

Review

# Explainable Artificial Intelligence Approaches in Primary Education: A Review

Jim Prentzas \* and Ariadni Binopoulou

Department of Education Sciences in Early Childhood Education, Democritus University of Thrace, Nea Chili, 68131 Alexandroupolis, Greece; abinopou@psed.duth.gr

\* Correspondence: dprentza@psed.duth.gr

**Abstract:** Artificial intelligence (AI) methods have been integrated in education during the last few decades. Interest in this integration has increased in recent years due to the popularity of AI. The use of explainable AI in educational settings is becoming a research trend. Explainable AI provides insight into the decisions made by AI, increases trust in AI, and enhances the effectiveness of the AI-supported processes. In this context, there is an increasing interest in the integration of AI, and specifically explainable AI, in the education of young children. This paper reviews research regarding explainable AI approaches in primary education in the context of teaching and learning. An exhaustive search using Google Scholar and Scopus was carried out to retrieve relevant work. After the application of exclusion criteria, twenty-three papers were included in the final list of reviewed papers. A categorization scheme for explainable AI approaches in primary education is outlined here. The main trends, tools, and findings in the reviewed papers are analyzed. To the best of the authors' knowledge, there is no other published review on this topic.

**Keywords:** explainable artificial intelligence; artificial intelligence in education; AI literacy; educational technology; E-learning; primary education; primary school; elementary education; elementary school



Academic Editors: Valentina E. Balas and Mohit Mittal

Received: 2 May 2025

Revised: 28 May 2025

Accepted: 1 June 2025

Published: 3 June 2025

**Citation:** Prentzas, J.; Binopoulou, A. Explainable Artificial Intelligence Approaches in Primary Education: A Review. *Electronics* **2025**, *14*, 2279. <https://doi.org/10.3390/electronics14112279>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

During the last few decades, there has been an effort to integrate artificial intelligence (AI) in education [1–3]. The integration of AI in education involves different aspects. One aspect is the teaching of AI to students. Another aspect concerns the use of AI to assist in teaching and learning tasks. An indicative example of this aspect is the incorporation of AI in interactive learning systems to provide real-time personalized interaction to students and teachers. An alternative example is the use of AI in processing data collected from specific learning settings. A third aspect is the use of AI to assist in policymaking and decision support. An example of using AI in policymaking is the processing of large amounts of data collected from national or international educational assessments and surveys in order to obtain information that could be reflected in future policies [4–6]. Examples of using AI in decision support include the identification of geographical regions requiring additional schools or teachers [7] and the determination of optimal school bus routes and stops [8].

One may note that there is sufficient experience in the integration of AI in higher education. First, the teaching of AI in higher education has been explored extensively during the last few decades in higher education, because AI is a learning subject in undergraduate and postgraduate levels. Consequently, there is an abundance of teaching experience,

learning content, and learning tools concerning AI. Second, many higher education institutions have explored the use of AI to assist in teaching and learning. On the one hand, this has been done with customized learning tools developed within higher education institutions [9,10] and with other available AI-based tools [11]. On the other hand, research is also conducted about the use of AI in processing data collected from students (e.g., data stored in e-learning platforms and student projects). Moreover, the use of AI in decision support within higher education has been explored (e.g., identification of students facing difficulties in their studies [12]).

There are studies about the integration of AI in the younger age groups as well, in addition to higher education. A recent development is an increasing interest in implementing relevant approaches in primary education, as demonstrated by the publications in this subject during the last few years. According to the International Standard Classification of Education by the United Nations Educational, Scientific, and Cultural Organization (UNESCO), primary education is an educational level following early childhood education and preceding secondary education [13]. The integration of AI in primary education is an emerging research direction for three main reasons: First, technology-enhanced learning is becoming increasingly popular in educational settings concerning the younger age groups. Teachers, children, and parents are now more experienced with technological resources compared to previous time periods, and this assists in the integration of technology in primary education. Educational applications and digital content are available for primary education, facilitating the implementation of technology-enhanced activities. Second, AI is becoming a hot topic at all levels of education, including primary education. The advances in AI during the last decade and the recent popularity of AI applications have provided an impetus for this. Third, the integration of AI in primary education has not been explored to the extent that it has in higher education. This leaves room for research work.

A prerequisite in many fields is the provision of explanations for the decisions made by AI-based systems. Education is not an exception. In certain cases (e.g., various types of rule-based systems and other comprehensible models) there are inherent mechanisms providing explanations for the produced outputs [14,15], or straightforward ways of doing this. However, several AI methods do not have inherent mechanisms for (or there are no straightforward ways of) explaining their outputs, and the knowledge they encompass is also not comprehensible. This results in difficulties. Explainable AI (XAI) is an AI field aiming to deal with such issues. Generally speaking, XAI aims to offer comprehensive explanations of the procedures and functions of AI applications, making AI more transparent and understandable to a wider audience [16,17]. Interest in XAI is gradually increasing because it provides the means to enhance the accountability and fairness of AI systems, resulting in increased trust in AI decisions [18]. For these specific reasons, XAI is also gradually being integrated in educational settings [19].

This paper reviews work concerning the integration of XAI in primary education. The main reasons for such a review are the increasing interest in integrating AI in primary education and the increasing interest in XAI in all sectors, including education. A review of the integration of XAI in primary education will attract the interest of researchers and teachers. It will also provide insight to researchers and teachers about the main trends of approaches integrating XAI in primary education and how XAI may generally be useful in primary education. To the best of the authors' knowledge, there is no other published review on this subject.

This paper is organized as follows: Section 2 presents the methodology, including the research questions and the search procedure. Section 3 outlines the main aspects of the reviewed papers. Section 4 presents the results, discussing the main trends, and briefly presenting reviewed papers according to the category to which they belong. Section 5

discusses aspects of the methodological quality of the reviewed papers. Section 6 presents the discussion. Finally, Section 7 presents the conclusions and the future research directions.

## 2. Materials and Methods

The preparation of this review was guided by explicit research questions. More specifically, these research questions were as follows:

- (1) In which main categories may the XAI approaches in primary education be discerned?
- (2) What are the main trends of the XAI approaches in primary education?
- (3) What are the main XAI tools or methods used in the XAI approaches in primary education?
- (4) In which primary education learning subjects are the XAI approaches used?
- (5) For which AI methods are XAI tools or methods used to provide explanations in XAI approaches in primary education?
- (6) Are single XAI tools and methods or a combination of them more preferred in XAI approaches in primary education?
- (7) Taking into consideration the XAI tools or methods most used in XAI approaches in primary education, which of the main functionalities offered are exploited?

An exhaustive search was carried out to retrieve relevant research using two search tools: Google Scholar and Scopus. The search was carried out from January to February 2025. These two search tools were selected because it is very likely that all relevant research can be retrieved from them. The default settings of both search tools were used. In both tools, the search was carried out in the full text of the documents and was not limited to titles, abstracts, and keywords. No limitation on the publication years was defined. All items retrieved by the two search tools were examined.

The search involved research about XAI and primary education. For the first term (i.e., explainable artificial intelligence), two alternative keywords were used in the search: (i) “explainable artificial intelligence” and (ii) “explainable ai”. For the second term (i.e., primary education), four alternative keywords were used in the search: (i) “elementary school”, (ii) “elementary education”, (iii) “primary school”, and (iv) “primary education”. This was done because, internationally, the alternative terms “primary education” and “elementary education” are used for the same educational level. Similarly, the alternative terms “primary school” and “elementary school” are used for schools at this educational level. Based on the aforementioned, the following logical expression was given as an input to both search tools:

(“explainable artificial intelligence” OR “explainable ai”) AND (“elementary school” OR “elementary education” OR “primary school” OR “primary education”).

In total, 1372 items were retrieved by using Google Scholar and Scopus. Duplicate retrieved items were excluded. The set of remaining items was examined to derive the work that would be reviewed. Retrieved items were excluded from the list of reviewed items according to specific criteria. More specifically, retrieved items were excluded in the following cases:

- (a) They involved papers published in journals that were retracted;
- (b) They involved work that was authored in a language other than English;
- (c) They were not accessible in general or through our institution;
- (d) They involved reviews, position papers, overviews, or editorials, and not research studies;
- (e) They did not involve education but some other field;
- (f) They did not involve primary education;
- (g) They did not involve the use of XAI;
- (h) They involved theses or technical reports.

A total of 140 duplicate items were found and were omitted. Further on, items meeting criteria (a)–(c) were excluded. The numbers of items that were excluded due to criteria (a), (b), and (c) were 1, 47, and 69, respectively. For several of the remaining items, a decision could be made to exclude them from the list of reviewed items (due to criteria (d)–(h)) by reading their title and/or abstract. In total, 901 were thus excluded. Another 214 items remained, whose full contents had to be checked; 191 of these remaining items were excluded due to criteria (e)–(g). In the end, 23 items remained and were used for the review. Table 1 depicts the relevant data. The table provides numerical data about the items excluded due to criteria (d)–(h). As already mentioned, the two search tools were used to search the full text of the papers for the combination of terms. The tools' scope of search was not restricted to the titles, abstracts, and keywords of the items. This enabled the retrieval of the twenty-three reviewed papers because in several of them the combination of terms is not included in the title, abstract and keywords. However, this also meant the initial retrieval of many items that were not relevant and in which the given terms were merely mentioned somewhere in the text.

**Table 1.** Data involving the number of items retrieved with the search tools.

Type of Items	Number
Number of total items initially retrieved with both search tools (without omitting duplicate items)	1372
Number of duplicate items	140
Number of papers published in journals that were retracted	1
Number of items not authored in English	47
Number of items that were not accessible	69
Number of items that were excluded by reading their title and/or abstract	
Number of items that did not involve education but some other field: 397	
Number of items that did not involve primary education: 194	901
Number of items that did not involve the use of XAI: 39	
Number of items that involved theses or technical reports: 40	
Number of items that involved surveys, overviews, or reviews: 231	
Number of items that were excluded by reading their full content	
Number of items that did not involve education but some other field: 56	191
Number of items that did not involve primary education: 83	
Number of items that did not involve the use of XAI: 52	
Number of items that were included in the list of reviewed items	23

### 3. Main Aspects of the Reviewed Papers

Before presenting the list of reviewed items, a categorization scheme for XAI approaches in primary education will be outlined here. The categorization scheme was derived based on the reviewed work and provides an answer to Research Question 1. More specifically, approaches using XAI in primary education may be divided into three main categories: (a) approaches that use XAI to assist in learning and teaching, (b) approaches that use XAI in the context of AI as a learning subject, and (c) approaches that use XAI in policymaking, decision support, and/or administrative tasks in an educational context. Henceforth, the three aforementioned categories will be referred to as categories A, B, and C, respectively. Note that this categorization scheme was implied in the Introduction.

Generally speaking, approaches in category A may be further divided into (i) approaches using XAI during learning and teaching and (ii) approaches using XAI for analysis and support tasks for learning and teaching. The former may involve XAI in face-to-face, blended, or distance learning activities. The latter usually involve the processing of data collected in the learning environment(s). This is mainly done after the learning procedure, but in some cases it may be done prior to the learning procedure (e.g., processing

of pre-tests and questionnaires to obtain information about the status of learners before the learning procedure). Furthermore, XAI may be used in tasks requiring the processing of available data from other sources (e.g., datasets) to acquire information that is useful for teaching and learning. Moreover, XAI may be used to create new content and manage existing content. Approaches in category C may involve the handling of data derived from large-scale research studies at a regional, national, or international level. In addition, educational data may be processed in combination with other types of data. In this category, administrative tasks for teachers, educational personnel, executives, and policymakers are also included. Results derived from studies in category C may indirectly affect teaching and learning.

Figure 1 shows the categorization scheme for XAI approaches in primary education. Table 2 shows indicative tasks concerning XAI in each category. Tasks that have not been presented in the reviewed works (and, thus, constitute unexplored directions) are shown in italics.

Table 3 shows the list of the reviewed items. The reviewed items are shown in alphabetical order, based on the surname of the first author. For each item, the following are shown: the citation and the publication type, the category to which it belongs, the country (or countries) of the author(s), the AI method(s) explained, and the XAI tools or methods applied. In case of a study with multiple authors whose institutions are from different countries, all of these countries are shown in the corresponding cell. This was done for two studies [20,21]. The initials “UAE” stand for the United Arab Emirates. The term “Pub. Type” stands for “Publication Type”. The initials “CP” stand for “Conference Proceedings”, and the initials “JP” stand for “Journal Publication”. Table 3 provides answers to Research Questions 3 and 5.

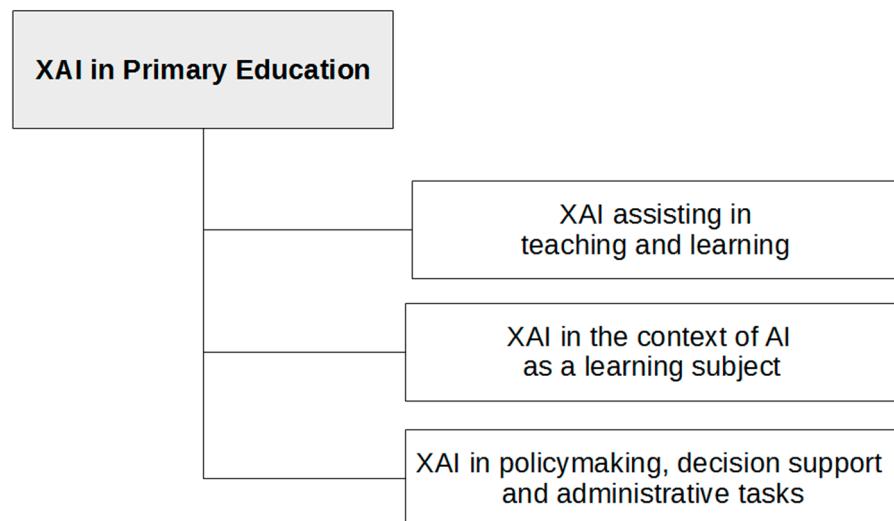
**Table 2.** Indicative tasks concerning XAI in each category of approaches.

Category	Indicative Tasks Concerning XAI
A	Educational applications, presentations, and demonstrations using XAI <i>Lab activities using XAI</i>
	E-learning tool mechanisms (e.g., selection of learning activities, assessment, feedback to teachers, feedback to students, <i>collaboration</i> ) using XAI
	Classroom co-orchestration using XAI
	Processing of data collected in the specific learning environment(s) using XAI
	Processing of data previously accumulated from other studies using XAI
B	Preparation of new teaching content and management of existing teaching content using XAI <i>Lesson planning using XAI</i>
	Educational applications, presentations, and demonstrations about XAI
	<i>Lab activities about XAI</i>
	E-learning tools about XAI <i>Lesson plans about XAI</i> <i>Design of AI curriculum including XAI</i>
C	Analysis of educational data, perhaps in combination with other types of data (family-related, demographics, financial, government data, etc.) using XAI
	Analysis of large-scale educational assessments using XAI
	Assessment of educational unit(s) using XAI
	Administrative tasks of teachers, educational personnel, executives, and policymakers using XAI

**Table 3.** The list of reviewed items.

ID	Citation, Pub. Type	Category	Country (Countries)	AI Method(s) Explained	XAI Tool(s) or Method(s)
1	[22], CP	B	Spain	Voting scheme of J48, RepTree, RandomTree, and FURIA	ExpliClas
2	[23], JP	A	Germany	LightGBM	SHAP
3	[24], JP	C	UK	A black-box AI model	Logical rules
4	[25], JP	C	India	Random Forest	LIME, SHAP, FAMeX,
5	[26], CP	A	South Korea	XGBoost	SHAP
6	[27], CP	A	Uganda	Transformer models	SHAP, BertViz
7	[7], JP	C	South Korea	XGBoost	Feature importance, partial dependence plots, SHAP
8	[28], CP	A	UK	XGBoost	SHAP
9	[21], CP	B	Sweden, Spain	CNN with LSTM	Grad-CAM
10	[29], JP	C	UAE	XGBoost	SHAP
11	[30], CP	A	India	XGBoost	SHAP
12	[5], JP	C	UAE	CatBoost	SHAP
13	[20], JP	A	Switzerland, USA, Germany	Not specifically mentioned	Not specifically mentioned
14	[31], JP	A	Japan	Transformer model	Visualization of attention weights
15	[32], CP	A	Japan	LLM and non-LLM transformer models	Visualization methods
16	[33], JP	A	Czech Republic	Isolation Forest	Explainable outlier detection
17	[34], CP	A	Germany	Transformer models	SHAP
18	[35], JP	A	USA	Reinforcement learning	Integrated gradient analysis
19	[6], JP	C	Japan	KNN, SVM, Random Forest	SHAP
20	[4], CP	C	Brazil	Random Forest	Feature importance
21	[36], CP	A	Netherlands	Open Learner Model, probability models	Text-based explanations
22	[37], JP	A	Netherlands	Open Learner Model, probability models	Text-based explanations
23	[38], JP	A	China	CNN	SHAP

The approaches were published in the time period 2019–2025. This means that they are recent approaches. This was expected, because in the last few years there has been an increasing interest in integrating AI in primary education. A consequence of this interest is the publication of approaches that specifically deal with XAI in primary education. The majority of the studies (thirteen) were published in journals. The other ten studies were published in conference proceedings.



**Figure 1.** A categorization scheme for the XAI approaches in primary education.

As far as the categorization of the studies is concerned, the number of studies in categories A, B, and C is fourteen, two, and seven, respectively. Therefore, most approaches belong to category A. Fewer studies belong to category B compared to the other two categories.

One may note that almost all continents are involved in these studies. This is a positive aspect. Most of the studies are from Asia and Europe (ten studies in each continent). Three studies are from the Americas, and one is from Africa. The continent with the largest number of countries involved in the studies is Europe (seven countries), followed by Asia (five countries). Fifteen countries are involved in the studies. The countries with the most studies (i.e., three studies) are Germany and Japan. Seven other countries (i.e., India, the Netherlands, South Korea, Spain, the UAE, the UK and the USA) are involved in two studies, and the remaining countries are involved in a single study.

According to the data shown in Table 3, various XAI tools or methods are used in these studies. Most of the studies use a single XAI tool or method. A few studies use multiple XAI tools or methods. SHAP is the tool that was used in the most studies. In [20], a specific XAI tool or method is not mentioned. Provision of feature-based explanations is the function that XAI performs in most of the studies. In the following sections, the specific functions that XAI implements in the studies are further analyzed. Below, the functionality of the XAI tools or methods shown in Table 3 will be outlined.

SHAP (SHapley Additive exPlanations) is a tool that provides local and global explanations for any AI model. It is based on the SHAP values of features. SHAP values are calculated for each feature and show the positive or negative contribution of the feature in producing each output. SHAP conveniently provides visual explanations in the form of plots. The role of certain types of SHAP plots used in the reviewed papers will be briefly outlined. The global bar plot explains global feature importance by ordering the input features according to their mean absolute SHAP values and using a separate bar for each feature. Note that the bar plot may provide local explanations as well, by passing a row of SHAP values as a parameter to the plot function. In this case, the specific feature value is shown, along with the specific SHAP value. The beeswarm summary plot combines global and local explanations by ordering features according to their contribution to the output and showing the distribution of SHAP values and a coloring of the feature values for each feature across all dataset instances. For each feature, a dataset instance corresponds to a dot in the plot, with blue or red color to denote a low or high feature value, respectively. The individual force plot provides local explanations by showing the strength of the con-

tribution of each feature to the specific output by ordering, coloring, and sizing features accordingly in a single bar line. The collective force plot provides global explanations by using the absolute mean SHAP values. The waterfall plot provides local explanations by showing the positive or negative contribution of each feature in separate bars. Scatter plots show dependencies (e.g., how a specific feature affects the output, or how two features interact). The term “force plot” used henceforth refers to the individual force plot, as this is the most frequently used or mentioned type of force plot in the specific studies, in other research, and on the Web. The term “individual force plot” is used henceforth only in the Discussion, to avoid ambiguities with the collective force plot. The term “collective force plot” is explicitly stated henceforth. Details about SHAP are available in the relevant paper [39] and the online documentation of SHAP [40].

Integrated gradients constitute an XAI method introduced in [41] to determine the contribution of features in the output. This is a feature attribution method for local explanations. Integrated gradients cumulate gradients along the path from the baseline to an input. This satisfies two axioms: sensitivity, and implementation invariance. According to sensitivity, differing single features that result in different predictions are given non-zero attributions. Implementation invariance signifies that the same attributions are given to two functionally equivalent models that produce the same outputs for all inputs.

Grad-CAM (Gradient-Weighted Class Activation Mapping) provides visual explanations (heatmaps) denoting the important parts of an image in producing the output of the model. It was introduced in [42]. It is more associated with image-related applications (e.g., image classification and image captioning) and deep neural networks. It focuses on local explanations.

Partial dependence plots are used to provide graphical explanations. More specifically, they denote how the values of one or two features affect the output. Partial dependence plots were introduced in [43]. They can be useful in showing specific values of input features that have a great effect on the output [7]. They provide global explanations.

BertViz is an XAI tool specifically designed to visualize and interpret the inner working of transformer-based models, particularly BERT (Bidirectional Encoder Representations from Transformer). According to [44], BertViz plays a crucial role in demystifying black-box behavior in Large Language Models (LLMs), thus contributing to more transparent and trustworthy AI.

ExpliClas is a Web service providing explanations for Weka classifiers. It was introduced in [45] and provides text and graphical explanations. Local and global explanations may be given. The specific version of ExpliClas provides explanations for four classifiers: J48, RepTree, RandomTree, and FURIA (Fuzzy Unordered Rule Induction Algorithm) [46].

LIME (Local Interpretable Model-Agnostic Explanations) is a tool providing local explanations for any AI model. It was introduced in [47]. It identifies the influential features in producing the output for a specific instance. It works by approximating the AI model in the local region of the given instance. To do this, it analyzes how the output of the AI model changes with variations in the given instance.

FAMeX (FEature iMportance-based eXplanable AI algorithm) is an XAI method used in [25], citing other work that introduces it. This method uses a feature importance process to identify the most relevant features in producing the output. A criticality score is computed for each feature for this purpose.

## 4. Results

This section presents the main trends and a brief description of the reviewed studies. The presentation is organized according to the category to which the studies belong. First, approaches using XAI to assist in teaching and learning are presented in Section 4.1. In

Section 4.2, approaches using XAI in the context of AI as a learning subject are presented. In Section 4.3, approaches using XAI in policymaking, decision support, and administrative tasks are presented.

#### 4.1. Approaches Using XAI to Assist in Teaching and Learning

Various approaches using XAI in teaching and learning may be implemented. Two main types of these approaches may be discerned: (i) approaches in which XAI is used during learning and teaching, and (ii) approaches in which XAI is used in analysis and support tasks for learning and teaching. Four studies involved the use of XAI during learning and teaching [20,26,36,37]. The other studies concerned the use of XAI in analysis and support tasks for learning and teaching.

Relevant approaches are shown in Table 4, along with the corresponding main tasks performed by (X)AI, the learning subject, and the type of the approach. In the second column, the tasks implemented by the AI method that need to be explained are shown in plain text. The specific AI method(s) tested (or used) is (are) shown within parentheses. The tasks implemented by XAI are shown in italics in the corresponding column. The XAI tools explicitly mentioned are shown within parentheses. As shown in Table 4, multiple AI methods are used in some studies. In such studies, if the XAI tasks are based on one of the AI methods and not all of them, the specific AI method is shown in italics in Table 4. If the XAI tasks are based on all AI methods used, the AI methods are shown in plain text. Below, the main trends in these studies are outlined. Afterwards, the main aspects of several studies are briefly described in order to provide further details to the reader.

**Table 4.** Main aspects of approaches using XAI to assist in learning and teaching.

Citation	(X)AI Tasks	Learning Subject	Type
[23]	Gender prediction using eye movement (SVM, Random Forest, logistic regression, XGBoost, <i>LightGBM</i> ) <i>Feature selection for all tested AI methods (SHAP)</i> <i>Contribution of features in output (SHAP)</i>	Computational thinking	(ii)
[26]	Student's knowledge state prediction (deep knowledge tracing model) Prediction of student's response to a question in the next step (deep knowledge tracing model) Prediction of final test score ( <i>XGBoost</i> ) <i>Explanation to teachers about final test score prediction (SHAP)</i> <i>Advice to student about proposed activities to carry out (SHAP)</i>	Mathematics	(ii)
[27]	Machine translation of learning material (transformer models) <i>Visualization of the contribution of features in model output (SHAP)</i> <i>Visualization of the attention mechanism of transformer models (BertViz)</i>	Social studies, English as a second language	(ii)
[28]	Access inequalities in online learning ( <i>XGBoost</i> ) <i>Contribution of features in model output (SHAP)</i>	Mathematics	(ii)

**Table 4.** Cont.

Citation	(X)AI Tasks	Learning Subject	Type
[30]	Prediction of students' motivation in learning (XGBoost) <i>Contribution of features in prediction (SHAP)</i>	English as a second language in learning after school	(ii)
[20]	Classroom co-orchestration (teacher and AI system) <i>Explanations about the system decisions</i>		(i)
[31]	Classroom dialogue analysis (transformer model) <i>Visualization of attention weights</i>	Mathematics	(ii)
[32]	Classroom dialogue analysis (LLM and non-LLM models) <i>Visualization methods</i>	Mathematics	(ii)
[33]	Identification of educational items for revision Outlier detection (Isolation Forest), <i>Listing of item properties with extreme values for each outlier</i>	Mathematics, programming, English as a second language	(ii)
[34]	Proficiency and readability modeling of Portuguese (transformer models) <i>Contribution of features in output (SHAP)</i>	Portuguese as a second language	(ii)
[35]	Adaptive selection of a pedagogical strategy (reinforcement learning) <i>Feature contribution in strategy selection (integrated gradient analysis)</i>	Mathematics	(ii)
[36,37]	Alternative recommendations to students (Open Learner Model, probability models) <i>Personalized text-based explanations for the recommendations</i>	Support of students' self-regulated skills	(i)
[38]	Natural Language Processing (NLP) of interaction data in online learning platforms Prediction of academic performance using features extracted from NLP (decision tree, artificial neural network, CNN) <i>Contribution of features in prediction (SHAP)</i>	Language, English as a second language, mathematics	(ii)

#### 4.1.1. Main Trends in the Reviewed Studies Reviewed Studies Using XAI to Assist in Teaching and Learning

Three studies concerned the use of XAI to provide real-time interaction in the context of e-learning systems [26,36,37]. One of these studies was a revised and extended version of another one [37]. These approaches may support purely distance learning activities and blended learning activities (i.e., a combination of face-to-face and Internet-based learning activities). In this context, XAI may be used to provide real-time explanations and advice to teachers and students. These aspects have been explored in several studies [26,36,37]. Two alternative approaches to the provision of explanations to students and teachers may be implemented. In the former, students and teachers acquire explanations about different aspects [26]. In the latter, teachers acquire the explanations provided to students [36,37].

In Internet-based activities, and especially in purely distance learning ones, teachers need to dedicate time in order to supervise students, deduce their progress, and interact with them in a timely and concise manner. Teachers may obtain feedback from XAI about

the students' overall estimated knowledge, their deficiencies, and their performance in the specific implemented tasks [26,37]. The availability of this information may reduce the amount of time that teachers would have to dedicate in order to deduce it by themselves. This enables them to take the necessary actions and interact accordingly with students. Students may directly benefit from XAI in their interaction with e-learning tools. As they work on their own, and usually in a different space and/or time from teachers, they need assistance in selecting the next tasks to implement. XAI may provide this assistance [26,37].

XAI may be used in classroom and lab activities. Teachers may employ educational presentations and demonstrations using XAI. Educational applications based on XAI may be shown to students by teachers, and students may also use them in technology-enhanced activities. Assessment and interaction in the learning procedure may benefit from XAI. XAI may be also used for classroom orchestration. In [20], the design of a classroom co-orchestration approach is discussed, in which the orchestration load is shared between the teacher and an AI system. In this context, XAI explains decisions to the teacher.

XAI may be used to perform analysis and support tasks for learning and teaching. As shown in Table 2, a type of task concerns the processing of data collected in the specific learning environment(s) using XAI. Relevant data may be collected before (e.g., pre-tests), during, and after learning. Different research directions may be explored. One main direction may concern the analysis of data collected from students. Data about individual student performance and interaction with the e-learning tool may be stored in e-learning tool databases [28]. In [28], data from completed math exercises in a math intelligent tutoring system were analyzed to assess access inequalities in online learning. SHAP was used to identify contributing features. Data in e-learning tools may also concern the interactions among students [38]. For instance, in [38], data from chat-based interactions among students who used an e-learning platform within the classroom were analyzed to predict students with low and high academic performance. SHAP identified the contribution of features in the prediction. Data may be collected from students via other methods as well (e.g., questionnaires, observations, recordings, interviews, discussions, projects, pre-tests, and post-tests) and processed with the assistance of XAI. In [30], students used tablets to interact with educational apps in rural learning centers after school. SHAP plots (i.e., waterfall and beeswarm summary plots) were used to explain the predictions of XGBoost (eXtreme Gradient Boosting) [48] with respect to students' motivation. In [23] data about students' eye movements in a virtual reality classroom were collected, with the purpose of identifying gender differences in learning. SHAP was used to perform feature selection in order to improve the performance of the tested classification model, and to provide explanations for the outputs of the best model using global bar and beeswarm summary plots. A second main direction concerns the analysis of data collected from teachers after the learning procedure, mainly using questionnaires, interviews, and discussions. A third direction is to analyze data concerning the interactions between students and teachers during learning [31]. A fourth main direction is to analyze aspects regarding the mechanisms of an AI-based e-learning system (that is, to explain the decisions made by the e-learning system). In this context, an e-learning approach using reinforcement learning to provide adaptive pedagogical support to students was presented in [35]. It is difficult to interpret the function approximation of the reinforcement algorithm. Therefore, XAI (i.e., integrated gradients) was used to determine the key features affecting the selection of the pedagogical policy by reinforcement learning.

Among the studies that use XAI for analysis and support tasks are the ones concerning language data in the form of text or oral data. Relevant studies include those in [27,31,34]. Various research directions may be explored in this concept. One direction is to analyze language data collected from the specific learning environment and acquire information

about the learning process. Such an approach is presented in [31], in which students' and teachers' classroom dialogues were analyzed using a transformer model, i.e., the Global Variational Transformer Speaker Clustering (GVTSC) model. The XAI method used concerns the visualization of attention weights. This work was extended in [32] using visualization methods for the GVTSC model and a Large Language Model. Another direction is to process available language data from previous studies in order to acquire information assisting in language learning. This direction was explored in [34]. More specifically, linguistic modeling (i.e., proficiency and readability modeling) using available corpora was performed to identify linguistic properties that play an important role in language learning. SHAP was used to identify the most important features. Another direction concerns the creation of new learning content from available text or oral data. In [27] work towards this direction by translating learning content is discussed.

XAI may be used to support teaching and learning by assisting in the creation of new learning content and the management of existing learning content. Automated and semi-automated processes save time that may be dedicated to other tasks. This is particularly true in cases of creating/managing text-based content and a large number of educational items. It should be mentioned that the content's authors may be unavailable or may have limited spare time. Relevant studies include [27,33]. In [27], machine translation with transformer models was used to translate learning content from English to the first language of students, and vice versa. XAI was used to explain the transformer models. SHAP provided plots to denote the contribution of input text features in producing output text. BertViz was used to visualize the attention mechanism of the models (i.e., the strength of weights in connections between tokens). In [33], the identification of educational items in an e-learning platform requiring revision was explored. Outlier detection with explanations and interpretable clustering were applied.

#### 4.1.2. Brief Description of Reviewed Studies Using XAI to Assist in Teaching and Learning

In [23] gender differences in computational thinking skills in the context of an immersive virtual reality classroom were investigated. Eye movements were used as a biometric input to predict gender classification. Sensors were used to collect data from the participants. Five classification models were evaluated: Support Vector Machine, logistic regression, Random Forest, XGBoost, and LightGBM (Light Gradient-Boosting Machine) [49]. LightGBM exhibited the best performance. SHAP was used for two purposes: First, SHAP was used for feature selection. The performance of all five models was improved after carrying out feature selection, but LightGBM still outperformed the other models. Second, SHAP was used to explain the outputs of LightGBM, the model with the best performance. A global bar plot was used to classify the features according to their importance, and to denote their average impact on the output. A beeswarm summary plot was also used to summarize the positive and negative contribution of each feature across the dataset. This specific approach can be used to provide personalized learning support and assist in the design of the educational setting and adaptive tutoring systems.

In [26], an approach within the context of an e-learning platform is presented. The student takes a diagnostic test when they initially register on the platform. The approach models a student's knowledge state using a deep knowledge tracing model [50] and predicts the response to a question in the next step. XGBoost is used to predict a student's score in the final test based on the results of a diagnostic test. The learning subject is mathematics. Proprietary data from the KOFAC (Korean Foundation for the Advancement of Science and Creativity).were used. Knowledge concepts included in the diagnostic test are associated with Shapley values, which are used to explain the predicted final scores to teachers. Teachers obtain a viewpoint on students' knowledge state and performance,

saving time and assisting them in their online interaction with students. Shapley values are also used to provide advice to students about which activities they ought to do, and in which order. Therefore, the explainable component provides online feedback to teachers and students.

English is the official language in Uganda and is used in school education, but various ethnic languages are the first languages of many students. Luganda is a language spoken in rural central Uganda. This makes it difficult for students to comprehend the concepts taught in schools. The translation of learning material assists in its comprehension by students. A machine translation approach was carried out to translate primary-school social studies notes from English to Luganda, and vice versa [27]. Transformer models based on the Marian machine translation framework [51] were used for the translation. XAI was used to provide explanations for the outputs produced by the transformers. Two XAI tools were used: SHAP and BertViz. SHAP was used to provide a visualization of the contribution of features to the output. The force plot is explicitly mentioned in the paper. BertViz provided a visualization of the attention mechanism of the transformer models. Connections between tokens were visualized according to the strength of the corresponding attention weights. This study was one of the few to use multiple XAI tools.

Tablets are a convenient technological resource for areas with limited resources. In this context, a mobile learning approach using tablets was implemented after school in seven learning centers in rural parts of India [30]. The learning subject was English as a second language. The approach explored the factors affecting students' motivation. The ARCS model [52], encompassing four main traits for motivation (i.e., attention, relevance, confidence, and satisfaction), was used. Educational apps for language learning were installed on the tablets, and a specific curriculum was followed. Records were kept for the attendance of students and their usage of tablets. The students also replied to a survey. XGBoost was used to predict the time the students dedicated to using the educational applications. SHAP was used to provide explanations for the derived predictions. Waterfall and beeswarm summary plots were used to show the contribution of the four features in the prediction.

In [20], the design of a co-orchestration approach in technology-enhanced classrooms is presented. It explores how responsibilities for managing social transitions in learning environments can be shared between teachers and AI systems. It aims to reduce teachers' orchestration load while maintaining their sense of control. It emphasizes balancing teacher and AI system responsibilities to enhance the fluidity of learning transitions. The design and a prototype of the AI system were evaluated by seven teachers from different schools that teach in various grades. A prerequisite in the design of the AI system is to explain its decisions using the notion of XAI. This will enable teachers to assess whether the decision of the system is incorrect, in which case further action is needed by them. The specific type of XAI is not explicitly mentioned in the paper. The approach focuses on supporting the orchestration of social transitions, and not on other types of support.

The evaluation of educational items in e-learning platforms is generally an interesting topic, due to the large amount of items, the large number of users, the large amount of interaction data, and the variety of item content. Automated or semi-automated item evaluation methods enable content improvement, saving time for content authors, and may enhance learning and teaching. An approach to identify educational items for revision is presented in [33]. Explainable outlier detection and interpretable clustering were used to identify candidate items for revision. The educational items under consideration cover a wide range of school subjects (i.e., computing, mathematics, and English as a second language) and are available in an e-learning platform. The Isolation Forest algorithm [53] was used for outlier detection. Explanations were provided for the detected outliers to

assist content authors. The researchers explored different methods for outlier explanations, including 2D scatter plots, as proposed in the LOOKOUT algorithm [54]. These 2D scatter plots are a type of lossless XAI visualization method that preserves all information in the data [17]. The researchers decided to use simple explanation methods. For each outlier, a listing of item properties with extreme values was presented. The basic boxplot method was used to identify properties with extreme values, and afterwards, related properties were grouped, assigning them comprehensible labels. Interpretable clustering concerns clustering approaches in which concise descriptions are given for clusters. This could be based on combinations of properties that often occur together.

An approach that could benefit the learning of Portuguese as a second language is presented in [34]. The researchers analyzed available corpora in Portuguese to perform linguistic modeling regarding the proficiency and readability levels. Linguistic complexity measures are associated with the proficiency and readability levels. Types of complexity measures include superficial, lexical, morphological, syntactic, and discourse-based measures. Four types of classifiers (i.e., Support Vector Machine, linear regression, Random Forest, and a multi-layer neural network) were tested for the proficiency and readability levels separately. SHAP was used to provide explanations by examining the contribution of features in the derived results. For this purpose, global bar plots and beeswarm summary plots were used.

In [35], an e-learning system is presented using reinforcement learning to adaptively assist in math learning. The system employs a narrative-based educational approach. Reinforcement learning is used in an AI guide to adaptively select a pedagogical strategy for a specific student (i.e., direct hints, generic encouragement, guided scaffolding prompts, or passive positive acknowledgment). The goal of reinforcement learning is to learn a decision policy by maximizing a reward function that takes into account the learning objectives. Features taken into consideration by reinforcement learning include fixed student features, learning activity features, and features concerning student interaction and performance during learning. The system was tested with students whose knowledge was evaluated with pre-tests and post-tests. The results were positive for students with low performance in the pre-tests. XAI was used offline to explain the decisions of reinforcement learning. This is needed because the policy optimization algorithms used in reinforcement learning to learn a decision policy that maximizes the reward are often difficult to interpret. This is the case for the proximal policy optimization algorithm [55] used in the study. Integrated gradient analysis was used to identify the most important features in policy selection. The results showed that the most important features are math anxiety and pre-test scores.

Tsiakas et al. [36,37] presented the architecture of a system intended to support students' self-regulated skills. The overall concept of the system is based on a cognitive game including parameters whose configuration corresponds to a variety of skills and abilities. The main self-regulated skills supported are goal setting, self-efficacy, and task selection. The training process consists of sessions organized in rounds. An Open Learner Model [56] is used to record the student's progress. Recommendations are provided to students at the beginning of each session (i.e., target score) and at the end of each session round (i.e., next task configurations). Alternative recommendations are provided based on the students' profiles and context. Recommendations are based on probability models. For each recommendation, personalized explanations may be provided. Teachers may view and edit recommendations and explanations.

#### 4.2. Approaches Using XAI in the Context of AI as a Learning Subject

AI may be integrated in primary education as a learning subject. It is considered to be useful to teach AI at all educational levels, due to the importance of AI in everyday tasks. AI as a learning subject involves various dimensions. Among others, these concern the acquisition of knowledge about (i) basic AI concepts, technologies, and processes; (ii) the role of AI in society and the responsible use of AI; and (iii) AI as a learning tool. The term “AI literacy” is used to define the knowledge about AI that every student needs to possess as a member of society.

Appropriately designed learning activities that correspond to the age of students are needed for teaching AI. This is also the case for relevant educational applications, presentations, and demonstrations that may be available to students in the classroom and through the Internet. Laboratory activities concerning AI may also be implemented. Alternative approaches may be implemented in the teaching of AI. On the one hand, the teaching of AI may be performed autonomously. On the other hand, the teaching of AI may be carried out in combination with or in the context of the teaching of other subjects (e.g., computational thinking, robotics, science, mathematics, language). XAI may be part of the AI curriculum.

Two studies were found in which XAI was used in the context of AI as a learning subject. Table 5 shows the main aspects of these studies. More specifically, for each study, the following are outlined: (a) the taught subject; (b) the tasks implemented by the AI method(s) shown in parentheses, for which explanations are provided; and (c) the tasks implemented by the XAI tools shown in parentheses. The cell contents concerning (a) and (b) are shown in plain text, whereas the contents concerning (c) are shown in italics. The following paragraphs include a short presentation of the three studies.

**Table 5.** Main aspects of approaches using XAI in the context of AI as a learning subject.

Citation	Main Goals
[22]	Workshops teaching (X)AI to children in combination with visual programming and sports Classification of selected players (voting scheme of J48, RepTree, RandomTree, and FURIA) <i>Text and visual explanations for classification (ExpliClas)</i>
[21]	Raise awareness and improve students' comprehension of bias and fairness in AI decision-making Image classification (CNN with LSTM) <i>Visual explanations for classification (Grad-CAM)</i>

Bias inherent in AI is an issue worthy of being taught to children, as AI affects many social sectors. Melsión et al. [21] focused on educating students about gender bias in supervised learning. Gender bias creates discrimination in AI decisions. An online educational platform was used for learning. A dataset consisting of images with gender bias was used to train the model, i.e., a combination of a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM). More specifically, the dataset used was a subset of the Microsoft COCO (Common Objects in Context) public dataset. XAI was incorporated in the platform to facilitate students' learning. More specifically, Grad-CAM provided visual explanations to help preadolescents understand gender bias in supervised learning. The primary goal was to raise awareness and improve students' comprehension of bias and fairness in AI decision-making.

Alonso [22] presented an approach for teaching AI to children in the context of workshops in a research center. Children were initially introduced to AI and its applications.

They then interacted with a specialized application implemented in Scratch, a popular visual programming environment for children. The application identified the roles of basketball players. For a selected player, four WEKA classifiers (i.e., J48, RepTree, RandomTree, and FURIA) were employed to perform classification. Voting was used to produce an output. Explanations generated by ExpliClas were shown in text and visual format. Children evaluated the explanations and then learned how they were generated. This specific approach combines XAI learning with programming concepts and sports.

#### 4.3. Approaches Using XAI in Policymaking, Decision Support, and Administrative Tasks

XAI may be used to obtain information from available data, assisting in educational policies, decision-making, and administrative tasks in primary education. Approaches that belong to this category are shown in Table 6. The main goals of these approaches concern the following: (i) student performance prediction, (ii) student performance analysis, (iii) identification of key factors affecting student performance, (iv) prediction of the required number of teachers by region, and (v) prediction of student attendance.

**Table 6.** Main aspects of studies using XAI in policymaking, decision support, and administrative tasks.

Citation	AI Tasks
[24]	Prediction of student attendance (a black-box AI model) <i>Logical rule-based explanations (Isabelle Insider and Infrastructure framework and the precondition refinement rule algorithm)</i>
[25]	Prediction of students' adaptability to online education (Random Forest) <i>Identify learning parameters increasing students' adaptability (LIME, SHAP, FAMeX)</i>
[7]	Prediction of the supply and demand of teachers by region (XGBoost) <i>Feature importance</i> <i>How numerical changes in a feature affects prediction (partial dependence plots)</i> <i>Contribution of features in the prediction (SHAP)</i>
[29]	Investigation of predictive factors for student approaches to math learning (XGBoost), math taught in second language <i>Contribution of predictive factors to math performance (SHAP)</i>
[5]	Modeling of students' math performance (CatBoost) <i>Identification of key factors affecting student math performance (SHAP)</i>
[6]	Prediction of student math learning outcomes (KNN, SVM, Random Forest regressor) <i>Identification of key predictive features (SHAP)</i>
[4]	Student performance prediction involving primary education learning subjects (linear regression, Random Forest, neural network) <i>Feature importance</i>

##### 4.3.1. Main Trends in the Reviewed Studies Using XAI in Policymaking, Decision Support, and Administrative Tasks

Table 6 summarizes the main aspects of the relevant studies. In the second column, the tasks performed by XAI are shown in italics, and the XAI tools explicitly mentioned are shown within parentheses. In the same column, the tasks implemented by the AI method(s) that need to be explained are shown in plain text. The specific AI method(s) tested (or used) is (are) shown within parentheses. In studies using multiple AI methods, if the XAI tasks are based on one of the AI methods, the specific AI method is shown in italics in Table 6. If the XAI tasks are based on all of the AI methods used, the AI methods are shown in plain text.

One may note that student performance is a key aspect in most of these studies (i.e., four studies). An explanation for this is the fact that student performance is an aspect that is frequently investigated in education. A further explanation is that performance in primary education affects performance at subsequent educational levels. Each of the aforementioned goals (iv) and (v) were pursued by a single study. Generally speaking, it is interesting that various aspects concerning education were explored.

Obviously, educational data may provide information for decision-making in primary education. However, in addition to educational data, other types of data (e.g., family, financial, and population data) may prove useful as well. Within specific countries, educational and other types of data are available at the national, regional, municipal, or school level. Open government data may assist researchers in conducting research that combines educational data and other types of data. Large-scale educational assessments may also be sources of educational and other types of data [4,5]. Typical examples of large-scale international educational assessments include the PISA (Program for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study), and PIRLS (Progress in International Reading Literacy Study). Nadaf et al. [5] used data from TIMSS assessments. Silva et al. [4] used data from large-scale educational assessments within Brazil. Sanfo [6] used data from an international survey concerning French-speaking countries.

Several indicative studies combined educational data and other types of data [4–7,25]. Lee [7] presented an approach predicting the number of teachers required in each region. He used basic educational data such as the numbers of primary education teachers, students, schools, and classes. Further data used included data affecting the size and structure of the population (e.g., numbers of births, deaths, marriages, and divorces), economically active population data, and population movement data. A finding was that the economically active population is the most important factor affecting the prediction of the required number of teachers. Silva et al. [4] combined educational data (e.g., average number of students per class, teachers' education, average daily teaching hours of teachers, students' performance in large-scale assessments), economic data (i.e., investments in education), and social well-being data. A finding was that the teachers' characteristics play the most important role in students' performance. Sanfo [6] used data from an international survey combining educational data with learning environment data and data about students' and teachers' backgrounds. The purpose of the approach was to predict students' learning outcomes. Nadaf et al. [5] combined educational data with family-related and demographic data. Kar et al. [25] used a dataset derived from a survey in Bangladesh involving students from all educational levels and, among others, including the financial status of the family.

Research may be carried out with datasets that are available beforehand. This was the usual case in almost all of the studies. However, datasets may be also created as a result of a specific study. For instance, in [29], the dataset used was created by the researchers from several schools in a country-level study based on students' results in a math and an English test, according to international standards and students' responses to a questionnaire.

Learning subjects explicitly stated in the studies are mentioned in Table 6. Obviously, two of the studies [7,24] do not concern a specific learning subject. In another study [4], specific learning subjects are not mentioned, but it may be assumed that a combination of learning subjects are involved.

#### 4.3.2. Brief Description of Reviewed Studies Using XAI in Policymaking, Decision Support, and Administrative Tasks

Attendance is an important factor in education, as it affects overall student performance and participation in learning activities. By predicting when a student will be absent and the reasons for this, actions can be taken to prevent it. In this context, a theoretical

approach for the prediction of school attendance is presented in [24]. The explainable approach is based on the Isabelle Insider and Infrastructure framework and the precondition refinement rule algorithm. A logical rule-based explanation is provided for the output of a black-box AI method (not specifically mentioned). The overall framework is based on computation tree logic (i.e., a type of temporal logics using branching) [57], attack trees (i.e., a model for risk analysis), and Kripke structures [58].

In the study of Lee [7], an approach predicting the number of required teachers in each region was applied. XGBoost was used for the predictions. Data sources involved Korean national public databases, the Korean Educational Development Institute (KEDI) and the National Statistical Office (NSO). The XAI methods used were feature importance, partial dependence plots, and SHAP. Feature importance showed that the most important features are the following (in descending order of significance): economically active population by city, number of classrooms by city, birth rate by city, number of students by city, and population migration. Partial dependence plots were used to show how numerical changes in a feature (e.g., economically active population) affected the prediction. The number of required teachers increased as the economically active population increased. Four different types of SHAP plots were used: beeswarm summary plots, force plots, scatter plots, and plots accumulating Shapley's influence on all data. Beeswarm summary plots showed that the economically active population and the number of classrooms were the most important features in the prediction, and the variance of their SHAP values was greater. Force plots showed the contribution of features in teacher number prediction for specific cities and years. Scatter plots showed an above-average increase in the required number of teachers in case a specific value of the economically active population was surpassed.

Miao et al. [29] used SHAP extensively to explore the contribution of predictive factors to students' math performance. A country-level study was carried out. XGBoost was used for performance predictions. Various SHAP plots were used. A force plot was used to provide a viewpoint on factors contributing positively and negatively to a specific student's performance. A global bar plot and a beeswarm summary plot were used to derive the most important factors in performance prediction and the range of the SHAP values. The four most important factors were language (i.e., vocabulary and grammar skills), the grade level, reading (i.e., reading comprehension), and self-efficacy in math. It should be mentioned that math is taught in English (second language) in the specific country. The researchers then segmented student performance into four groups (very poor, poor, normal, and outstanding), according to test scores, and studied the SHAP values of various factors separately for each group. Scatter plots were used for this purpose. More specifically, scatter plots were created for each category for the following: (a) language test scores and SHAP values for the language feature, (b) reading test scores and SHAP values for the reading feature, (c) self-efficacy level in math and self-efficacy SHAP values, (d) the ability of students to use reading and problem-solving strategies and corresponding SHAP values, and (e) the ability of students to exercise PISA math learning strategies and corresponding SHAP values. Thresholds in the levels of these features for the corresponding SHAP values to become positive were derived. As expected, better (worse) levels in these features resulted in better (worse) math performance.

In [5], the factors affecting UAE students' math performance in the TIMSS international assessment were examined. A boosted regression tree model, i.e., CatBoost [59], was used to model students' performance. SHAP was used to assess the contribution of features to the students' performance. Two types of plots produced by SHAP were used: beeswarm summary and dependence plots. According to the beeswarm summary plot, the three most important key features contributing to math performance were students' confidence in mathematics, language difficulties, and familiarity with measurements and geometry. As

far as the feature “language difficulties” is concerned, the distribution of SHAP values was wide, demonstrating the heterogeneity in its contribution (from very negative to very positive). This observation led the researchers to use a dependence plot for the contribution of language difficulties to math performance. A nonlinear relationship between them was observed.

Sanfo [6] used data from Burkina Faso’s 2019 Program for the Analysis of CONFEMEN Education Systems (PASEC). The initials CONFEMEN stand for Conference of Ministers of Education of French-Speaking States and Governments. Educational data were combined with other types of data (i.e., learning environment data and data about students’ and teachers’ backgrounds). Machine learning models were used to predict students’ learning outcomes. SHAP was used to identify the main features contributing to the results of three machine learning models exhibiting the best performance in classification (KNN, SVM) and regression (Random Forest regressor) from the set of models tested. The global bar plot (i.e., bars are the mean absolute SHAP values for the whole dataset) was used for global explainability, and the waterfall plot (i.e., SHAP values for a specific dataset instance) was used for local explainability. The features identified as most important for KNN and SVM were in the following order of importance: local development, community involvement, school infrastructure, and teacher experience. The features identified as most important for Random Forest regressor were in the following order of importance: school infrastructure, grade repetition, local development, and teacher experience. The use of different machine learning models verified the importance of three of the four features and provided a different viewpoint on one of the four features (grade repetition or community involvement) and the importance ordering.

In the work of Silva et al. [4], open government data and data from large-scale assessments were used for three purposes: (a) to explore the correlation between student performance and other parameters (e.g., teachers’ education, class size, social well-being, municipality investments), (b) to predict student performance based on past data, and (c) to explore the consistency of the correlation among student performance and other parameters and the analysis of feature importance. They used artificial neural networks, linear regression, and Random Forest for student performance prediction. Feature importance analysis was performed only for the Random Forest, which exhibited the best performance. It should be mentioned that the specific learning subjects involved are not mentioned in the specific paper, but they may be deduced. Student performance prediction is based on data concerning the school achievement rate and national exams. A search on the Web showed that the specific national exams involve at least language (Portuguese) and mathematics. Obviously, school achievement depends on all or most of the taught learning subjects.

## 5. Methodological Quality of the Reviewed Papers

An aspect of interest is to discuss the methodological quality of the reviewed studies. One may assess the methodological quality of each study in the following dimensions: (i) study design, (ii) sample size, (iii) data source, and (iv) limitations identified. Table 7 presents relevant data.

The studies concern varying numbers of participants, ranging from less than ten to hundreds or thousands. There are also studies in which the number of participants is not mentioned or details are missing. There are also studies in which there were no participants to evaluate the approach.

**Table 7.** Main aspects of the methodological quality of the reviewed studies.

ID	Citation	Study Design	Sample Size	Data Source	Limitations Identified
1	[22]	Design and implementation case study as a workshop (XAI for kids)	Not precisely defined (20 school students in the age range from 6 to 17 years old)	Training data (80 samples) created by the researcher User feedback data	Gender bias in training data Slightly better accuracy of black-box models Simplified interaction design Lack of detailed evaluation with humans
2	[23]	Quantitative experimental study using machine learning and eye-tracking data	381 sixth-grade students (final 280 participants, M:140, F:140)	Eye-tracking data (>3600 samples) created by the researchers Behavioral and physiological measures Questionnaires	Specific context No longitudinal measurement Interpretability challenges No direct correlation to actual computational thinking learning outcomes
3	[24]	Explanatory modeling	Synthetic data (gender, location, transport, ethnicity, special needs)	Synthetic data (not sourced from an actual dataset)	No real data used Certain abstract definitions about explanations Not empirically validated
4	[25]	AI analysis of parameters affecting	1205 students at all educational levels in Bangladesh	Survey data from 1205 students in all educational levels	Generalizability
5	[26]	Experimental design using AI in a real-world educational setting	59 first-grade and 227 second-grade students (5 primary schools in South Korea)	Proprietary data from the KOFAC	Data availability Limited evaluation
6	[27]	Machine translation model designed to translate social studies notes from Luganda to English, and vice versa	The number of participants is not specified	A dataset (4000 words) created from notes by the researchers (publicly available) and another publicly available dataset	Future extensions to other subjects No extensive evaluation by teachers
7	[7]	Predictive modeling	Structured public datasets (data from 17 cities and provinces, 2001–2019)	Korean national public databases, KEDI and NSO	No human evaluation Further research needed with additional features
8	[28]	Data mining from e-learning platform (quantitative data)	Students from three countries	Dataset from e-learning platform (random subsample of 5000)	Study prior to and during COVID-19 Quantitative and aggregative approach
9	[21]	Mixed-methods experimental study	76 5th- and 6th-grade students from a primary school	A subset of the Microsoft COCO public dataset	Gender bias in the dataset Sample size and scope Online study constraints Short-term exposure to tool

**Table 7.** Cont.

ID	Citation	Study Design	Sample Size	Data Source	Limitations Identified
10	[29]	AI-based analysis of test results and questionnaires	5th- to 9th-grade students in 20 public schools in Abu Dhabi (high-quality data from 1660 students, at least 305 in primary education)	Data from math diagnostic tests and questionnaires (included measures of student self-efficacy, metacognitive strategies, instructional language skills, and math performance outcomes)	Data availability Lack of data on teachers' knowledge/skills, metacognitive strategies Dependence on data from student questionnaires
11	[30]	AI-based analysis of data from students using educational tablets (quantitative study)	135 1st- to 5th-grade students (after school centers, three states in India)	Interaction data Student questionnaire	No qualitative data Generalizability beyond India
12	[5]	AI-based analysis of the UAE students' math performance in the TIMSS 2019 assessment	22,163 4th- and 27,342 8th-grade students from the UAE	TIMSS 2019 for the UAE	Economic and metacognitive factors not covered in TIMSS Detailed impact of each learning factor not explored
13	[20]	Classroom orchestration and transition taxonomy	7 teachers from 5 different primary schools teaching various grades (convenience sample)	Insights and feedback gathered from teachers during co-design workshops and prototype evaluations	Sample size Prototype scope No data collected from classrooms
14	[31]	Classroom dialogue analysis and impact of classroom utterances in student learning	Data from a 45 min classroom session, 2 classes, 4th and 6th grades in a primary school	Collected from recorded teacher and student dialogues: 193 and 274 utterances (4th and 6th grades, respectively)	Sample size Data from one classroom session
15	[32]	Classroom dialogue analysis and impact of classroom utterances in student learning using two models	Data from classroom	Samples of dialogues	Sample size Size of dialogue data
16	[33]	Mixed (analyzed item properties, employed both simulation and real-world data analysis, conducted two case studies)	Data from a large-scale learning environment used by tens of thousands students (8–15 years old)	Pool of items Item metadata Students' responses to items	Sample size details are missing Bias in data (student/system behavior) not considered Detailed analysis only for text items

**Table 7.** Cont.

ID	Citation	Study Design	Sample Size	Data Source	Limitations Identified
17	[34]	Linguistic modeling (i.e., proficiency and readability modeling) using available corpora	The number of participants is missing	European Portuguese Learner Corpus, Brazilian school materials	Complexity of measures Missing details about human evaluators Class imbalance in corpora Need for enhanced corpora metadata
18	[35]	Evaluate the effectiveness of a learning system based on reinforcement learning to enhance math concept learning, 2 case studies	Approximately 300 students	Interaction data from students using the system	Sample size details are missing Remote learning due to COVID-19 Long-term impact
19	[6]	AI-based analysis of math learning using PASEC 2019 data	Students from Burkina Faso who participated in PASEC 2019	Data from the 2019 PASEC for Burkina Faso	Generalizability of findings Selection of features
20	[4]	Statistical and AI-based analysis of student performance using open government and large-scale assessment data	Data from Brazilian public education institutions	Open government data (Brazil) Large-scale assessment data (Brazil)	Generalizability beyond Brazil Selection of features Selection of calendar years of collected data
21, 22	[36,37]	Conceptual study of a system aiming to enhance self-regulated skills	Conceptual study	Comprehensive review	Empirical validation missing Implementation details are missing
23	[38]	Analysis of online live interaction for student performance prediction	>100,000 students in online platform	About 10 million interactive texts	Collection of data from an online platform only Need for more fine-grained dialogue analyses Student only dialogues

As far as the data sources are concerned, one may note that, in most studies relying on available datasets, there was an effort to mainly use publicly available datasets. This facilitated other researchers to conduct research with other methods and compare the results. In some cases, datasets created in the context of studies were made publicly available, also facilitating further research. In a specific study, a proprietary dataset was used.

The main limitations indicated concerned missing details (e.g., about the participants and their number, implementation details). Limitations concerning the evaluation were also indicated. There are studies lacking an evaluation with humans, or this evaluation is not thorough. A further limitation concerned the selection of features while implementing the AI models.

## 6. Discussion

A number of studies concerning XAI in primary education have been presented. XAI has shown its usefulness in different contexts. Obviously, teachers and parents of students may have some reservations about the use of AI in educational settings. However, some of these reservations may be surpassed due to XAI. The following sections discuss various aspects of the reviewed studies.

### 6.1. Answers to the Research Questions

In Section 2, seven research questions were mentioned, guiding the preparation of this review. Answers to these questions have been given in previous parts of this paper. In this section, we mention previous parts providing the answers, and we also discuss further aspects.

Research Question 1 involves the main categories in which XAI approaches in primary education may be discerned. Three main categories are described in Section 3 (text, Figure 1, Tables 2 and 3). The relevant studies are presented in Section 4, according to the category to which they belong.

Most studies belong to category A or C. Relatively fewer studies belong to category B. There are a number of reasons for this. The teaching of AI to young students is a more recent development compared to the use of AI to assist in teaching, learning, policymaking, decision support, and administrative tasks at the specific educational level. Furthermore, it is more straightforward to implement approaches belonging to category A or C. Approaches belonging to category B require careful design and the establishment of the corresponding curriculum. According to Chiu [18], there does not seem to be a consensus among researchers about what needs to be taught concerning AI and how it should be taught to children. A further comment is that school teachers and students are obviously not experts in AI. This means that they need to interact with learning content and tools that they comprehend. This is especially the case for students, due to their young age. During the last few decades, experience in teaching AI in higher education has accumulated based on relevant learning outcomes, strategies, content, and tools. A corresponding process is required for primary schools (that is, gaining experience in teaching AI to young students according to learning outcomes and strategies, and using appropriate content and tools that need to be prepared).

Research Question 2 was as follows: “What are the main trends of the XAI approaches in primary education?” Answers to this question are given in Sections 3 and 4.

Research Question 3 was as follows: “What are the main XAI tools or methods used in the XAI approaches in primary education?” Table 3 summarizes the XAI tools or methods used. In Section 3, the general functionality of the main XAI tools is outlined. In Section 4, it is explained how XAI tools or methods are used in the context of the reviewed studies. The relevant aspects are outlined in Tables 4–6.

Research Question 4 was as follows: “In which primary education learning subjects are the XAI approaches used?” In Section 4, learning subjects are mentioned for the reviewed studies, and they are outlined in Tables 4–6. These aspects are summarized in Table 8, which outlines the studies per learning subject. The studies involving a specific teaching subject explicitly mentioned are shown in Table 8, along with the two studies concerning general skills (i.e., self-regulated skills). The learning subjects are listed in alphabetical order in the table. The reviewed approaches deal with various learning subjects in primary education, including studies in all categories. It is useful to apply XAI in various learning subjects and examine the derived benefits.

**Table 8.** Studies per learning subject.

Learning Subject	Studies
(X)AI	[21,22]
Computing	[22,23,33]
Language	[4,38]
Mathematics	[4–6,26,28,29,31–33,35,38]
Second language	[27,29,30,33,34,38]
Social sciences	[27]
Self-regulated skills	[36,37]
All or most of primary education learning subjects	[4]

Note that not all reviewed studies involve a specific learning subject. These studies are those of Lee [7] and Kammüller and Satija [24], which concern the prediction of the supply and demand of teachers by region and the prediction of student attendance, respectively. Therefore, they may not be associated with a specific learning subject. Furthermore, in the work of Kar et al. [25] analyzing survey data, no learning subject is mentioned in the dataset.

Furthermore, there are studies involving a combination of learning subjects [4,27,29,33]. It should be mentioned that the study of Silva et al. [4] concerns a combination of learning subjects because student performance was predicted using data from a large-scale educational assessment within Brazil. The assessment concerned the school achievement rate, which obviously depends on all (or most) of the teaching subjects and national exams, which involve at least language (Portuguese) and mathematics. For these reasons, the study of Silva et al. [4] is mentioned in the subjects of language and mathematics in Table 8. There is also a separate row in the table mentioning that Silva et al. (2024) [4] are associated with all or most of the learning subjects.

One may note that there is a tendency for using XAI in mathematics. More specifically, eleven of the twenty studies in Table 8 concern mathematics. This may be attributed to the difficulties that certain students face in mathematics specifically. Learning tools and analysis of math learning have attracted the interest of researchers. Examples of interactive learning tools for mathematics are presented in several studies [26,33,35]. Kim et al. [26] considered XAI in dynamically assisting in interactive learning activities. The other four approaches [28,33,35,38] involve XAI in the analysis of data stored in learning tools to provide offline assistance to teachers. Onishi et al. [31,32] also considered offline assistance of XAI to teachers (i.e., analysis of classroom dialogues). Mathematics is also a subject of large-scale assessments in education. Examples of relevant approaches include the works of Miao et al. [29], Nadaf et al. [5], and Sanfo [6]. Some of the reviewed studies have shown the interrelation of mathematics and language; that is, the knowledge level in language affects math learning.

XAI in second-language learning seems to be a direction followed by certain approaches. In Table 8, six relevant studies are included in the list of reviewed papers. Most countries are multilingual, and second-language learning is important for communication among their inhabitants. There are also countries in which the official language taught in schools differs from the first language of a percentage of the population. In addition, there are examples of countries (e.g., the UAE) in which the curriculum is taught in English [29], which differs from the first language of the country's population. XAI has shown the ability to provide solutions in different scenarios involving second-language learning.

Research Question 5 was as follows: "For which AI methods were XAI tools or methods used to provide explanations in XAI approaches in primary education?" The specific AI methods are outlined in Table 3. Section 4 discusses which AI methods were

used in the studies, along with their functionality. Tables 4–6 outline the AI methods used and their functionality. In some studies, multiple AI methods were used. In these cases, it is explained in Section 4 whether XAI tasks were implemented for all of these multiple AI methods or for specific ones of them. This is also shown in Tables 4–6.

Research Question 6 was as follows: “Were single XAI tools or methods or a combination of them more preferred in XAI approaches in primary education?” The answer to this question is given in Table 3, which outlines the XAI method(s) or tool(s) used in each study. Section 4 analyzes the tasks implemented by XAI, and relevant aspects are outlined in Tables 4–6. Based on the aforementioned, it can be seen that most of the studies used a single XAI tool or method. Few studies explored the combination of multiple XAI tools or methods. Indicative such studies include those of Kobusingye et al. [27] and Lee [7].

Nevertheless, the combination of different XAI tools or methods may provide advantages. In fact, this research question stems from the first author’s previous work in the combination of intelligent methods (i.e., neuro-symbolic approaches [60] and other types of combinations [61]) which offers benefits by exploiting the advantages of the combined methods. Specifically, the combination of different XAI tools or methods may provide a variety of explanatory viewpoints. A typical example is the work of Lee [7], which combined the explanations provided by feature importance, partial dependence plots, and SHAP. The information provided by feature importance may be also provided by SHAP with the global bar plot. However, the information provided by partial dependence plots about how changes in the value of a feature affect the output value may be usefully blended with the information provided by SHAP plots. In the work of Kobusingye et al. [27], SHAP and BertViz were combined, providing different explanatory viewpoints once again. SHAP plots showed the contribution of input text features, and BertViz visualized the attention mechanism of the transformer models. In Kar et al. [25], LIME, SHAP, and FAMeX were used to compare their results in the identification of features contributing most to the output.

Research Question 7 concerned the following: “Taking into consideration the XAI tool or method mostly used in XAI approaches in primary education, which of the main functionalities offered were exploited?” As shown in Table 3 and already mentioned in Section 3, SHAP is the tool that is used more frequently compared to other XAI tools or methods. The functionality of SHAP used in these studies involves, among others, the following:

- (a) Calculation of SHAP values. This is performed in all studies using SHAP.
- (b) Generation of various types of plots (Table 9).
- (c) Depiction of feature contribution in producing the outputs. This is done in all studies using SHAP. SHAP values and the generated plots were used for this purpose.
- (d) Feature selection in order to improve the performance of the tested AI methods. This was done in the work of Gao et al. [23].
- (e) Provision of local and global explanations. Global explanations are provided with the global bar, beeswarm summary, collective force plot, and scatter plots. Local explanations are provided with waterfall and force (and individual) plots, and to a certain degree with the beeswarm summary plots. Local explanations are also considered to be the online feedback to teachers and students provided in the study of Kim et al. [26].

**Table 9.** Types of SHAP plots mentioned in the reviewed studies.

Citation	Global Bar Plot	Beeswarm Summary Plot	(Individual) Force Plot	Collective Force Plot	Waterfall Plot	Scatter Plot
[23]	✓	✓				
[25]	✓					
[27]			✓			
[7]		✓	✓			
[28]	✓	✓		✓		✓
[29]	✓	✓	✓			✓
[30]		✓			✓	
[5]		✓				✓
[34]	✓	✓				
[6]	✓				✓	
[38]	✓	✓				

An interesting aspect is to explore the reasons for which SHAP was the most used tool in the studies. A major reason for this is that SHAP conveniently generates various types of plots that may be interpreted. Another reason is the fact that SHAP provides both global and local explanations. In education, it is necessary to examine both types of explanations in order to explore the overall tendencies and parameters affecting specific outputs. For instance, it is useful to examine features affecting a specific decision for a specific student in a specific learning situation. However, it is also useful to examine how features generally affect all students or specific categories of students. The examination of features affecting specific categories of students yielded useful insights in certain studies.

As mentioned in Section 4, different types of SHAP plots were found to be useful in the studies. Table 9 depicts the types of SHAP plots mentioned in the reviewed studies. It should be noted that the contents of this table are an indication of researchers' tendencies in the types of SHAP plots that they use. However, it may be possible that further types of SHAP plots were used in certain studies but this was not mentioned in the corresponding papers (e.g., due to space limitations in the presentation of the work). One can note that the beeswarm summary plot is the type of plot used in most studies. This specific type of plot combines local and global explanation, as it shows results for every dataset instance. The second most used type of plot is the global bar plot. This plot provides convenient global explanation (similar to feature importance) by ordering the input features according to their mean absolute SHAP values in the whole dataset. The (individual) force and scatter plots share the position of the third most used type of plot. Scatter plots are useful to show dependencies. Individual force plots conveniently provide local explanations, showing the magnitude of the contribution of each feature.

## 6.2. Further Issues

There seems to be a tendency of using XAI to analyze results derived from large-scale educational assessments. The large amount of derived data enables the application of various AI methods. The analysis of the results may be conducted by the organizations or institutions that perform these assessments. However, individual researchers or groups of researchers may also analyze the results, because the relevant datasets are publicly

available. The availability of these large datasets creates opportunities to test various AI (and specifically XAI) methods.

A few studies concerned rural regions [27,30], underprivileged regions [30], or regions with limited resources [30]. An interesting aspect in AI-based learning is implementing approaches handling difficulties in such regions and providing solutions to certain problems that education faces. Such approaches demonstrate the social role that AI (and specifically XAI) may play in providing equal opportunities for learning.

The approach discussed by Alonso [22] may provide ideas about how to integrate XAI as a learning subject. The teaching of XAI concepts in combination with other subjects, such as programming, may prove beneficial. Another useful conclusion deriving from this work is the incorporation of XAI in popular learning tools used in education, adjusting the presentation of explanations in order to make them comprehensible to children. Comprehensible AI models may also assist in the generation of adjusted explanations.

It should be mentioned that interactions between humans and AI systems affect the users' level of trust in AI. In [62], it was demonstrated that VIRSTSI (Variability and Impact of Reciprocal Trust States towards Intelligent Systems) could effectively simulate trust transition in human–AI interactions. The model's ability to represent and adapt to individual trust dynamics suggests its potential for improving the design of AI systems that are more responsive to user trust levels. This approach captures how users' trust levels evolve in response to AI-generated outputs. As XAI increases trust in AI systems, it would be useful to employ ideas from VIRSTSI in order to assess the trust transition in human–XAI interactions in primary education.

An interesting aspect is to obtain measurable data about the impact of XAI on educational outcomes. For instance, these outcomes could involve reductions in workload, improvements in education due to policy changes, and improved student performance. Generally speaking, such data items are not yet available, and this represents a research gap. Obviously, information derived due to XAI could not have been obtained otherwise. It is useful to assess the impact of having such information.

As limitations of this research, one could mention the following: First, research items not authored in English were excluded. This is typical in reviews, for practical reasons, but it may exclude insightful work. Second, research items that were generally not accessible or were not accessible through our institution's infrastructure during the time period of the search were excluded. This is also typical in reviews, but there is always a possibility that relevant work may be excluded. Third, the term "XAI" is generally recently used in research due to the increased interest. It may be possible that certain research carried out in the context of XAI is not explicitly identified as such by the corresponding researchers. Fourth, the search was carried out using two tools (i.e., Scopus and Google Scholar). Further tools could be used to perform the search, such as Web of Science, which is not accessible by members of our institution due to the unavailability of the corresponding subscription.

## 7. Conclusions and Future Directions

This paper reviewed research integrating XAI in primary education. XAI holds incredible potential to transform education by making AI-driven decisions transparent, fair, and easier to trust. By helping educators and students understand how AI tools produce their output, XAI fosters greater confidence and enables more effective learning experiences. The categorization scheme proposed here discerns three main types of XAI approaches in primary education. It provides a structured and simple framework through which the application of XAI in primary education can be understood. This framework not only highlights the diverse roles that XAI can play in educational contexts but also offers a

foundation for future research and practice, guiding the development of more transparent, ethical, and effective AI solutions in primary education.

Studies from almost all continents have been presented here. Fifteen countries were involved in the studies. As the use of AI (and specifically XAI) in society and education increases, it is expected that more studies will be implemented in the specific countries and beyond. It should be noted that no studies from Oceania (e.g., Australia and New Zealand) were retrieved. It is expected that studies on this continent will be implemented in the near future.

XAI has been used in various learning subjects, as shown in Table 8. Further approaches using XAI in these subjects are likely to be implemented. A future direction that is likely to involve a lot of work involves first-language learning. This is evident when taking into consideration the needs in this subject and the work that has been done for second-language learning. It is likely that XAI will also be used in further subjects. Subjects for which XAI may provide benefits include science, geography, and history, because there is abundant learning material, and students face difficulties. Interactive learning tools encompassing XAI may be implemented for these subjects.

Relatively few studies have applied XAI in the context of interactive e-learning tool mechanisms used during learning and teaching. This is a promising direction for conducting research, because the number of e-learning tools developed for primary education is increasing. Furthermore, one study concerned the use of XAI to explain the decisions made offline and what was learned by the e-learning system. This is another promising research direction, due to the complexity of certain AI methods used in e-learning tools.

In [33], XAI was used to support the identification of educational items requiring revision. A trend in education is the implementation of Massive Open Online Courses (MOOCs). A future direction is to use XAI to support the creation and management of content in the context of MOOCs. MOOCs also involve the fostering of a learning community. XAI may assist in providing feedback to students, assisting them in learning collaboratively. Feedback is also useful to teachers to assist them in their interactions with students, instructional design, and acquisition of information about students' performance and interaction.

Table 2 shows unexplored directions (in italics) in tasks implemented by XAI. These unexplored directions will be briefly outlined here. First, approaches discussing lab activities using XAI seem to be missing. XAI has been used in the context of online tools. These activities were implemented through the Internet, but it would be useful to implement corresponding face-to-face lab activities. This would assist in viewing the reactions of students in their interaction with XAI. One could also mention the approach discussed by Alonso [22], but this has not been implemented in the context of school infrastructure. Second, approaches using XAI in lesson plan generation seem to be missing. AI tools may be used for this purpose [63], but teachers would like to receive explanations about the generated plans. Third, lesson plans about the teaching of XAI need to become available and disseminated to teachers.

Educational policies assisting in the integration of XAI in primary education are needed. Generally speaking, educational policies play an important role in the integration of technology in education. There seems to be a willingness of policymakers to integrate AI in primary education. This willingness needs to include XAI as well. This will be a first step for a general integration of XAI in primary education.

Relatively few studies involve the use of XAI in the context of AI as a learning subject. However, as the interest in teaching AI concepts to younger age groups increases, it is expected that the number of studies researching the teaching of XAI concepts in younger age groups will increase. This will involve face-to-face activities in classrooms and laboratories,

as well as Web-based learning tools. In this context, the design of AI-related curricula for K-12 education presented by Chiu [18] may be applied. The overall design was derived through interaction with teachers. The curriculum design includes two main themes: (a) curriculum as content and product, and (b) curriculum as process and praxis. The first main theme includes further themes such as knowledge in AI (i.e., what is AI, history of AI, and latest developments), how AI technologies work, and the impact of AI in society. The second main theme includes further themes, such as students' ways of learning AI, communication between students and teachers and between students and teaching material, and ways of addressing students' needs. It is also useful to employ comprehensible AI models in case their performance is good.

Combining AI with prosocial educational games can enhance the development of essential 21st-century skills, such as problem-solving, critical thinking, collaboration, and ethical reasoning [64]. XAI in combination with games could be employed in the context of interactive learning tools in primary education.

**Author Contributions:** Conceptualization, J.P.; methodology, J.P.; validation, J.P. and A.B.; data curation, J.P. and A.B.; writing—original draft preparation, J.P. and A.B.; writing—review and editing, J.P. and A.B.; visualization, J.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The search data are available upon request from the authors.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformer
CNN	Convolutional Neural Network
COCO	Common Objects in Context
CONFEMEN	Conference of Ministers of Education of French-Speaking States and Governments
CP	Conference Proceeding
FAMeX	FEature iMportance-Based eXplanable AI algorithm
FURIA	Fuzzy Unordered Rule Induction Algorithm
Grad-CAM	Gradient-Weighted Class Activation Map
GVTS	Global Variational Transformer Speaker Clustering
JP	Journal Publication
KEDI	Korean Educational Development Institute
KNN	K-Nearest Neighbors
KOFAC	Korean Foundation for the Advancement of Science and Creativity
LLM	Large Language Model
LightGBM	Light Gradient-Boosting Machine
LIME	Local Interpretable Model-Agnostic Explanations
LSTM	Long Short-Term Memory
MOOC	Massive Open Online Course
NLP	Natural Language Processing
NSO	National Statistical Office
PASEC	Program for the Analysis of CONFEMEN Education System
PIRLS	Progress in International Reading Literacy Study
PISA	Program for International Student Assessment
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine

TIMSS	Trends in International Mathematics and Science Study
UAE	United Arab Emirates
UNESCO	United Nations Educational, Scientific, and Cultural Organization
USA	United States of America
VIRTSI	Variability and Impact of Reciprocal Trust States towards Intelligent Systems
Weka	Waikato Environment for Knowledge Analysis
XAI	Explainable Artificial Intelligence
XGBoost	eXtreme Gradient Boosting

## References

- Chen, L.; Chen, P.; Lin, Z. Artificial intelligence in education: A review. *IEEE Access* **2020**, *8*, 75264–75278. [[CrossRef](#)]
- Bozkurt, A.; Karadeniz, A.; Baneres, D.; Guerrero-Roldán, A.E.; Rodríguez, M.E. Artificial intelligence and reflections from educational landscape: A review of AI Studies in half a century. *Sustainability* **2021**, *13*, 800. [[CrossRef](#)]
- Prentzas, J. Artificial Intelligence methods in early childhood education. In *Artificial Intelligence, Evolutionary Computing and Metaheuristics. In The Footsteps of Alan Turing; Studies in Computational, Intelligence*; Yang, X.S., Ed.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 427, pp. 169–199, ISBN 978-364-229-693-2. [[CrossRef](#)]
- Silva, M.; Ferreira, A.; Alves, K.; Valenca, G.; Brito, K. A new perspective for longitudinal measurement and analysis of public education in Brazil based on open data and machine learning. In Proceedings of the Seventeenth International Conference on Theory and Practice of Electronic Governance, Pretoria, South Africa, 1–4 October 2024; Chun, S., Karuri-Sebina, G., Przyebilowicz, E., Barbosa, F., Braga, C., Eds.; ACM: New York, NY, USA, 2024; pp. 130–138. [[CrossRef](#)]
- Nadaf, A.; Monroe, S.; Chandran, S.; Miao, X. Learning factors for TIMSS math performance evidenced through machine learning in the UAE. In *Artificial Intelligence in Education Technologies: New Development and Innovative Practices, Proceedings of the Third International Conference on Artificial Intelligence in Education Technology, Wuhan, China, 1–3 July 2022; Lecture Notes on Data Engineering and Communications Technologies*; Cheng, E.C.K., Wang, T., Schlippe, T., Beligiannis, G.N., Eds.; Springer: Berlin/Heidelberg, Germany, 2022; Volume 154, pp. 47–66. [[CrossRef](#)]
- Sanfo, J.B. Application of Explainable Artificial Intelligence approach to predict student learning outcomes. *J. Comput. Soc. Sci.* **2025**, *8*, 9. [[CrossRef](#)]
- Lee, Y. Applying Explainable Artificial Intelligence to develop a model for predicting the supply and demand of teachers by region. *J. Educ. e-Learn. Res.* **2021**, *8*, 198–205. [[CrossRef](#)]
- Ozmen, M.; Sahin, H. Real-time optimization of school bus routing problem in smart cities using genetic algorithm. In Proceedings of the 6th International Conference on Inventive Computation Technologies, Coimbatore, India, 20–22 January 2021; IEEE: New York, NY, USA, 2021; pp. 1152–1158. [[CrossRef](#)]
- Perikos, I.; Grivokostopoulou, F.; Hatzilygeroudis, I. Assistance and feedback mechanism in an Intelligent Tutoring System for teaching conversion of Natural Language into Logic. *Int. J. Artif. Intell. Educ.* **2017**, *27*, 475–514. [[CrossRef](#)]
- Chrysafiadi, K.; Virvou, M. Evaluating the integration of fuzzy logic into the student model of a web-based learning environment. *Expert Syst. Appl.* **2012**, *39*, 13127–13134. [[CrossRef](#)]
- Markos, A.; Prentzas, J.; Sidiropoulou, M. Pre-Service teachers' assessment of ChatGPT's utility in higher education: SWOT and content analysis. *Electronics* **2024**, *13*, 1985. [[CrossRef](#)]
- Hoffait, A.S.; Schyns, M. Early detection of university students with potential difficulties. *Decis. Support Syst.* **2017**, *101*, 1–11. [[CrossRef](#)]
- UNESCO Institute for Statistics. International Standard Classification of Education (ISCED) 2011. Available online: <https://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf> (accessed on 10 May 2025).
- Hatzilygeroudis, I.; Prentzas, J. HYMES: A HYbrid Modular Expert System with efficient inference and explanation. In Proceedings of the 8th Panhellenic Conference on Informatics, Nicosia, Cyprus, 8–10 November 2001; Manolopoulos, Y., Evripidou, S., Eds.; Livanis Publications: Athens, Greece, 2001; Volume 1, pp. 422–431.
- Hatzilygeroudis, I.; Prentzas, J. Symbolic-neural rule based reasoning and explanation. *Expert Syst. Appl.* **2015**, *42*, 4595–4609. [[CrossRef](#)]
- Saeed, W.; Omlin, C. Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowl.-Based Syst.* **2023**, *263*, 110273. [[CrossRef](#)]
- Hassija, V.; Chamola, V.; Mahapatra, A.; Singal, A.; Goel, D.; Huang, K.; Scardapane, S.; Spinelli, I.; Mahmud, M.; Hussain, A. Interpreting black-box models: A review on Explainable Artificial Intelligence. *Cogn. Comput.* **2024**, *16*, 45–74. [[CrossRef](#)]
- Chiu, T.K. A holistic approach to the design of Artificial Intelligence (AI) education for K-12 schools. *TechTrends* **2021**, *65*, 796–807. [[CrossRef](#)]

19. Khosravi, H.; Shum, S.B.; Chen, G.; Conati, C.; Tsai, Y.S.; Kay, J.; Knight, S.; Martinez-Maldonado, R.; Sadiq, S.; Gašević, D. Explainable artificial intelligence in education. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100074. [[CrossRef](#)]
20. Olsen, J.K.; Rummel, N.; Aleven, V. Designing for the co-orchestration of social transitions between individual, small-group and whole-class learning in the classroom. *Int. J. Artif. Intell. Educ.* **2021**, *31*, 24–56. [[CrossRef](#)]
21. Melsión, G.I.; Torre, I.; Vidal, E.; Leite, I. Using explainability to help children understand gender bias in AI. In Proceedings of the Twentieth Annual ACM Interaction Design and Children Conference, Athens, Greece, 24–30 June 2021; ACM: New York, NY, USA, 2021; pp. 87–99. [[CrossRef](#)]
22. Alonso, J.M. Explainable Artificial Intelligence for kids. In Proceedings of the 11th Conference of the European Society for Fuzzy Logic and Technology, Prague, Czech Republic, 9–13 September 2019; Atlantis Press: Dordrecht, The Netherlands, 2019; pp. 134–141. [[CrossRef](#)]
23. Gao, H.; Hasenbein, L.; Bozkir, E.; Göllner, R.; Kasneci, E. Exploring gender differences in computational thinking learning in a VR classroom: Developing machine learning models using eye-tracking data and explaining the models. *Int. J. Artif. Intell. Educ.* **2023**, *33*, 929–954. [[CrossRef](#)]
24. Kammüller, F.; Satija, D. Explanation of student attendance AI prediction with the Isabelle Infrastructure Framework. *Information* **2023**, *14*, 453. [[CrossRef](#)]
25. Kar, S.P.; Das, A.K.; Chatterjee, R.; Mandal, J.K. Assessment of learning parameters for students' adaptability in online education using machine learning and explainable AI. *Educ. Inf. Technol.* **2024**, *29*, 7553–7568. [[CrossRef](#)]
26. Kim, S.; Kim, W.; Jang, Y.; Choi, S.; Jung, H.; Kim, H. Student knowledge prediction for teacher-student interaction. In Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence, 2–9 February 2021; AAAI Press: Palo Alto, CA, USA, 2021; Volume 35, No. 17. pp. 15560–15568. [[CrossRef](#)]
27. Kobusingye, B.M.; Dorothy, A.; Nakatumba-Nabende, J.; Marvin, G. Explainable machine translation for intelligent e-learning of social studies. In Proceedings of the Seventh International Conference on Trends in Electronics and Informatics, Tirunelveli, India, 11–13 April 2023; IEEE: New York, NY, USA, 2023; pp. 1066–1072. [[CrossRef](#)]
28. McIntyre, N.A. Access to online learning: Machine learning analysis from a social justice perspective. *Educ. Inf. Technol.* **2023**, *28*, 3787–3832. [[CrossRef](#)]
29. Miao, X.; Mishra, P.K.; Nadaf, A. Evidence and promises of AI predictions to understand student approaches to math learning in Abu Dhabi K12 public schools. *Gulf Educ. Soc. Policy Rev.* **2021**, *1*, 109–134. [[CrossRef](#)]
30. Mohan, A.; Mathew, M.; Malavika, K.; Gutjahr, G.; Menon, R.; Nedungadi, P. An Explainable AI model for student engagement in tablet learning. In Proceedings of the Sixth IEEE International Conference on Recent Advances in Intelligent Computational Systems, Kothamangalam, Kerala, India, 16–18 May 2024; IEEE: New York, NY, USA, 2024. [[CrossRef](#)]
31. Onishi, S.; Yasumori, T.; Shiina, H. Visualization of the impact of classroom utterances using generative dialogue models. *IIAI Lett. Inform. Interdiscip. Res.* **2023**, *4*, LIIR181. [[CrossRef](#)]
32. Onishi, S.; Kojima, S.; Shiina, H.; Yasumori, T. Estimating the impact of classroom speech using a Large Language Model. In Proceedings of the Sixteenth IIAI International Congress on Advanced Applied Informatics, Takamatsu, Japan, 6–12 July 2024; IEEE: New York, NY, USA, 2024; pp. 409–414. [[CrossRef](#)]
33. Pelánek, R.; Effenberger, T.; Kukučka, A. Towards design-loop adaptivity: Identifying items for revision. *J. Educ. Data Min.* **2022**, *14*, 1–25. [[CrossRef](#)]
34. Ribeiro-Flucht, L.; Chen, X.; Meurers, D. Explainable AI in language learning: Linking empirical evidence and theoretical concepts in proficiency and readability modeling of Portuguese. In Proceedings of the Nineteenth Workshop on Innovative Use of NLP for Building Educational Applications, Mexico City, Mexico, 20 June 2024; Kochmar, E., Bexte, M., Burstein, J., Horbach, A., Laarmann-Quante, R., Tack, A., Yaneva, V., Yuan, Z., Eds.; Association for Computational Linguistics: Kerrville, TX, USA, 2024; pp. 199–209.
35. Ruan, S.; Nie, A.; Steenbergen, W.; He, J.; Zhang, J.Q.; Guo, M.; Liu, Y.; Nguyen, D.K.; Wang, C.Y.; Ying, R.; et al. Reinforcement learning tutor better supported lower performers in a math task. *Mach. Learn.* **2024**, *113*, 3023–3048. [[CrossRef](#)]
36. Tsiakas, K.; Barakova, E.; Khan, J.V.; Markopoulos, P. BrainHood: Towards an explainable recommendation system for self-regulated cognitive training in children. In Proceedings of the 13th ACM International Conference on PErvasive Technologies related to Assistive Environments, Corfu, Greece, 30 June–3 July 2020; ACM: New York, NY, USA, 2020. [[CrossRef](#)]
37. Tsiakas, K.; Barakova, E.; Khan, J.V.; Markopoulos, P. BrainHood: Designing a cognitive training system that supports self-regulated learning skills in children. *Technol. Disabil.* **2020**, *32*, 219–228. [[CrossRef](#)]
38. Zhen, Y.; Luo, J.D.; Chen, H. Prediction of academic performance of students in online live classroom interactions—An analysis using natural language processing and deep learning methods. *J. Soc. Comput.* **2023**, *4*, 12–29. [[CrossRef](#)]
39. Lundberg, S.; Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems, Proceedings of the Thirty-First International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017*; Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R., Eds.; Curran Associates: Red Hook, NY, USA, 2017; Volume 30, pp. 4765–4774.

40. Lundberg, S. SHAP Documentation. Available online: <https://shap.readthedocs.io/en/latest/index.html> (accessed on 15 January 2025).
41. Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic attribution for deep networks. In *Proceedings of Machine Learning Research, Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017*; Precup, D., Teh, Y.W., Eds.; MLR.org; 2017; Volume 70, pp. 3319–3328.
42. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations From Deep Networks via Gradient-Based Localization. In *Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017*; IEEE: New York, NY, USA, 2017; pp. 618–626. [CrossRef]
43. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
44. Vig, J. BertViz: A tool for visualizing multihead self-attention in the BERT model. In *Proceedings of the International Conference on Learning Representations Workshop: Debugging Machine Learning Models*, New Orleans, LA, USA, 6 May 2019; Volume 3.
45. Alonso, J.M.; Bugarín, A. ExpliClas: Automatic generation of explanations in natural language for Weka classifiers. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, New Orleans, LA, USA, 23–26 June 2019; IEEE: New York, NY, USA, 2019. [CrossRef]
46. Hühn, J.; Hüllermeier, E. FURIA: An algorithm for unordered fuzzy rule induction. *Data Min. Knowl. Disc.* **2009**, *19*, 293–319. [CrossRef]
47. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why should i trust you?” Explaining the predictions of any classifier. In *Proceedings of the Twenty-Second ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 1135–1144. [CrossRef]
48. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794. [CrossRef]
49. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems, Proceedings of the Thirty-First International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017*; Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates: Red Hook, NY, USA, 2017; Volume 30, pp. 3146–3154.
50. Piech, C.; Bassen, J.; Huang, J.; Ganguli, S.; Sahami, M.; Guibas, L.J.; Sohl-Dickstein, J. Deep knowledge tracing. In *Advances in Neural Information Processing Systems, Proceedings of the Twenty-Ninth International Conference on Neural Information Processing Systems, Montreal, Canada, 7–12 December 2015*; Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R., Eds.; MIT Press: Cambridge, MA, USA, 2015; Volume 28, pp. 505–513.
51. Junczys-Dowmunt, M.; Grundkiewicz, R.; Dwojak, T.; Hoang, H.; Heafield, K.; Neckermann, T.; Seide, F.; Germann, U.; Aji, A.F.; Bogoychev, N.; et al. Marian: Fast Neural Machine Translation in C++. *arXiv* **2018**, arXiv:1804.00344.
52. Keller, J.M. Development and use of the ARCS model of instructional design. *J. Instr. Dev.* **1987**, *10*, 2–10. [CrossRef]
53. Liu, F.T.; Ting, K.M.; Zhou, Z.-H. Isolation-based anomaly detection. *ACM Trans. Knowl. Discov. Data* **2012**, *6*, 1–39. [CrossRef]
54. Gupta, N.; Eswaran, D.; Shah, N.; Akoglu, L.; Faloutsos, C. Beyond outlier detection: LOOKOUT for pictorial explanation. In *Machine Learning and Knowledge Discovery in Databases, Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, Dublin, Ireland, 10–14 September 2018*; Lecture Notes in Computer Science; Berlingerio, M., Bonchi, F., Gärtner, T., Hurley, N., Ifrim, G., Eds.; Springer: Cham, Switzerland, 2019; Volume 11051, pp. 122–138. [CrossRef]
55. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347.
56. Bull, S.; Kay, J. Open Learner Models. In *Advances in Intelligent Tutoring Systems; Studies in Computational Intelligence*; Nkambou, R., Bourdeau, J., Mizoguchi, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 308, pp. 301–322. [CrossRef]
57. Clarke, E.M.; Emerson, E.A. Design and synthesis of synchronization skeletons using branching time temporal logic. In *Logic of Programs*; Lecture Notes in Computer Science; Kozen, D., Ed.; Springer: Berlin/Heidelberg, Germany, 1982; Volume 131, pp. 52–71. [CrossRef]
58. Clarke, E.M. The birth of model checking. In *25 Years of Model Checking*; Lecture Notes in Computer Science; Grumberg, O., Veith, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; Volume 5000, pp. 1–26. [CrossRef]
59. Prokhorenkova, L.O.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems, Proceedings of the Thirty-Second International Conference of the Neural Information Processing Systems, Montréal Canada, 3–8 December 2018*; Bengio, S., Wallach, H.M., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates: Red Hook, NY, USA, 2018; Volume 31, pp. 6639–6649.
60. Hatzilygeroudis, I.; Prentzas, J. Integrated rule-based learning and inference. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1549–1562. [CrossRef]
61. Prentzas, J.; Hatzilygeroudis, I. Assessment of life insurance applications: An approach integrating neuro-symbolic rule-based with case-based reasoning. *Expert Syst.* **2016**, *33*, 145–160. [CrossRef]

62. Virvou, M.; Tsirhrintzis, G.A.; Tsichrintzi, E.-A. VIRTSI: A novel trust dynamics model enhancing Artificial Intelligence collaboration with human users—Insights from a ChatGPT evaluation study. *Inform. Sciences* **2024**, *675*, 120759. [[CrossRef](#)]
63. Prentzas, J.; Sidiropoulou, M. Integrating OpenAI Chat-GPT in a University Department of Education: Main types of use and preliminary assessment results. In *Extended Selected Papers of the 14th International Conference on Information, Intelligence, Systems, and Applications*; Lecture Notes in Networks and Systems; Bourbakis, N., Tsirhrintzis, G.A., Virvou, M., Jain, L.C., Eds.; Springer: Cham, Switzerland, 2024; Volume 1093, pp. 306–326, ISBN 978-3-031-67425-9. [[CrossRef](#)]
64. Papadimitriou, S.; Virvou, M. *Artificial Intelligence-Based Games as Novel Holistic Educational Environments to Teach 21st Century Skills, Intelligent Systems Reference Library*; Springer: Cham, Switzerland, 2025; Volume 93, ISBN 978-3-031-77463-8.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.