, empowering educators and students with actionable insights and fostering a deeper understanding of complex network systems.

## 6 Conclusion

In this paper, we argued that integrating XAI within educational applications, particularly for VET, is pivotal in fostering trust and transparency between the software and its users. The successful application of XAI goes beyond the complexity of underlying machine learning models and educational contexts. The simplicity of implementing XAI, when underpinned by robust architecture and design principles oriented towards explainability, is instrumental in realizing its benefits. The adaptability of the proposed XAI approach in the VET use case demonstrates its potential for broader educational applications. The exploration of diverse educational scenarios will further enrich this dynamic field, contributing to the refinement and enhancement of XAI strategies in education. The ongoing studies and experiments in this field are expected to lead to significant breakthroughs, making sure that the tools used in educational technology advance in step with the rapid developments in AI and machine learning.

**Acknowledgment.** This work has been co-financed by Greece and the European Union (European Regional Development Fund-ERDF) through the Regional Operational Program "Attiki" 2014–2020.

## References

1. Ali, S., et al.: Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. Inf. Fusion **99** (2023). https://doi.org/10.1016/j.inffus.2023.101805
2. Mccarthy, J.: What is artificial intelligence? (2007). https://dl.acm.org/doi/pdf/10.1145/1283920.1283926
3. Wang, P.: On defining artificial intelligence. J. Artif. Gener. Intell. **10**, 1–37 (2019). https://doi.org/10.2478/jagi-2019-0002
4. IBM: What is machine learning? https://www.ibm.com/topics/machine-learning. Accessed 19 Nov 2023
5. Ayodele, T.O.: X types of machine learning algorithms. New Adv. Mach. Learn. **3**, 19–48 (2010)
6. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (2000). https://doi.org/10.1007/978-1-4757-3264-1
7. Baxter, J.: Theoretical models of learning to learn. In: Thrun, S., Pratt, L. (eds.) Learning to Learn, pp. 71–94. Springer, Boston (2020). https://doi.org/10.1007/978-1-4615-5529-2_4
8. Wu, J.: Machine learning in education. In: Proceedings - 2020 International Conference on Modern Education and Information Management, ICMEIM 2020, pp. 56–63. Institute of Electrical and Electronics Engineers Inc. (2020). https://doi.org/10.1109/ICMEIM51375.2020.00020
9. Zhou, Y., Song, Z.: Effectiveness analysis of machine learning in education big data. In: Journal of Physics: Conference Series. IOP Publishing Ltd. (2020). https://doi.org/10.1088/1742-6596/1651/1/012105
10. Kishan Das Menon, H., Janardhan, V.: Machine learning approaches in education. In: Materials Today: Proceedings, pp. 3470–3480. Elsevier Ltd. (2020). https://doi.org/10.1016/j.matpr.2020.09.566

11. Hilbert, S., et al.: Machine learning for the educational sciences (2021). https://doi.org/10.1002/rev3.3310
12. Castelvecchi, D.: Can we open the black box of AI? Nature **538**, 20–23 (2016). https://doi.org/10.1038/538020a
13. Preece, A., Harborne, D., Braines, D., Tomsett, R., Chakraborty, S.: Stakeholders in explainable AI (2018)
14. Barredo Arrieta, A., et al.: Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities, and challenges toward responsible AI. Inf. Fusion **58**, 82–115 (2020). https://doi.org/10.1016/j.inffus.2019.12.012
15. Jonathon P., Hahn, C.A., Fontana, P.C., Broniatowski, D.A.: Draft NISTIR 8312 - four principles of explainable artificial intelligence (2020). https://doi.org/10.6028/NIST.IR.8312-draft
16. Lundberg, S.M., Allen, P.G., Lee, S.-I.: A unified approach to interpreting model predictions (2017)
17. Molnar, C.: Interpretable machine learning a guide for making black box models explainable (2019)
18. Angelov, P.P., Soares, E.A., Jiang, R., Arnold, N.I., Atkinson, P.M.: Explainable artificial intelligence: an analytical review. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **11** (2021). https://doi.org/10.1002/widm.1424
19. Longo, L., Goebel, R., Lecue, F., Kieseberg, P., Holzinger, A.: Explainable artificial intelligence: concepts, applications, research challenges and visions. In: Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-MAKE 2020. LNCS, vol. 12279, pp. 1–16. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-57321-8_1
20. Khosravi, H., et al.: Explainable artificial intelligence in education. Comput. Educ.: Artif. Intell. **3** (2022). https://doi.org/10.1016/j.caeai.2022.100074
21. Clancey, W.J., Hoffman, R.R.: Methods and standards for research on explainable artificial intelligence: lessons from intelligent tutoring systems. Appl. AI Lett. **2** (2021). https://doi.org/10.1002/ail2.53
22. Conati, C., Porayska-Pomsta, K., Mavrikis, M.: AI in Education needs interpretable machine learning: lessons from open learner modelling (2018)
23. Bull, S., Pain, H.: "Did I Say What I Think I Said, and Do You Agree with Me?": Inspecting and Questioning the Student Model. AACE (1995)
24. Alonso, J.M.: Explainable Artificial Intelligence for Kids (2019)
25. Manhica, R., Santos, A., Cravino, J.: The use of artificial intelligence in learning management systems in the context of higher education: systematic literature review. In: 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–6. IEEE (2022). https://doi.org/10.23919/CISTI54924.2022.9820205
26. Zhang, Y., Ghandour, A., Shestak, V.: Using learning analytics to predict students performance in moodle LMS. Int. J. Emerg. Technol. Learn. **15**, 102–114 (2020). https://doi.org/10.3991/ijet.v15i20.15915
27. Ogata, H., Flanagan, B., Takami, K., Dai, Y., Nakamoto, R., Takii, K.: EXAIT: educational eXplainable artificial intelligent tools for personalized learning (2024)
28. Adnan, M., Uddin, M.I., Khan, E., Alharithi, F.S., Amin, S., Alzahrani, A.A.: Earliest possible global and local interpretation of students' performance in virtual learning environment by leveraging explainable AI. IEEE Access. **10**, 129843–129864 (2022). https://doi.org/10.1109/ACCESS.2022.3227072
29. Swamy, V., Radmehr, B., Krco, N., Marras, M., Käser, T.: Evaluating the explainers: black-box explainable machine learning for student success prediction in MOOCs (2022)
30. Holmes, W., et al.: Ethics of AI in education: towards a community-wide framework. Int. J. Artif. Intell. Educ. **32**, 504–526 (2022). https://doi.org/10.1007/s40593-021-00239-1

31. Swamy, V., Du, S., Marras, M., Kaser, T.: Trusting the explainers: teacher validation of explainable artificial intelligence for course design. In: ACM International Conference Proceeding Series, pp. 345–356. Association for Computing Machinery (2023). https://doi.org/10.1145/3576050.3576147
32. Arias-Duart, A., Pares, F., Garcia-Gasulla, D., Gimenez-Abalos, V.: Focus! Rating XAI methods and finding biases. In: 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–8. IEEE (2022). https://doi.org/10.1109/FUZZ-IEEE55066.2022.9882821