*technologies*

MDPI

*Systematic Review*

# A Systematic Literature Review of the Latest Advancements in XAI

**Zaid M. Altukhi [1,2,*], Sojen Pradhan [1] and Nasser Aljohani [2,*]**

[1] Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia; sojen.pradhan@uts.edu.au

[2] Faculty of Computer and Information Systems, Islamic University of Madinah, Al Jamiah, Madinah 42351, Saudi Arabia

* Correspondence: zaid.m.altukhi@student.uts.edu.au or zaid@iu.edu.sa (Z.A.); naljohani@iu.edu.sa (N.A.)

**Abstract:** This systematic review details recent advancements in the field of Explainable Artificial Intelligence (XAI) from 2014 to 2024. XAI utilises a wide range of frameworks, techniques, and methods used to interpret machine learning (ML) black-box models. We aim to understand the technical advancements in the field and future directions. We followed the PRISMA methodology and selected 30 relevant publications from three main databases: IEEE Xplore, ACM, and ScienceDirect. Through comprehensive thematic analysis, we categorised the research into three main topics: 'model developments', 'evaluation metrics and methods', and 'user-centred and XAI system design'. Our results uncover 'What', 'How', and 'Why' these advancements were developed. We found that 13 papers focused on model developments, 8 studies focused on the XAI evaluation metrics, and 12 papers focused on user-centred and XAI system design. Moreover, it was found that these advancements aimed to bridge the gap between technical model outputs and user understanding.

**Keywords:** XAI; explainable AI; black box; framework; methodology; PRISMA

check for updates

Academic Editor: Lipo Wang

Received: 7 October 2024
Revised: 19 February 2025
Accepted: 25 February 2025
Published: 1 March 2025

**Citation:** Altukhi, Z.M.; Pradhan, S.; Aljohani, N. A Systematic Literature Review of the Latest Advancements in XAI. *Technologies* **2025**, *13*, 93. https://doi.org/10.3390/technologies13030093

## 1. Introduction and Background

In recent years, the field of Explainable Artificial Intelligence (XAI) has garnered significant attention due to its crucial role in interpreting the opaqueness of machine learning (ML) black-box models. As AI systems become increasingly widespread, understanding how these systems make decisions is essential for fostering trust [1] and ensuring ethical considerations [2]. Users often struggle to comprehend AI reasoning and output, particularly when the underlying algorithms and logic are hidden and treated as a "black box", presenting challenges of explainability [3]. The lack of explainability and transparency in these algorithms can extend issues of justice and bias, ultimately reducing user acceptance and satisfaction.

The XAI is described as a domain within AI, which is dedicated to developing tools, techniques, and algorithms capable of producing high-quality, interpretable, intuitive, and human-understandable explanations of AI decisions [4]. Das et al. [4] assert that XAI encompasses methods and algorithms intended to enhance the trustworthiness and transparency of AI systems. XAI seeks to clarify the internal logic and predictions of complex ML models [5].

ML models are generally categorised into two types: white-box and black-box models [6,7]. White-box models are transparent, and they allow most data scientists and domain experts to understand how inputs are transformed into outputs, making it easier to explain the reasoning behind predictions, although they typically generate less accurate results [6].

In contrast, black-box models produce highly accurate results but are opaque, making it challenging to comprehend their inner workings and the rationale behind their predictions or decisions [6,7].

When the black box is used in AI systems, XAI methods can be used to explain black-box behaviour and decisions [8]. XAI methods address the question of why ML black-box algorithms predict specific outcomes. These methods are primarily divided into model-specific and model-agnostic categories. Model-specific methods are tailored to particular types of models, leveraging their internal structures [9]. whereas model-agnostic methods can be applied to any model, treating them as black boxes.

Among the common XAI methods, SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) are widely used to interpret black-box outputs. SHAP can provide explanations of model predictions, making it versatile for understanding overall model behaviour and specific outcomes [10]. Whereas, LIME focuses on providing explanations for individual predictions by perturbing the input data and observing the changes in predictions [11]. Both SHAP and LIME can be applied to any ML model, making them flexible and widely applicable [12,13]. These techniques facilitate a better understanding of the underlying mechanisms driving the predictions of black-box models, thus enhancing transparency and trust in AI systems.

Many AI frameworks position XAI in the final stage of their processes, as shown in Figure 1. They start by collecting data, training and testing these data using ML models, and then applying XAI techniques to generate explanations or interpretations of the ML model outputs [9,14].

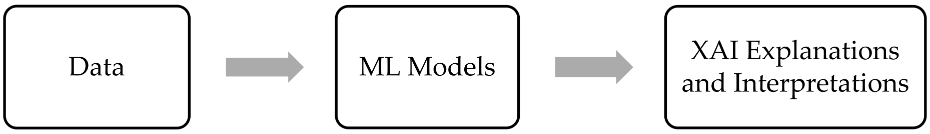

**Figure 1.** XAI framework process flow chart. Adapted from [9].

### *1.1. XAI Stakeholders*

Stakeholders in XAI can include AI regulators, developers, managers, users, individuals affected by AI decisions, domain experts, service providers, individuals affected by AI decisions, and customers across different systems [15,16]. Effective communication and collaboration between XAI end-users and developers is crucial for developing successful XAI systems. Both stakeholder groups have distinct needs and perspectives that must be considered to create a useful and trustworthy XAI solution.

End-users are the individuals who interact with the AI system and rely on its outputs for decision-making. They can be categorised into non-technical users and domain experts. Non-technical users are those such as laypersons, customers, and professionals from various fields who may not have a deep understanding of AI. Their primary need is to understand the AI's decisions in a straightforward and intuitive manner [17,18]. Other users, such as domain experts, are professionals with expertise in specific fields, such as educators, doctors, financial analysts, or forensic experts, who use AI to augment their decision-making processes. They require explanations that are detailed and relevant to their domain to increase trustworthiness and effectively utilise the AI system [19].

Developers are the individuals and teams involved in designing, developing, and maintaining AI systems, including AI developers and engineers, designers, UX specialists, ethicists, and compliance officers. AI developers and engineers are the technical experts who build the AI models [17] and ensure that the models are not only accurate but also interpretable and explainable. They often focus on algorithmic transparency and the inner workings of the models. Within the development team, the designers and UX specialists are the professionals who are responsible for creating user interfaces that present the AI's explanations in a user-friendly manner [20]. Their goal is to bridge the gap between complex AI outputs and user comprehension. Ethicists and compliance officers who ensure that the AI systems adhere to ethical standards and regulatory requirements. Utilising XAU systems allows them to verify that the AI system operates within legal and ethical matters [21].

### *1.2. XAI Explanation Types*

There are three common ways to generate explanations in the context of XAI: contrastive explanations, counterfactual explanations, and natural language explanations. These are different approaches to make AI-based systems more transparent and understandable.

'Contrastive explanations' highlight the key features that differentiate the predicted output from an alternative output or reference point. They aim to answer questions like "Why was this instance classified as X instead of Y?" by identifying the most relevant features that led to the specific prediction [22].

'Counterfactual explanations' provide insights into how and why an AI model made a certain decision by illustrating what changes to the input features could alter the outcome. They present a scenario that differs minimally from the current scenario but leads to a different outcome [23].

'Natural language explanations' in the context of XAI refer to using human language to explain the decisions and workings of AI models. This approach aims to make the AI's decision-making process more understandable and accessible to people, particularly to those without a technical background [24].

### *1.3. Purpose of the Study*

This review investigates the latest advancements in XAI, focusing on frameworks and methodologies to understand the contributions of recent research, particularly in the education sector. However, due to the limited number of studies specifically addressing XAI advancements in education, we broadened our scope to include general XAI research while incorporating a few relevant studies in education. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [25] methodology to collect recent XAI studies and employed thematic analysis to categorise emerging topics of XAI studies as advancements by identifying key research directions. Finally, we analysed the selected articles from three perspectives: what each study investigated, how it was conducted, and why it was developed.

## 2. Methodology

PRISMA systematic review methodology was employed to fulfil the study's objectives and address the research questions. PRISMA principles [25] were utilised to establish a priori framework for conducting a rigorous systematic review. Additionally, these principles guided the processes of searching for identifying and selecting articles for inclusion in the research. Adopting a systematic approach is crucial for minimising bias and creating a transparent, reproducible work for analysis. This review followed PRISMA guidelines [26] as follows:

### *2.1. Phase1: Identification of Relevant Studies (Identification)*

#### 2.1.1. Definition of the Research Questions

This section explores the latest advancements in XAI by addressing two key research questions (RQ). These questions are designed to uncover the approaches shaping the current landscape of XAI, the innovative achievements that have been made, the novel findings resulting from these advancements, and the challenges these methodologies and frameworks have encountered.

RQ1. What are the latest XAI advancements developed focused on?

This question seeks to identify and describe the most recent methodologies and frameworks developed in the field of XAI. Classifying these approaches into several categories allows us to explore the key aspects of these innovations.

RQ2. What are these advancements, and how and why have they been developed?

This question discusses the advancements found in RQ1 from three perspectives, providing a high-level understanding of their purpose and the methods used to achieve their objectives. It explores what the innovations are, why they were created, and how they were developed.

#### 2.1.2. Eligibility Criteria

Table 1 shows the inclusion and exclusion criteria followed to conduct this review.

**Table 1.** Inclusion and exclusion criteria.

| **Inclusion Criteria (IC)** | **Exclusion Criteria (EXC)** |
|---|---|
| IC-1: Peer-reviewed paper: Published in journals or conferences with peer reviews before January 2024. | EXC-1: Not Peer-Reviewed: Publications that are not peer-reviewed (e.g., white papers, opinion articles, or blog posts) or published outside the specified time range (before January 2014 or after January 2024). |
| IC-2: Primary study: Papers that have directly contributed to XAI advancements and XAI in education. | EXC-2: Secondary Studies: Papers that are secondary studies, such as surveys, systematic literature reviews, theoretical papers, or literature reviews, as well as papers not directly contributing to XAI achievements or limited to a specific domain (healthcare, cyber security, energy, etc.) |
| IC-3: Studies using empirical data. Research methods, which include quantitative, qualitative, mixed-method studies, case studies, and experimental designs. | EXC-3: Focused Solely on Images or Spatial Data: Frameworks and methodologies that exclusively focus on images, spatial data, or any domain not relevant to general XAI applications are excluded. |
| IC-4: The article is published in English. | EXC-4: Articles not written or published in English.<br>EXC-5: Papers that do not directly relate to XAI or that do not contribute to the goals of explainability or interpretability in AI. |

#### 2.1.3. Information Searching Period

A comprehensive search strategy was employed to locate publications within three primary academic databases shown in Figure 2 for articles published from January 2014 to January 2024.

### *2.2. Phase 2: Selection of Relevant Studies (Screening)*

#### 2.2.1. Search

Specific search terms were employed, incorporating combinations like ‘explainable AI’ OR ‘XAI’ AND (‘methodology’ OR ‘framework’ OR ‘novel’). These keywords, besides the logical operators compatible with the chosen database search functions, were systematically applied to identify relevant studies.

#### 2.2.2. Title and Abstract Exploration

Initially, a total of 1649 articles were retrieved. To assess the initial relevance of the search results, a detailed review article information was conducted. This step involved screening the article titles and abstracts and running them through Python 3.11.10 code to eliminate any publication whose title or abstract that did not contain one of the following statements: 'XAI' OR 'Explainable AI' OR 'Explain*'. Then, titles were first carefully examined, and those focusing on domains outside of education were immediately disregarded.

#### 2.2.3. Potentially Relevant Studies Selection

The studies selected during the previous phase were thoroughly evaluated to determine their alignment with the established inclusion criteria. Particular attention was given to whether they focused on new methodologies and frameworks. We found that there were 218 studies that did not provide any novelty findings or advancements. Studies that fulfilled these criteria advanced to the next stage of the review.

### *2.3. Phase 3: Study Inclusion (Inclusion)*

During this phase, potentially relevant studies were thoroughly reviewed. Titles, abstracts, and conclusions were rigorously evaluated against the inclusion and exclusion criteria. Furthermore, a table was created to organise the included papers, identifying the key characteristics that justified their inclusion. This table also served to evaluate each study's methodology and main contributions. In the process of using PRISMA for selecting articles, we did not identify any potential biases.

As previously discussed, the article selection process was divided into three key stages. Figure 2 presents the distribution of the publications over different selected databases and the numbers of publications in each phase.

Table 2 shows the number of articles included to this review from their database. The Figure 3 provides a visual representation of how PRISMA stages were applied throughout the study selection process.

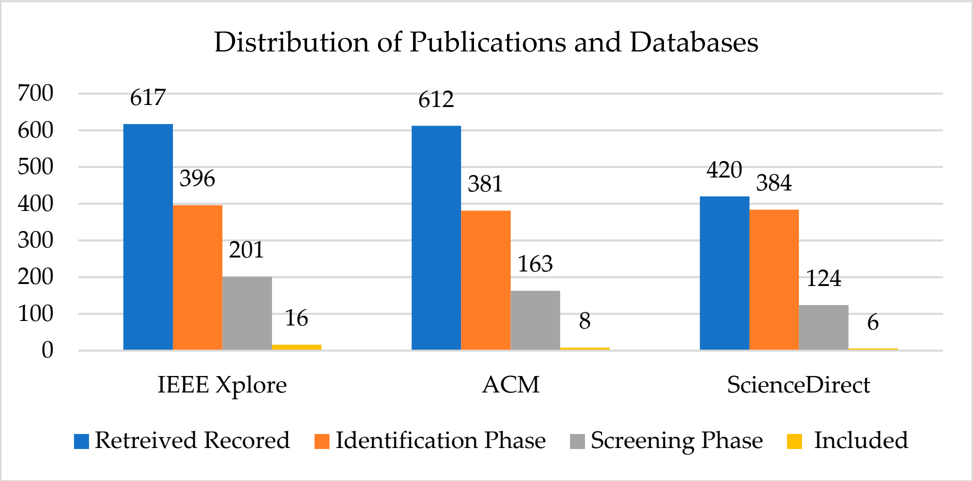

**Figure 2.** Distribution of publications and databases.

**Table 2.** Distribution of publications and databases used.

| Database | Number of Papers Remaining After Meeting the Inclusion and Exclusion Criteria |
|---|---|
| IEEE Xplore | 16 |
| ACM | 8 |
| ScienceDirect | 6 |
| **Total** | **30** |

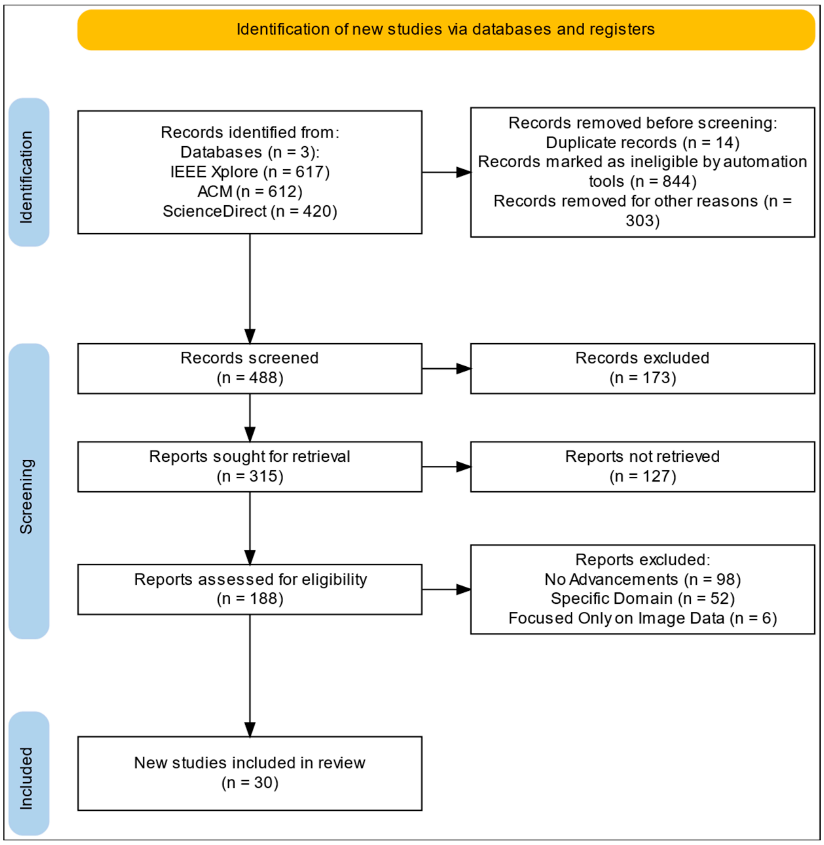
Identification of new studies via databases and registers
Identification
Records identified from: Databases (n = 3): IEEE Xplore (n = 617) ACM (n = 612) ScienceDirect (n = 420)
Records removed before screening: Duplicate records (n = 14) Records marked as ineligible by automation tools (n = 844) Records removed for other reasons (n = 303)
Screening
Records screened (n = 488)
Records excluded (n = 173)
Reports sought for retrieval (n = 315)
Reports not retrieved (n = 127)
Reports assessed for eligibility (n = 188)
Reports excluded: No Advancements (n = 98) Specific Domain (n = 52) Focused Only on Image Data (n = 6)
Included
New studies included in review (n = 30)

**Figure 3.** PRISMA flow diagram of paper selection. Created using [27].

### 2.4. *Categorising Advancements Using Thematic Analysis*

We used the thematic analysis framework proposed by Braun et al. [28] to categorise the selected papers into different categories to explore the latest advances in XAI, which answer the first research question.

In order to answer RQ2, we followed the research methodology established by Arrieta et al. [29] to discuss the results in answering the what, how, and why aspects of their studies from the selected articles.

## 3. Results

The selected papers explore various frameworks and methodologies applicable to different topics. For the purpose of this study, we utilise the thematic analysis framework proposed by Braun et al. [28] to categorise the relevant data from articles systematically.

### 3.1. *Advancements Categorisation Process*

The process of categorising the results aims to answer RQ1. This categorisation reveals the areas that the selected papers focus on, providing insights into the aspects that the XAI research community emphasises.

#### 3.1.1. Data Familiarisation

The initial phase of Braun and Clarke's framework [28] intends to familiarise us with the data collected by examining the selected papers to comprehend the context and relevance to our objectives. We extracted the purpose of the latest advancements related to XAI from these publications. Then, these data were arranged into a matrix in an MS Excel spreadsheet to discern thematic characteristics for further analysis in the next step.

### 3.1.2. Generating Initial Codes

In the second phase, we highlighted key terms based on the aims and objectives of each selected paper. These key terms were used to categorise their specified advancements. Codes are labelled as shown in Table 3 (ADV1, ADV2, ADV3) to distinguish each advancement.

**Table 3.** Initial codes of the categorisation.

| Code | Key Terms |
|---|---|
| ADV1 | tools, approaches, features, and/or techniques, to improve or enhance XAI models accuracy, performance, and/or efficient |
| ADV2 | Evaluation, metric, assessment, measurement, comparison, accuracy |
| ADV3 | User experience, interface, guides, practitioners, experts, stakeholders, design, |

### 3.1.3. Generating the Theme

The selected publications were categorised into the codes identified in Phase 2. Papers on developing models or improving the prediction of explainability were assigned to the ADV1 category. Those that concentrated on evaluating or proposing new metrics or comparing different models were classified under ADV2. Lastly, papers emphasising user experience, system design, and interface considerations were allocated to ADV3.

### 3.1.4. Reviewing Potential Theme

In this phase, previously established themes are reviewed and altered. We checked if the original themes adequately represent the important features of the data. The scope of each topic is examined for coherence and consistency within the dataset. There are two levels involved. Firstly, the themes were examined to ensure internal homogeneity at the level of the coded data extraction. The complete dataset was examined to ensure the themes appropriately capture the context and significance of the data. Secondly, themes were changed or combined to resolve any inconsistencies or ambiguities, guaranteeing an accurate and transparent depiction of the facts.

### 3.1.5. Naming Categories

After refining the themes, the subsequent step involves clearly defining and naming the categories. Each category name should encapsulate the essence of its respective theme and briefly convey the main idea. This process includes writing detailed descriptions for each category, specifying its scope and limitations. The category names are crafted to be concise yet comprehensive, capturing the essential aspects of the identified novelties.

ADV1, named "Models Development", encompasses advancements focusing on methods to enhance explainability and improve accuracy, performance, or efficiency by analysing input feature weights or employing other relevant techniques. ADV2, termed "Evaluation Metrics and Methods", pertains to studies that compare different models, evaluate various explanations, or develop new metrics for assessing the explainability, accuracy, or performance of generated explanations. Lastly, ADV3, titled "User-Centred and XAI Systems Design", includes papers concentrating on stakeholders in XAI, guidelines, and systematic approaches for developing XAI systems.

These names are intended to provide clear, intuitive labels that facilitate the understanding of the themes. Although some frameworks and methodologies belong to multiple categories, they are inter-connected.

### 3.1.6. Categories Result

The results of the thematic analysis indicate that the categories of Models Development (ADV1), Evaluation Metrics and Methods (ADV2), and User-Centred Design in XAI Systems (ADV3) have garnered nearly equal interest in the context of XAI advancements, as shown in Figure 4.

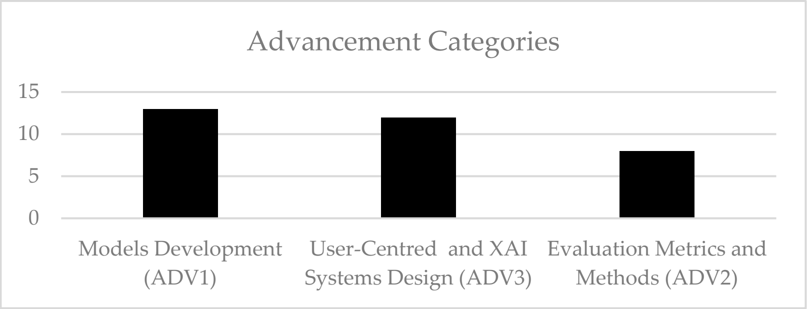

**Figure 4.** Number of articles within each category.

For further clarification, we established sub-categories under each of these main categories based on the specific research purposes of the selected articles.

The main categories and sub-headings, along with the number of articles contributing to each category, appear in Figure 4. The sub-headings are named based on the purpose of the advancement, which is mentioned explicitly in the article. Notably, some articles address multiple categories, reflecting the interdisciplinary nature of the research as shown in Figure 5.

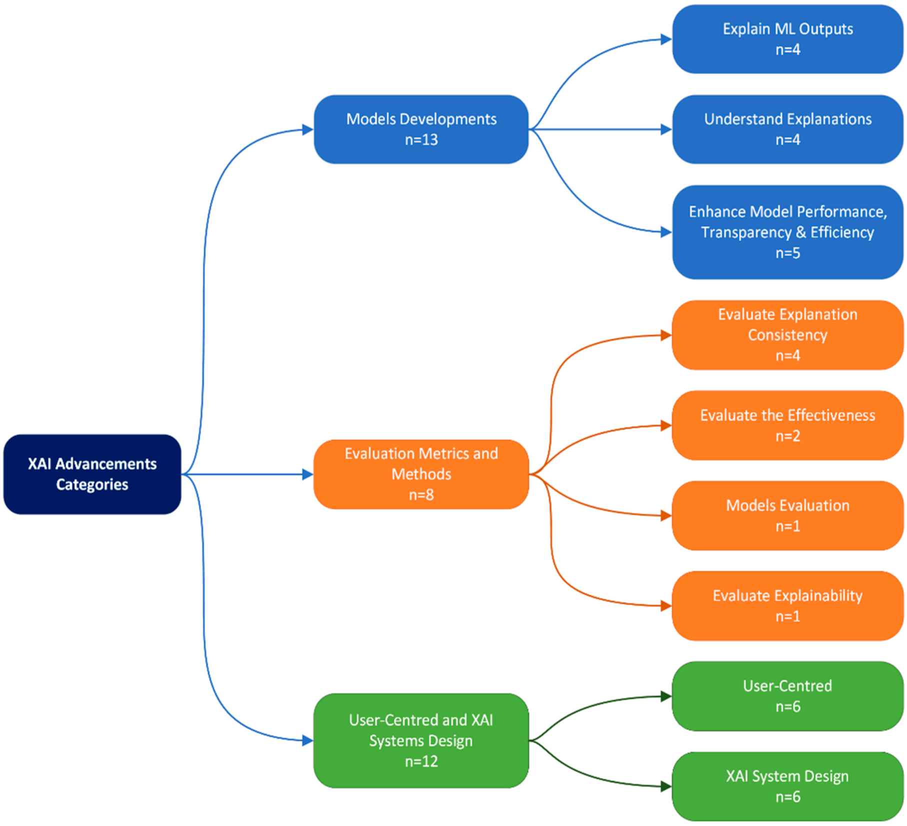
XAI Advancements Categories
Models Developments n=13
Explain ML Outputs n=4
Understand Explanations n=4
Enhance Model Performance, Transparency & Efficiency n=5
Evaluation Metrics and Methods n=8
Evaluate Explanation Consistency n=4
Evaluate the Effectiveness n=2
Models Evaluation n=1
Evaluate Explainability n=1
User-Centred and XAI Systems Design n=12
User-Centred n=6
XAI System Design n=6

**Figure 5.** XAI advancements categories across the 30 articles (five studies cover more than one category, as shown in Table 4).

**Table 4.** Details of selected publications.

| Main Category | Sub-Category | Reference |
|---|---|---|
| Model Developments | Explain ML outputs<br>Understand Explanations<br>Enhance Model Performance, Transparency, and Efficiency | [30–33]<br>[34–37]<br>[14,38–41] |
| Evaluation Metrics and Methods | Evaluate Explanation Consistency<br>Evaluate the Effectiveness<br>Models Evaluation<br>Evaluate Explainability | [9,42–44]<br>[45,46]<br>[31]<br>[47] |
| User-Centred and XAI Systems Design | User-Centred<br>XAI System Design | [15,33,36,48–50]<br>[43,51–55] |

### *3.2. Models Development (ADV1)*

This category encompasses papers proposing methodologies or frameworks to enhance XAI outputs. These approaches offer the necessary structure and techniques for developing, implementing, and improving XAI technologies. Innovations in this section have introduced novel methods to improve the interpretability of ML black box outputs, optimise the understanding explanations, and enhance model performance, transparency, and efficiency.

#### 3.2.1. Explain ML Outputs

The frameworks and methodologies in this section provide insights into why a particular model produced a specific result or prediction. Researchers utilise different approaches to explain ML black-box algorithms outputs such as, analysing the relationships between different features [30–32], simplifying the output of complex models using interpretable models [33], and understanding the cluster models' behaviours [32]. Table 5 shows the advancement types, techniques and main findings that found in the selected paper for explaining ML outputs.

**Table 5.** Advancements in explaining ML outputs.

| Advancement Type | Advancement Technique | Findings | Ref. |
|---|---|---|---|
| Arg-XAI framework | Represent and analyse logical relationships between features to generate human-readable explanations. | The framework could effectively clarify ML model outcomes and provided transparent explanations, aligning with the results from traditional interpretable models. | [30] |
| Local eXplanation of Dimensionality Reduction (LXDR) methodology | Represent interpretability for dimensionality reduction (DR) techniques. It addresses the critical gap in understanding non-linear and model-agnostic DR processes. | This development transforms non-interpretable DR techniques into more transparent ones. | [31] |
| Kauffmann et al. Framework | Trace the influence of input features on cluster assignments. | The framework significantly improves the interpretability of clustering decisions and provides clear and comprehensible explanations for why certain data points are grouped in clustering models. | [32] |

**Table 5.** *Cont.*

| Advancement Type | Advancement Technique | Findings | Ref. |
|---|---|---|---|
| Contributions Oriented Local Explanations (COLE) Framework | Simplify the outputs of complex black-box models such as Convolutional Neural Networks (CNNs), by using simpler and transparent models like k-Nearest Neighbours (k-NN). | This approach leverages the simpler model's more comprehensible logic to explain the complex model's decision-making process. | [33] |
| COLE-HP methodology | Extracting feature weights from CNNs and applying them to a k-NN model precisely and computationally efficiently. | This methodology enhances the twin-systems framework by providing a clear, implementable method for generating explanations. | [33] |

### 3.2.2. Understand Explanations

Although the main reason for utilising XAI is to generate explanations of the complex black-box model outputs, some explanations are difficult to interpret or understand. Researchers use different ways to clarify the explanations by providing contrastive, counterfactual explanations and natural language. Other researchers use a natural language approach to make the outputs readable [34,37]. For more clarification, Table 6 shows the advancements types, methods, and key findings in improving the understanding of XAI outputs.

**Table 6.** Advancements in understanding the XAI explanations.

| Advancement Type | Advancement Technique | Findings | Ref. |
|---|---|---|---|
| MERLIN Methodology | Generate contrastive explanations for two ML models, Naïve Bayes and Random Forest. | MERLIN is useful to understand how two models can have similar predictive accuracy but different underlying logic. | [34] |
| Methodology | Use contrastive and natural language to provide explanations through indicators and natural language. | The methodology shows high consistency in the fidelity of the surrogate, demonstrating its effectiveness in capturing the logic of the underlying model. | [34] |
| Learning to Counter (L2C) Framework | Generates counterfactual explanations by achieving high levels of diversity and validity. | The framework effectively produces counterfactual explanations that meet key criteria. | [35] |
| LEWIS Framework | Employ both contrastive and counterfactual explanations to elucidate model decisions by addressing how a model's decision could have differed and why a specific decision was made over another. | This framework is compatible with any decision-making algorithm and provides clear, useful explanations and produces more insightful and actionable explanations than state-of-the-art methods such as LIME and SHAP. | [36] |
| INTERACTION Framework | Generating natural language explanations for AI decisions, specifically for natural language inference (NLI) tasks. | The framework improves the interpretability and transparency of AI decision-making by providing diverse explanations for NLI tasks. | [37] |

#### 3.2.3. Enhance Model Performance, Transparency, and Efficiency

This section presents innovative approaches designed to enhance AI model performance, transparency, and efficiency. Several researchers have developed frameworks and methodologies aimed at achieving these objectives. For instance, ref. [38] combined various data types to improve model efficiency. In efforts to increase transparency, refs. [39,41] integrated different techniques and models. Other researchers focused on boosting model performance by integrating black-box and white-box models [14] or employing specific prediction techniques to estimate model performance for certain ensemble methods [40]. Table 7 presents the advancement types, techniques, and key findings of advancements that aim to enhance the model's performance, transparency, and efficiency from the selected articles.

**Table 7.** Advancements that enhance model performance, transparency, and efficiency.

| Advancement Type | Advancement Technique | Findings | Ref. |
|---|---|---|---|
| Qu et al. Methodology | Utilising multimodal data fusion to predict student performance. | The model exhibited strong generalisation capabilities across various datasets and enhanced the efficiency of classification models. | [38] |
| Han et al. Framework | Utilising SHAP and BERT (Bidirectional Encoder Representations from Transformers) techniques, to enhance model performance and transparency, identify significant features, and detect dataset inconsistencies. | The framework enhanced the accuracy and robustness of the model. The application of SHAP clarified the model's reasoning, which is essential for debugging and further enhancement, leading to better performance, increased transparency, and greater comprehensibility. | [39] |
| EduBoost Methodology | Combining elements of black-box and white-box models to formulate interpretable and high-performance grey-box models to improve the model performance | Grey-box models consistently surpassed white-box models in performance. | [14] |
| Shapley-Based Feature Augmentation (SFA) Methodology | Augmenting the original features with out-of-fold predictions and explanatory features derived from Shapley values, enriches the data with informative features that significantly boost model performance across various datasets to enhance predictive performance. | The SFA method significantly improved predictive performance using out-of-fold predictions and Shapley values as augmented features. | [40] |
| Nimmy et al. Framework | Detailing step-by-step explanations on how decisions are reached. It facilitates the identification of input features that exhibit strong correlations with various decision outputs. | The framework provides transparent explanations for time-series decision-making tasks. | [41] |

### *3.3. Evaluation Metrics and Methods (ADV2)*

Articles in this category concentrate on developing or enhancing metrics and methods for evaluating the consistency, effectiveness, accuracy, and reliability of explanations produced by XAI models. These metrics and methods are critical for assessing the performance of XAI models and the comprehensibility of their explanations.

### 3.3.1. Evaluate Explanation Consistency

Evaluating the consistency of explanations in XAI necessitates assessing the determinism and implementation invariance of the explanation methods. Evaluating the explanation's consistency can also increase the explanation's reliability [44]. The following approaches show various ways to evaluate this consistency, including comparison methods, assessment of importance scores for each feature influencing the output, measurement of feature contribution explanations within XAI methods, and development of quantitative measures. These studies developed multiple techniques to evaluate the XAI explanations consistency as shown in Table 8.

**Table 8.** Advancements in evaluating the explanation's consistency.

| Advancement Type | Advancement Technique | Findings | Ref. |
|---|---|---|---|
| Yeo et al. Framework | Used the comparison method to maintain consistency. The framework enables objective comparisons between different XAI methods by emphasising consistent elements and excluding subjective aspects of explainability. | The study revealed that when the classifier accurately identified the underlying pattern in the model, the LIME algorithm most accurately selected the ground truth features. | [9] |
| Ratul et al. Methodology | Assessing model-agnostic attribution procedures, such as SHAP and LIME, which provide importance scores for each feature in input data that influences the model's output. | The evaluation methodology maintained the consistency, precision, and generality of the explanations. | [42] |
| Huang et al. Framework | Calculating feature contribution values and explanation summaries, focusing on the consistency and stability of XAI methods. | The framework helps assuring XAI methods address both stability and consistency. Explanation stability assesses intra-XAI method consistency across multiple datasets, while explanation consistency compares different XAI methods on the same dataset. | [43] |
| Rokade et al. Framework | Providing quantitative measures to assess the consistency, efficiency, integrity, and preciseness of explanations, aiming to quantify the reliability of explanations generated by XAI algorithms. | The application of this framework demonstrated the feasibility of quantitatively evaluating the explainability of AI models. | [44] |

### 3.3.2. Evaluate the Effectiveness

This section presents two different approaches for evaluating effectiveness. Speith et al. [45] developed a method to assess the effectiveness of XAI systems, while Zhang et al. [46] introduced a framework to evaluate the effectiveness of explanation methods in ML.

To improve how XAI systems are assessed for effectiveness, Speith et al. [45] developed a framework for categorising evaluation methods (EMs) based on the aspects of the XAI process they targeted. The framework includes explanatory, understanding, and desiderata satisfaction. This framework categorises EMs into three distinct groups.

First, Explanatory Information Evaluation Methods concentrate on the quality and accuracy of the explanatory information provided by explainability approaches. These methods assess whether the information correctly describes system-related aspects and is perceived as useful and comprehensible by the recipients. Second, Understanding Evaluation Methods measure the extent to which the explanatory information aids in comprehending system-related aspects, evaluating the facilitation of understanding. Lastly, Desiderata Evaluation Methods determine whether explainability approaches meet relevant societal desiderata, such as trust or fairness, and examine the impact of the explainability approach on significant outcome variables.

They found that high fidelity and completeness are essential for evaluating explanatory information. They observed that focusing solely on explanatory information does not necessarily indicate whether users better understand system-related aspects. They concluded that a comprehensive assessment of explainability approaches should integrate previous classification perspectives with their new framework.

Furthermore, to assess the effectiveness of explanation methods in ML, particularly for neural networks and ensemble models, one approach involves quantifying the impact of removing or altering high-contributing features within a model's dataset on its predictive accuracy, as applied by Zhang et al. [46]. They introduced an evaluation framework called the Mean Degree of Metrics Change (MDMC), which provides an empirical and quantifiable means to validate the quality of explanations offered by XAI techniques, distinguishing the effects of various explanation methods such as SHAP and LIME on model predictions.

### 3.3.3. Models Evaluation

A study by Bardos et al. [31] introduced a technique to evaluate the model's performance and evaluated the performance of Local eXplainable Dimensionality Reduction (LXDR) against its global counterpart, Global eXplainable Dimensionality Reduction (GXDR), using specific metrics to assess effectiveness and accuracy.

LXDR and GXDR are techniques that provide interpretability for dimensionality DR methods, especially non-linear and opaque methods. For more details, LXDR is a model-agnostic technique that provides local interpretations of the results of any non-linear DR technique. GXDR, on the other hand, is a global approach to explaining DR techniques.

Bardos et al. [31] found that LXDR outperformed GXDR in the "weights difference" metric, which compares extracted weights to the ground truth DR weights for interpretable Principal Component Analysis (PCA), and the "instance difference" metric, which assesses reduced representations across multiple datasets. Although LXDR's time performance degraded with very high-dimensional datasets and larger neighbourhood sizes, they identified optimisation and parallelisation opportunities to enhance scalability, highlighting LXDR's robustness and applicability across various scenarios.

### 3.3.4. Evaluate Explainability

To evaluate the explainability of textual information outputs from XAI systems, Sovrano et al. [47] developed the Degree of Explainability (DoX) method. This method quantifies explainability as directly proportional to the number of relevant questions a piece of information can accurately address. They also formulated a mathematical expression to operationalise this concept and incorporated it into a tool named DoXpy. This formula assesses the number of relevant questions (from a predefined set) that the information can satisfactorily answer, examining various aspects to measure its explainability.

They determined that the DoX metric effectively assesses the explainability of AI systems and correlates well with other established explainability measures. This finding suggests that the DoX methodology can serve as a reliable and objective alternative to the more subjective, user-based studies traditionally used for evaluating explainability in AI.

### *3.4. User-Centred and XAI Systems Design (ADV3)*

Studies in this category cover methods to enhance the user experience, accessibility, and design methodologies for XAI systems. These studies aim to make XAI more user-friendly and comprehensible for a diverse user base, including non-experts. Additionally, these studies introduce systematic steps or guidelines for developing XAI systems from both conceptual and technical perspectives. This section is divided into two main sections: user-centred design and XAI system design.

#### 3.4.1. User-Centred

This section includes research focused on understanding and addressing users' needs, emphasising helping users comprehend the outputs of black box, improving the communication between ML experts and users, and enabling them to explore and evaluate various options provided by XAI systems without relying on a single recommendation. Table 9 presents the recent advancements in developing XAI systems from a user-centered perspective.

**Table 9.** Advancements in XAI user-centred systems.

| **Advancement Type** | **Advancement Technique** | **Findings** | **Ref.** |
|---|---|---|---|
| Ontology-Driven Conceptual Model (ODCM) Framework | To understand the stakeholders' needs by following the XAI Requirement Elicitation (REXAI), which involves identifying, gathering, and defining stakeholders' needs and expectations for XAI systems. | The framework aligns well with user needs and cognitive processes, enhancing understanding, reasoning, and decision-making objectives. They concluded that their framework leads to enhanced user experiences and improved system performance in XAI systems. | [48] |
| Karpagam et al. Framework | Developing conceptual framework addressing the needs of various stakeholders across different data-intensive systems. | The framework illustrates how XAI can bridge the gap between AI models and their users, fostering trust and understanding. | [15] |
| LEWIS framework | Showing how changing certain inputs could affect the outputs. This system is helpful for users because it explains decisions in a way that is easy to understand, regardless of their technical background. | The framework helps end-users understand the black-box algorithm decisions. | [36] |
| COLE-HP methodology | Leveraging the more understandable logic of the simpler models such as decision tree or KNN to explain the black-box decisions, which allows users to gain insights into how the complex model makes decisions. | They found that providing post hoc explanations by example, identified through the COLE-HP method, effectively helped users understand why certain predictions were made. | [33] |

**Table 9.** *Cont.*

| Advancement Type | Advancement Technique | Findings | Ref. |
|---|---|---|---|
| Adnan et al. Framework | Making data analysis results comprehensible in a human-readable manner to provide global and local interpretations of students' performance using ML and DL techniques. | The framework significantly enhanced the interpretability of predictions, enabling educators to comprehend the rationale behind specific predictions, which is crucial for fostering trust and deriving actionable insights. | [49] |
| Evaluative AI framework | Selecting which hypotheses to investigate, and to assist them in exploring and assessing various options without offering a singular recommendation. | The framework aids in achieving balanced decisions and it is well-aligned with human cognitive processes, offering decision-makers control and flexibility and supporting essential decision-making components. | [50] |

### 3.4.2. XAI System Design

This section presents studies focusing on developing XAI systems, emphasising simplifying their creation and automating the selection of appropriate XAI solutions. It also includes research that guides developing XAI systems by aligning user needs with feature contribution explanations. It also outlines general steps for generating explanations and interpretations of AI systems and establishing communication guidelines between ML experts and end-users. These studies' techniques and findings can be found in Table 10.

**Table 10.** Advancements in XAI system design.

| Advancement Type | Advancement Technique | Findings | Ref. |
|---|---|---|---|
| XAI4PublicPolicy framework | A no-code solution for XAI to facilitate the use of XAI in policymaking, enabling non-technical users to generate and understand AI explanations and enabling the creation of XAI dashboards without necessitating any programming skills. | The framework automates the creation of XAI dashboards through a model executor, which manages the selection of models, datasets, and charts, generating explanations. It also improves the functionality and performance of the XAI system. | [52] |
| AutoXAI framework | XAI solutions are recommended to users based on their specific parameters and context. | The framework provides XAI solutions to fits users' needs based on user context. | [53] |
| Huang et al. Framework | Evaluating intra-method and inter-method consistency to enhance user trust and understanding and provide effective and consistent AI explanations. | The framework was demonstrated to be broadly applicable across different types of models and XAI methods. Also, it can be used for large-scale XAI projects involving large datasets and many target models. | [43] |

**Table 10.** *Cont.*

| Advancement Type | Advancement Technique | Findings | Ref. |
|---|---|---|---|
| Question bank framework | Mapping user needs for explainability to specific questions that users might ask about an AI system. The framework guides XAI developers and help them identify and prioritise user needs for explainability. | The framework underscores the substantial variability in user needs for explainability within AI systems. | [54] |
| Palacio et al. Framework | It provides concrete definitions for XAI-related terms and outlines all steps necessary to produce explanations and interpretations, offering a structured approach to the process. | Its compliance with existing concepts and desiderata related to XAI and by showcasing its application in practical use cases. The framework is designed to be commensurable and universal, meaning it can be applied across different XAI domains and contexts. | [55] |
| Severes et al. Framework | By proposing communication guidelines to improve communication between machine learning experts and the end-users. | The framework makes XAI system more accessible and understandable to non-experts. | [51] |

### 3.5. Type of Data Used in the Developed Advancements

We observed that most of the approaches reviewed in this study predominantly address tabular and textual data, with less emphasis on image data. It is crucial to note that we excluded all papers focusing exclusively on image-based approaches, as these were primarily developed for applications within the medical sector.

Table 11 presents various frameworks and methodologies reviewed in this study based on the data type they handle, which is tabular, textual, image, and not applicable (NA). For the frameworks categorised as NA, the researchers who developed them did not explicit the type of data that their advancements deal with. This classification aids in understanding the direction of XAI trends, guiding researchers and practitioners in selecting the appropriate methods for their specific needs.

**Table 11.** Classification of advancements XAI by data type.

| Type of Advancement | Name | Type of Data | | | | Year | Source |
|---|---|---|---|---|---|---|---|
| | | Tabular | Textual | Image | NA | | |
| Framework | LEWIS | ✓ | | | | 2021 | [36] |
| Framework | - | ✓ | | | | 2022 | [9] |
| Framework | BRB | ✓ | | | | 2023 | [41] |
| Framework | L2C | ✓ | | | | 2023 | [35] |
| Framework | Arg-XAI | ✓ | | | | 2022 | [30] |
| Framework | MDMC | ✓ | | | | 2021 | [46] |
| Framework | XAI4PublicPolicy | ✓ | | ✓ | | 2023 | [52] |
| Framework | - | ✓ | ✓ | | | 2022 | [49] |
| Framework | - | | ✓ | | | 2023 | [39] |
| Framework | INTERACTION | | ✓ | | | 2022 | [37] |

**Table 11.** *Cont.*

| Type of Advancement | Name | Type of Data | | | | Year | Source |
|---|---|---|---|---|---|---|---|
| | | **Tabular** | **Textual** | **Image** | **NA** | | |
| Framework | - | ✓ | ✓ | ✓ | | 2021 | [55] |
| Framework | COLE | ✓ | ✓ | ✓ | | 2021 | [33] |
| Framework | Neuralization-Propagation | ✓ | ✓ | ✓ | | 2022 | [32] |
| Framework | - | ✓ | ✓ | ✓ | | 2022 | [43] |
| Framework | - | ✓ | ✓ | ✓ | | 2023 | [51] |
| Framework | - | ✓ | ✓ | ✓ | | 2021 | [44] |
| Framework | - | ✓ | ✓ | ✓ | | 2022 | [15] |
| Framework | AutoXAI | | | | ✓ | 2022 | [53] |
| Framework | Evaluative AI | | | | ✓ | 2023 | [50] |
| Framework | - | | | | ✓ | 2020 | [54] |
| Framework | ODCM | | | | ✓ | 2023 | [48] |
| Framework | EMs | | | | ✓ | 2023 | [45] |
| Methodology | EduBoost | ✓ | | | | 2023 | [14] |
| Methodology | SFA | ✓ | | | | 2023 | [40] |
| Methodology | MERLIN | ✓ | ✓ | | | 2024 | [34] |
| Methodology | RfBERT | ✓ | ✓ | | | 2022 | [38] |
| Methodology | DoX | | ✓ | | | 2023 | [47] |
| Methodology | LXDR | ✓ | ✓ | ✓ | | 2022 | [31] |
| Methodology | COLE-HP | ✓ | ✓ | ✓ | | 2021 | [33] |
| Methodology | - | ✓ | ✓ | ✓ | | 2021 | [42] |

## 4. Discussion

As we classified the advancements into three categories, we found that the selected papers are most concentrated on developing the current models and the user-centred and XAI system design.

Many XAI researchers use terms like “outputs”, “results”, and “decisions” interchangeably when referring to the information produced by black-box algorithms [36,56]. The variation in terminology reflects the diverse contexts in which black-box algorithms are discussed. The term “outputs” is often used in a general sense, referring to any information produced by the algorithm [36]. “Results” is commonly employed when discussing the outcomes of algorithmic processing, especially in scientific or analytical contexts [57]. “Decisions” is frequently used when the algorithm’s output directly influences or determines a course of action [56,57].

This discussion will examine the selected papers through the lens of three key questions: ‘What’, ‘How’, and ‘Why’. Arrieta et al. [29] proposed a clear, organised, and logical flow of information outlining the methods employed and, finally, the motivations and implications to develop XAI advancements.

In this review, the “What” question represents the main topics discussed in the papers. The “How” question details the methods and techniques to achieve these objectives. Finally, the “Why” question addresses the objectives and motivations behind selected research approaches.

We identified eight key research topics through our review within the field as shown in Table 12.

Additionally, upon reviewing the papers, we observed that the authors emphasise the data type utilised in their work. Therefore, we will discuss the data type associated with each approach and identify trends based on the data types employed in the selected studies.

**Table 12.** Key research topics through our review.

| Main Topics (What?) | References |
|---|---|
| Understanding black-box algorithms | [30–33,36] |
| Understanding the explanations | [34,35,37] |
| Improving ML model performance | [14,39] |
| Evaluation of XAI | [31,42–47] |
| User-centred conceptual frameworks | [15,48,50] |
| XAI system design guidelines | [43,51,54] |
| XAI methods | [9,41] |
| XAI solutions selection | [52,53] |

### *4.1. Understanding Black-Box Algorithms*

Various methods were used to identify the factors that influenced black-box algorithm outputs. By evaluating which factors significantly impact the model's predictions, researchers can gain insights into the model's learned patterns and relationships. Table 13 illustrates the methods developed by researchers to interpret black box outputs. Furthermore, Figure 6 highlights the rationale behind these methods and their role in enhancing the comprehension of ML results.

**Table 13.** How researchers developed approaches to understand how black-box algorithms work.

| What | How | Reference |
|---|---|---|
| Understanding Black-box Algorithms | Analysing logical relationships between features. | [30] |
| | Identifying input changes that are likely influence the output. | [36] |
| | Tracing the influence of input features on cluster assignments. | [32] |
| | Using simpler models to explain black-box decisions. | [33] |
| | Representing interpretability for dimensionality reduction (DR) techniques to highlight the most influential features in reduced dimensions | [31] |

While all works focused on dataset features, the studies were split between discussing human and technical considerations. Human considerations aimed to support greater accessibility such as generating human-readable explanations [30,33,36], while technical considerations aimed to explain decision-making systems [32]. Furthermore, Bardos et al. [31] aim to allow users to discern which features of the original dataset are most influential in the reduced dimensions.

### *4.2. Understanding the Explanations*

Although the primary goal of using XAI models is to generate explanations for the outputs of ML black-box algorithms, as shown in Table 14, many authors have developed approaches to enhance the understanding of these explanations. These methods include generating contrastive, counterfactual, and natural language explanations. Both counterfactual [35] and natural language explanations [37] aimed to achieve explanation diversity.

Overall, these approaches enhance the interpretability of AI models, making them more transparent and trustworthy, which is crucial for their adoption in various fields, including higher education. Figure 7 elucidates the rationale behind these methods and their significance in enhancing our comprehension of the XAI outputs.

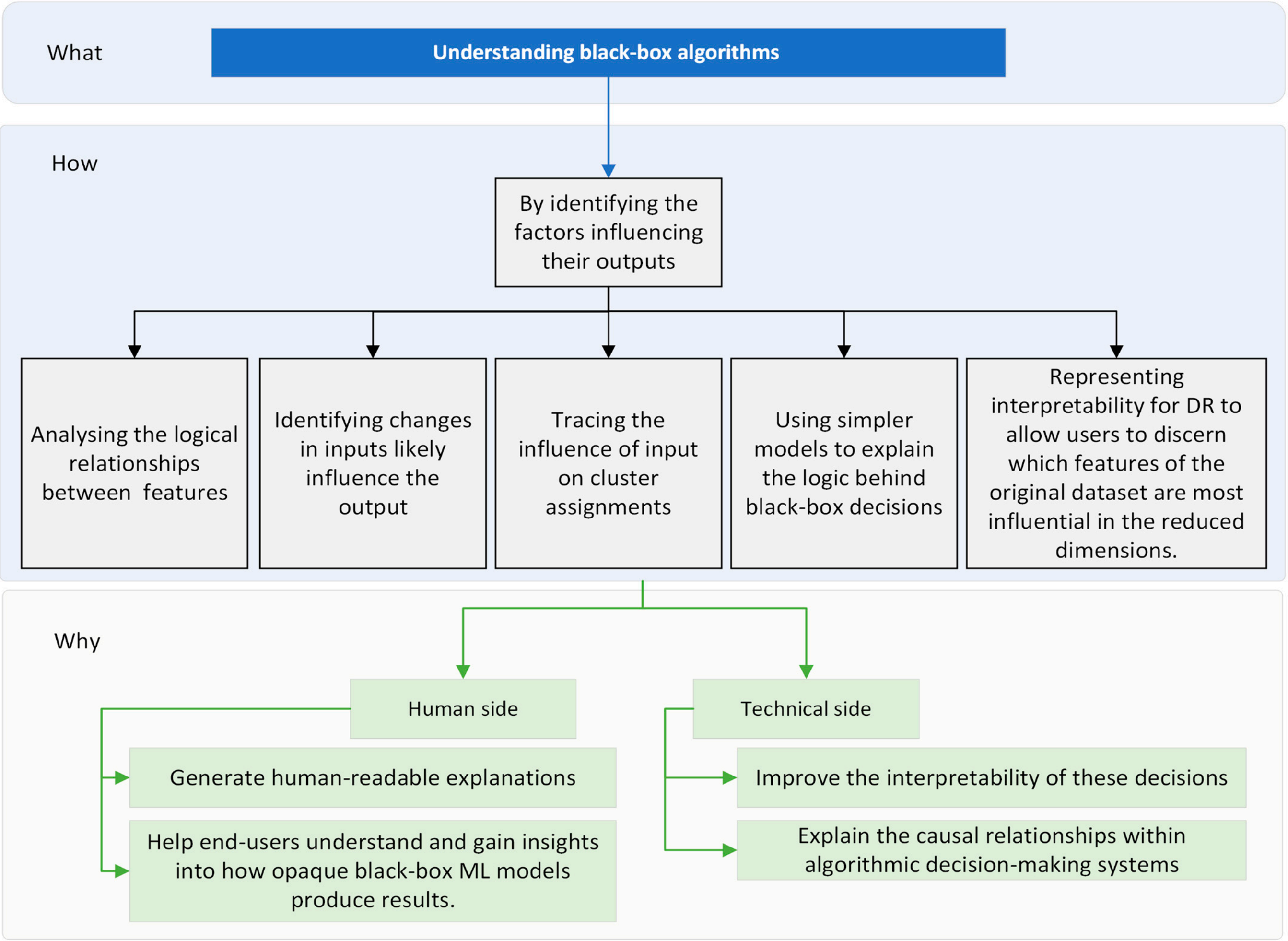
What
Understanding black-box algorithms
How
By identifying the factors influencing their outputs
Analysing the logical relationships between features
Identifying changes in inputs likely influence the output
Tracing the influence of input on cluster assignments
Using simpler models to explain the logic behind black-box decisions
Representing interpretability for DR to allow users to discern which features of the original dataset are most influential in the reduced dimensions.
Why
Human side
Generate human-readable explanations
Help end-users understand and gain insights into how opaque black-box ML models produce results.
Technical side
Improve the interpretability of these decisions
Explain the causal relationships within algorithmic decision-making systems

**Figure 6.** Flowchart for understanding how black-box algorithms work (What) by identifying the factors influencing their outputs (How) and their human or technical considerations (Why).

**Table 14.** Researcher approaches for understanding the explanations.

| What | How | Reference |
|---|---|---|
| Understanding the Explanations | By introducing contrastive explanations using symbolic reasoning to understand how two models can have similar predictive accuracy but different underlying logic. | [34] |
| | By generating natural language explanations for AI decisions. This is created using a transformer-based architecture combined with deep generative models. By analysing the posterior latent space, they produce multiple explanations reflecting the diversity of natural language, demonstrating how deep generative models can effectively process the premise, hypothesis, and explanation | [37] |
| | By producing counterfactual examples using a stochastic, feature-based approach for generating counterfactual explanations. This method achieves high levels of diversity and validity by describing how a model's output would change if certain input features were altered. | [35] |

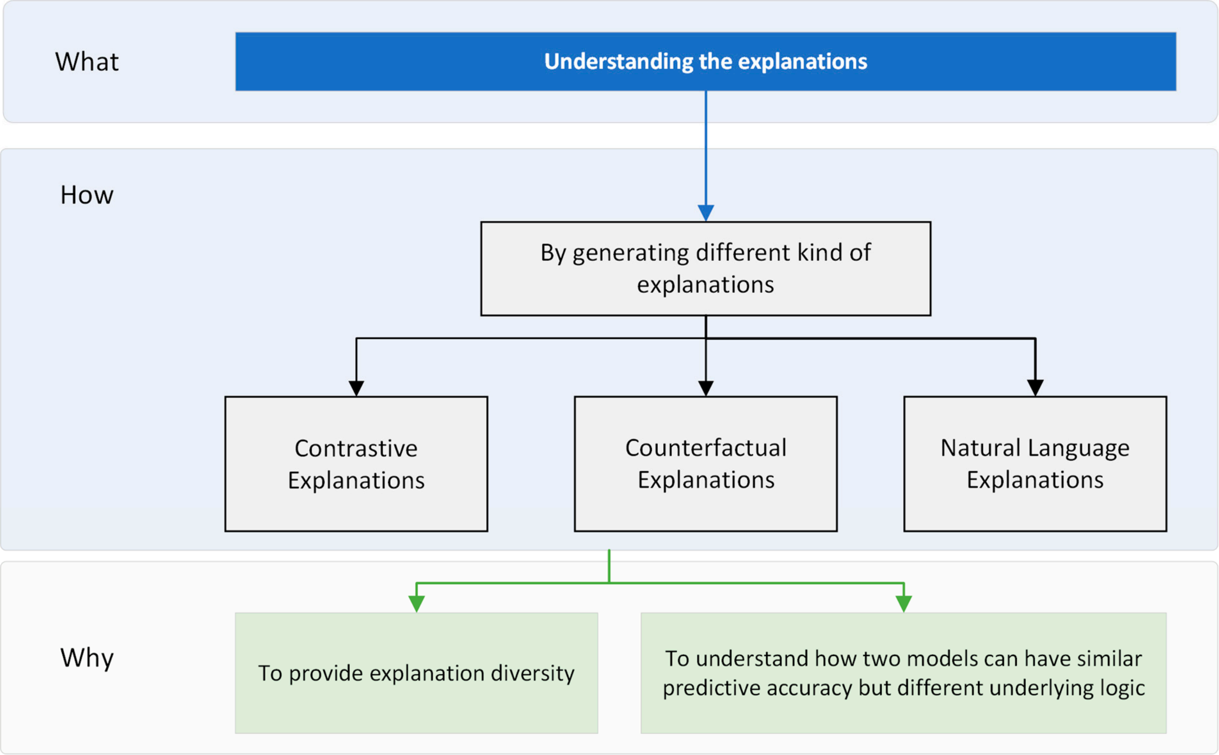
What
Understanding the explanations
How
By generating different kind of explanations
Contrastive Explanations
Counterfactual Explanations
Natural Language Explanations
Why
To provide explanation diversity
To understand how two models can have similar predictive accuracy but different underlying logic

**Figure 7.** Flowchart of the process of understanding XAI explanations (What) through diverse methods (How), and their goals (Why).

### *4.3. Improving ML Model Performance*

Some researchers have developed approaches to make the AI models more transparent and interpretable by improving the model's performance through technical means. XAI requires AI models to be more transparent and interpretable to provide insights into how they arrive at their outputs or decisions and build trust and understanding in AI systems [58]. This review discusses two approaches developed by Qin et al. [14] and Han et al. [39] that use the enhancement of XAI model performance to achieve model transparency and interpretability. Table 15 presents how researchers developed methods to increase the model performance.

**Table 15.** Research approaches for improving model performance.

| What | How | Reference |
|---|---|---|
| Improving ML Model Performance | Integrating XAI and ML techniques. They utilised SHAP with BERT, which allowed for a more precise understanding of the model's decisions and contributed to the refinement and accuracy of the classification process. | [39] |
| | Developing interpretable and high-performance grey-box models, which integrate aspects of both black-box and white-box models. | [14] |

The first approach developed by Han et al. [39], underscores the importance of XAI in developing robust and reliable ML models, demonstrating how transparency can lead to better performance and more trustworthy outcomes in various applications.

Qin et al. [14] illustrated the potential of grey-box models to balance the trade-off between interpretability and performance, making AI models more transparent and interpretable. Grey-box models balance the transparency of white-box models with the typical accuracy of black-box models [59]. Although they are not explicitly classified as XAI methods, they are highly relatable to the XAI domain due to their hybrid approach [14]. These models are essential to XAI efforts as they transform black-box models into forms that offer high performance and some interpretability. Figure 8 illustrates the rationale behind the development of these studies aimed at enhancing the performance of the ML model.

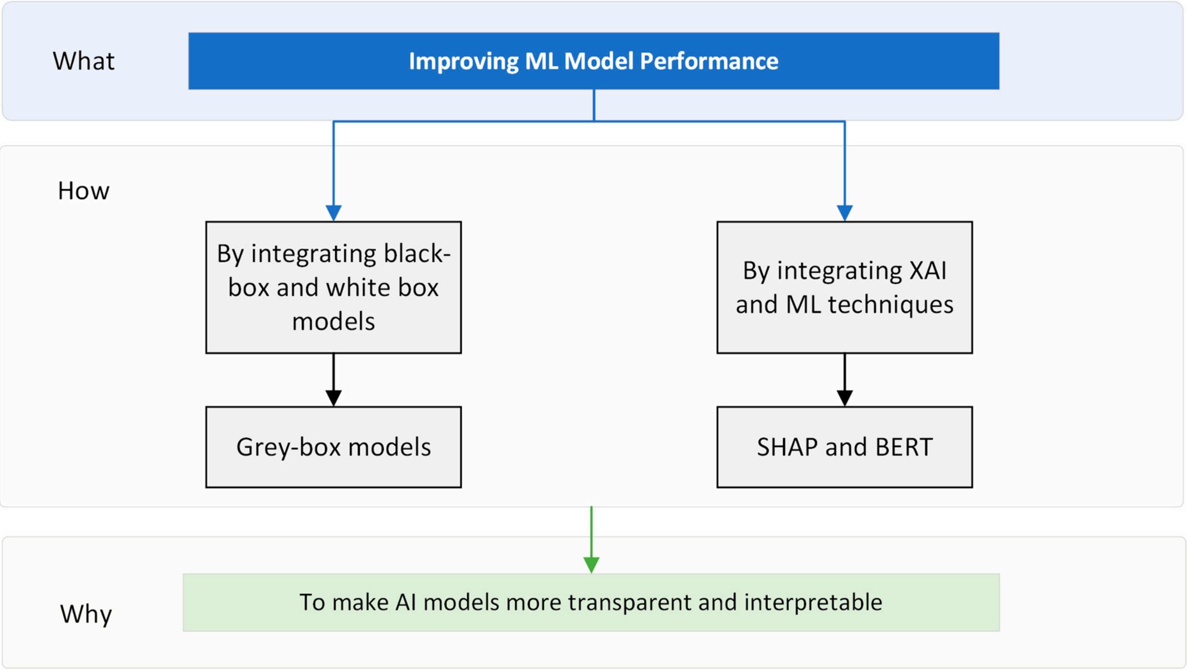
What
Improving ML Model Performance
How
By integrating black-box and white box models
By integrating XAI and ML techniques
Grey-box models
SHAP and BERT
Why
To make AI models more transparent and interpretable

**Figure 8.** Flowchart of the process of improving model performance (What) through specific techniques (How), to increase AI transparency and interpretability (Why).

### *4.4. Evaluations in XAI*

Evaluation approaches aim to ensure that XAI systems provide understandable, trustworthy, and actionable explanations for different stakeholders, ultimately enabling responsible and transparent AI deployment [60].

Many articles in this review mainly focus on developing evaluation approaches for various aspects of XAI, as shown in Table 16. Additionally, Figure 9 illustrates a flowchart of XAI evaluation studies using different methods to evaluate the effectiveness of XAI explanation systems.

**Table 16.** Research approaches for understanding XAI explanations.

| What | How | Reference |
|---|---|---|
| Evaluations in XAI | Incorporating the user-centric metrics in evaluating XAI systems. | [45] |
| | Developing a novel evaluation framework called the Mean Degree of Metrics Change (MDMC). | [46] |
| | Utilising different quantitative measures to evaluate the explanations consistency. | [44] |
| | Developing evaluation metrics for feature contribution explanations in the XAI methods. | [43] |
| | Assessment of model-agnostic attribution procedures. | [42] |
| | Developing a methodology to assess the explainability of textual information called the Degree of Explainability (DoX). | [47] |
| | Assessment of local explainable RD techniques. | [31] |

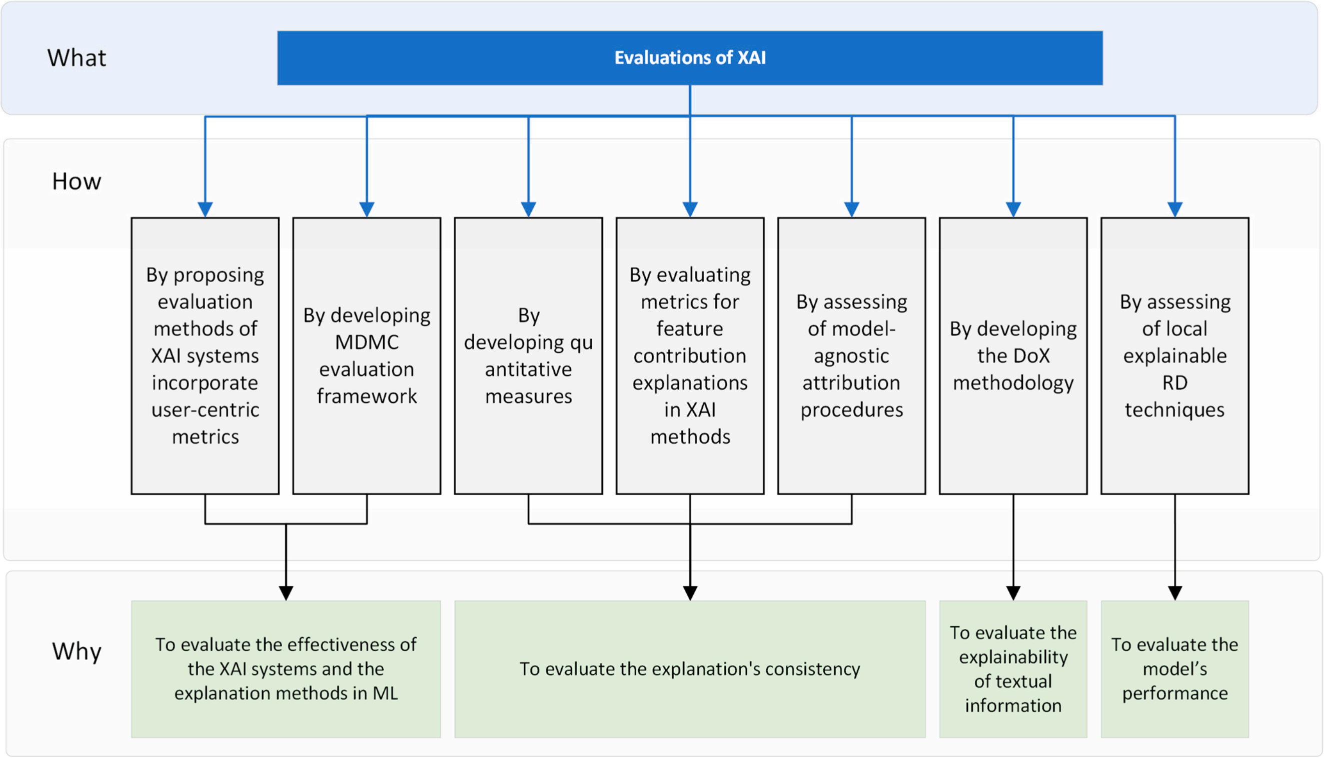
What
Evaluations of XAI
How
By proposing evaluation methods of XAI systems incorporate user-centric metrics
By developing MDMC evaluation framework
By developing qu antitative measures
By evaluating metrics for feature contribution explanations in XAI methods
By assessing of model-agnostic attribution procedures
By developing the DoX methodology
By assessing of local explainable RD techniques
Why
To evaluate the effectiveness of the XAI systems and the explanation methods in ML
To evaluate the explanation's consistency
To evaluate the explainability of textual information
To evaluate the model's performance

**Figure 9.** Flowchart of XAI evaluation articles (What) through various methods (How), to assess the effectiveness of XAI explanations and systems (Why).

### 4.5. User-Centred Conceptual Frameworks

We identified three articles that developed conceptual frameworks, all focusing on user-centred approaches, as shown in Table 17. The frameworks developed by Aslam et al. [48] and Karpagam et al. [15] primarily aim to understand stakeholders' needs. In contrast, Miller [50] introduces a framework designed to empower decision-makers. Figure 10 presents a flowchart illustrating user-centered conceptual frameworks developed through various research methods to enhance XAI processes.

**Table 17.** Types of user-centred conceptual frameworks.

| What | How | Reference |
|---|---|---|
| User-Centred Conceptual Frameworks | Using the XAI requirement elicitation (REXAI) to identify, gather, and define stakeholders' needs and expectations for XAI systems. | [48] |
| | Providing a general model to address the user's needs of various stakeholders. | [15] |
| | Allowing decision-makers to select hypotheses to investigate and assist them in exploring and assessing various options without offering a singular recommendation. | [50] |

### 4.6. XAI System Design Guidelines

Establishing guidelines for designing XAI systems is essential for developers to construct systems that perform effectively and meet user needs systematically. This review identifies three primary guidelines for developers of XAI systems, as shown in Table 18.

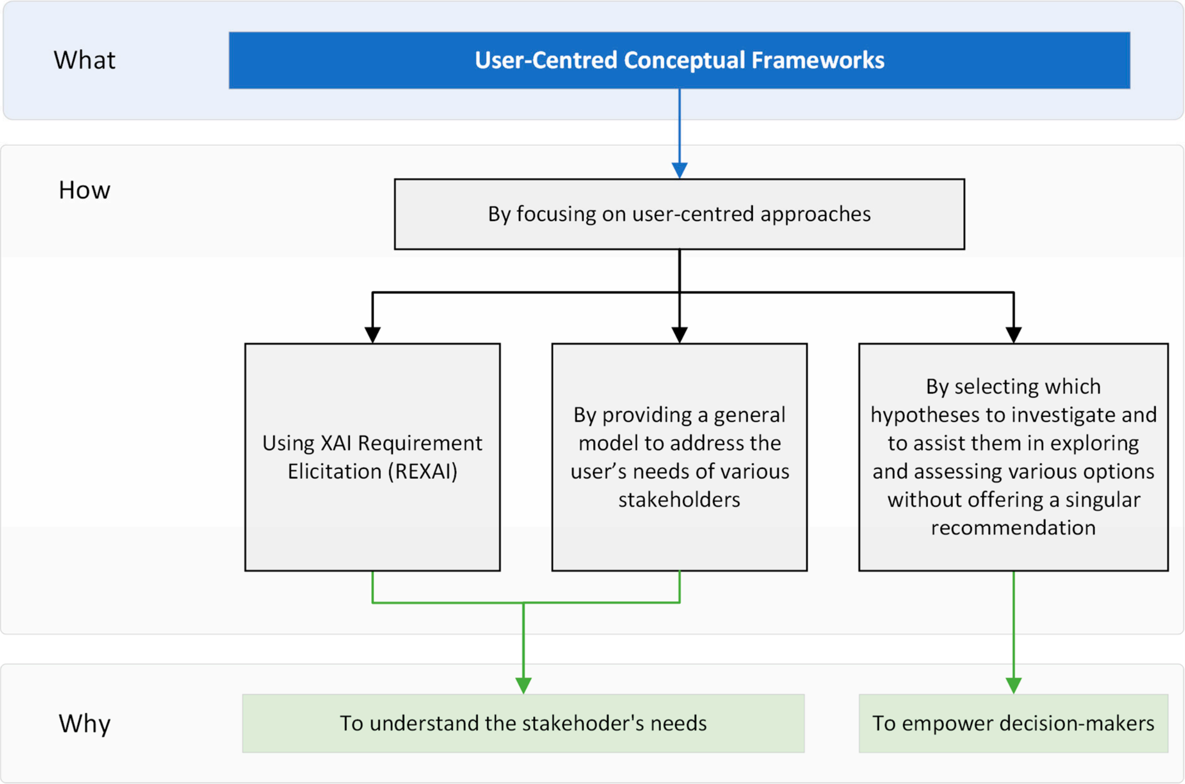
What
User-Centred Conceptual Frameworks
How
By focusing on user-centred approaches
Using XAI Requirement Elicitation (REXAI)
By providing a general model to address the user's needs of various stakeholders
By selecting which hypotheses to investigate and to assist them in exploring and assessing various options without offering a singular recommendation
Why
To understand the stakehoder's needs
To empower decision-makers

**Figure 10.** Flowchart on user-centred conceptual frameworks (What) devised through a variety of research approaches (How) to optimise XAI processes (Why).

**Table 18.** Methods of developing guidelines for building XAI systems.

| What | How | Reference |
|---|---|---|
| XAI System Design Guidelines | Developing question bank framework for developers and end-users who ask about an AI system, facilitated by an informed algorithm. | [54] |
| | Developing communication guidelines between the ML developers and non-expert users. | [51] |
| | Developing a general-purpose framework by evaluating intra-method and inter-method targeted the XAI developers. | [43] |

The first guideline focuses on guiding developers to identify and prioritise user needs for explainability [54], while the second guideline aims to make XAI systems accessible and understandable to non-expert users. Severes et al. [51] developed seven communication guidelines designed to ensure that the systems are user-friendly and comprehensible to those without technical expertise.

These guidelines facilitate developers' understanding of user expectations and needs and demonstrate to end-users the capabilities of the XAI system. This enables users to determine if the XAI system aligns with their requirements. The question bank and communication guidelines play a crucial role in comprehending the end-user's needs, thereby guiding the development of a system that meets these expectations and achieves the desired goals.

Designing XAI systems solely from the developer's perspective without considering the needs and capabilities of end-users can lead to ineffective and unreliable systems [61]. Therefore, incorporating user-centric design principles and clear communication is essential for developing effective and reliable XAI systems.

On the other hand, to help the developers select the most appropriate XAI method, Huang et al. [43] formulated a general-purpose framework by evaluating intra-method and inter-method consistency and explaining AI models through feature contribution explanations. Their framework addresses the critical need for effective and consistent AI explanations, enhancing user trust and understanding.

Figure 11 illustrates the connection between various established frameworks of XAI system design guidelines. These frameworks prioritize user needs, aim to make XAI systems accessible to non-expert users, and assist developers in choosing the most suitable option methods.

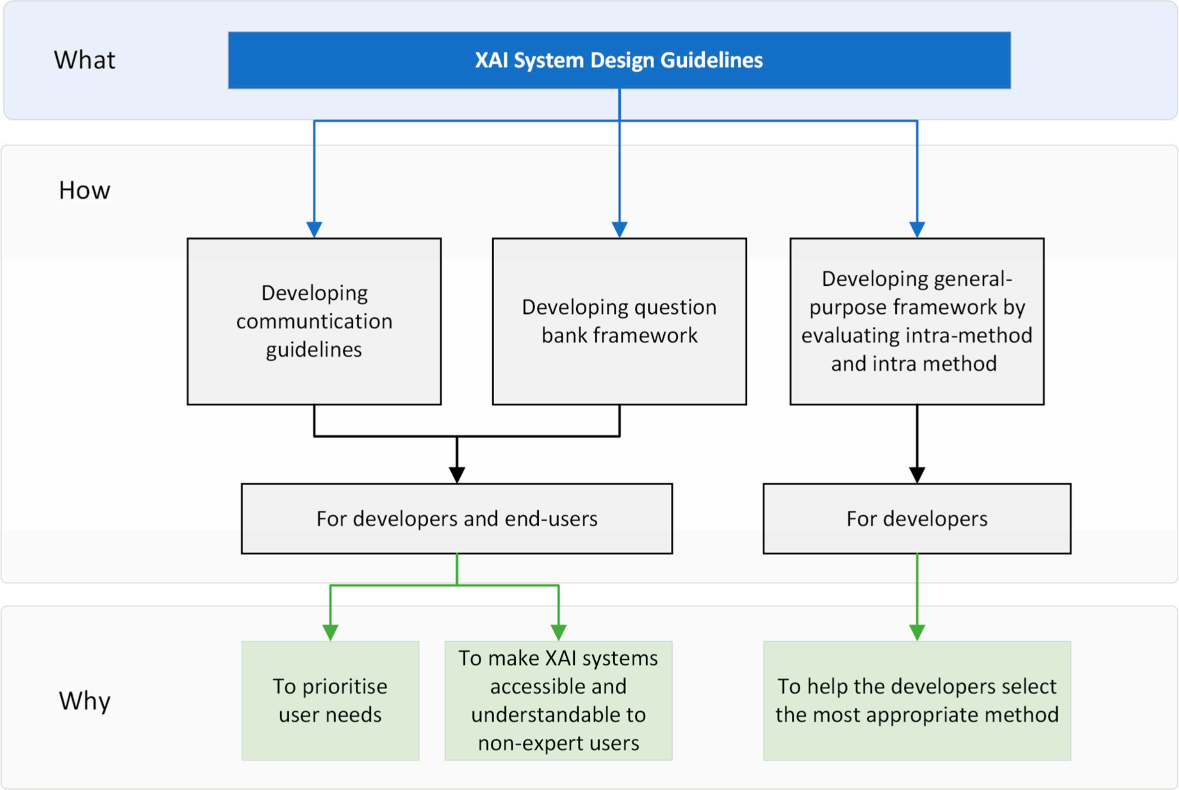
What
XAI System Design Guidelines
How
Developing communtication guidelines
Developing question bank framework
Developing general-purpose framework by evaluating intra-method and intra method
For developers and end-users
For developers
Why
To prioritise user needs
To make XAI systems accessible and understandable to non-expert users
To help the developers select the most appropriate method

**Figure 11.** Flowchart of XAI system design guidelines (What) outlined by developing frameworks and guidelines (How), with goals to prioritise user needs, make XAI systems accessible to non-expert users, and help developers select the most appropriate methods (Why).

### *4.7. XAI Methods*

XAI methods aim to make AI models transparent by providing understandable and interpretable explanations of their decisions. These methods can be categorised into intrinsic and post hoc techniques, each offering unique approaches to achieve transparency. This review has two articles that focus on the XAI methods, as shown in Table 19.

**Table 19.** How researchers enhance XAI methods.

| What | How | Reference |
|---|---|---|
| XAI Methods | Incorporating a Belief-Rule-Based (BRB) approach and utilising the casual links to model the evolution of feature values in progressive decisions. | [41] |
| | Focusing on consistent elements and disregarding subjective aspects of explainability. | [9] |

The paper by Nimmy et al. [41] establishes a framework, which addresses the several key limitations of existing XAI methods. Designed to provide glass-box explanations for time-series decision-making tasks, BRB systems have emerged as a promising approach in the field of XAI, particularly for decision support and prediction tasks in complex domains. They are designed to handle various types of uncertainties in knowledge representation and inference procedures, including vagueness, imprecision, randomness, ignorance, and incompleteness [62].

Nimmy et al. [41] utilised causal links, which focus on identifying causal features of a response variable by leveraging data from heterogeneous environments [63]. The evolution of feature values in progressive decision-making is a complex process involving dynamic adaptation and refinement of features over time to improve decision outcomes [64]. This concept is particularly relevant in ML and AI applications, where models continuously learn and update their parameters based on new data and feedback.

A progressive decision output is described as one where the inputs to the ML model at a given time $t_z$ are used to recommend a decision output $d$ at a future time slot $t_{z+n}$ [41]. In such cases, the model must determine how each input feature evolves from $t_z$ to $t_{z+n}$ before determining the output d at $t_{z+n}$. Examples of progressive decision outputs include risk assessment and predicting a financial position at a future time period. ML models that utilise causal links to model the evolution of feature values in progressive decisions can provide more transparent and interpretable explanations in XAI. This approach allows a better understanding of how different factors influence outcomes over time.

In contrast, Yeo et al. [9] introduces a comparison framework for XAI methods that emphasises consistent elements and excludes subjective aspects of explainability. This approach aims to enable objective comparisons between different XAI methods. The study presents consistency from a different angle, which was discussed in the Section 3.3.1, and shows how model evaluation can achieve this consistency. Also, Yeo, et al. [9] seeks to establish explanation consistency by comparing different XAI methods while excluding subjective aspects. The subjective aspects of explainability in XAI refer to the psychological perceptions and cognitive factors that influence how humans interpret and understand the explanations provided by AI systems [65]. Explainability is fundamentally a human perception influenced by factors such as understandability, interpretability, and transparency.

These advancements utilised different approaches to provide transparent explanations and provide objective comparisons between different XAI methods as shown in Figure 12.

### *4.8. XAI Solutions Selections*

Selecting the appropriate XAI solution to aid stakeholders in achieving their objectives is complex. Table 20 shows various research papers highlight how stakeholders seeking XAI solutions to ensure transparency and trust in AI models. XAI aims to enhance the interpretability of AI models, making them more transparent and comprehensible to a wide range of stakeholders [66,67]. This systematic review examines two articles focused on simplified XAI solutions.

**Table 20.** XAI methods frameworks.

| What | How | Reference |
|---|---|---|
| XAI Methods | Developing a framework that recommends optimal XAI solutions and their hyperparameters based on specified XAI evaluation metrics. | [53] |
| | Proposing a no-ode XAI solution to facilitate, which enables non-technical users to create XAI dashboards without necessitating programming skills. | [52] |

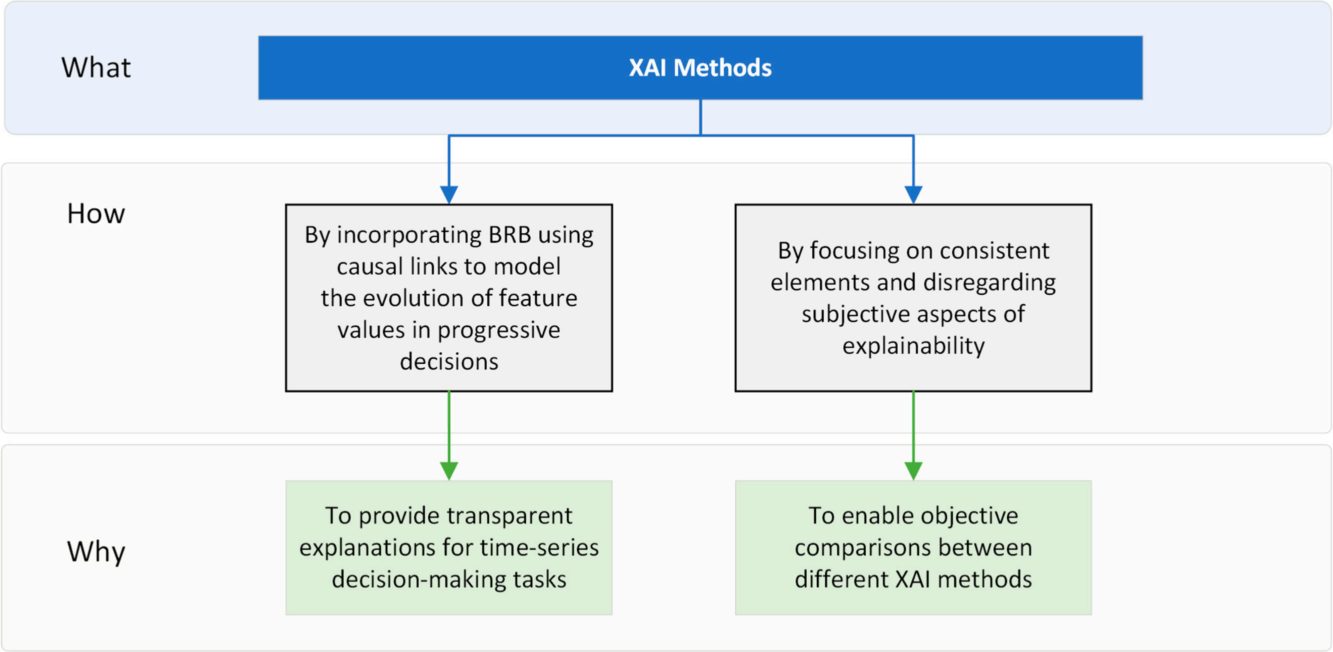
What
XAI Methods
How
By incorporating BRB using causal links to model the evolution of feature values in progressive decisions
By focusing on consistent elements and disregarding subjective aspects of explainability
Why
To provide transparent explanations for time-series decision-making tasks
To enable objective comparisons between different XAI methods

**Figure 12.** Flowchart of XAI methods (What) that utilise technical features (How) to optimise XAI processes (Why).

The paper by Cugny et al. [53] presents a framework that tailors the most suitable XAI solutions to specific user contexts, considering factors such as the dataset, ML model, and XAI needs and constraints. This approach is particularly advantageous for developers and technical users who require precise and context-specific XAI recommendations.

In contrast, Martinez et al. [52] proposes a no-code XAI solution, which enables non-technical users to generate and comprehend AI explanations.

These two studies, as shown in Figure 13, highlight the importance of selecting XAI solutions based on the stakeholders' context and needs. Cugny et al. [53] concentrates on optimising the selection process for technical users, whereas Martinez et al. [52] prioritises accessibility for non-technical users. Collectively, these approaches highlight the necessity of context-specific XAI solutions and the importance of tailoring these tools to meet the diverse requirements of various stakeholders.

### *4.9. Data Types in Developed Advancements*

As the field of XAI continues to evolve, developing methods capable of handling diverse data types will be essential in addressing the challenges of explainability in complex AI systems.

Exploring XAI frameworks and methodologies based on the data types they support is crucial for researchers and practitioners. For instance, frameworks like AutoXAI and Evaluative AI, which handle multiple data types, offer a significant advantage in projects requiring comprehensive data analysis across different formats.

Moreover, some researchers found that the fusion of multiple data types can increase model prediction accuracy. Qu et al. [38] claim that combining tabular and textual data within a Transformer-based framework markedly increased prediction accuracy. The study demonstrated that integrating tabular data with unstructured textual data produced a more robust predictive model. This fusion approach surpassed traditional methods that depend on a single data type, emphasising the advantages of a multimodal strategy in educational data mining.

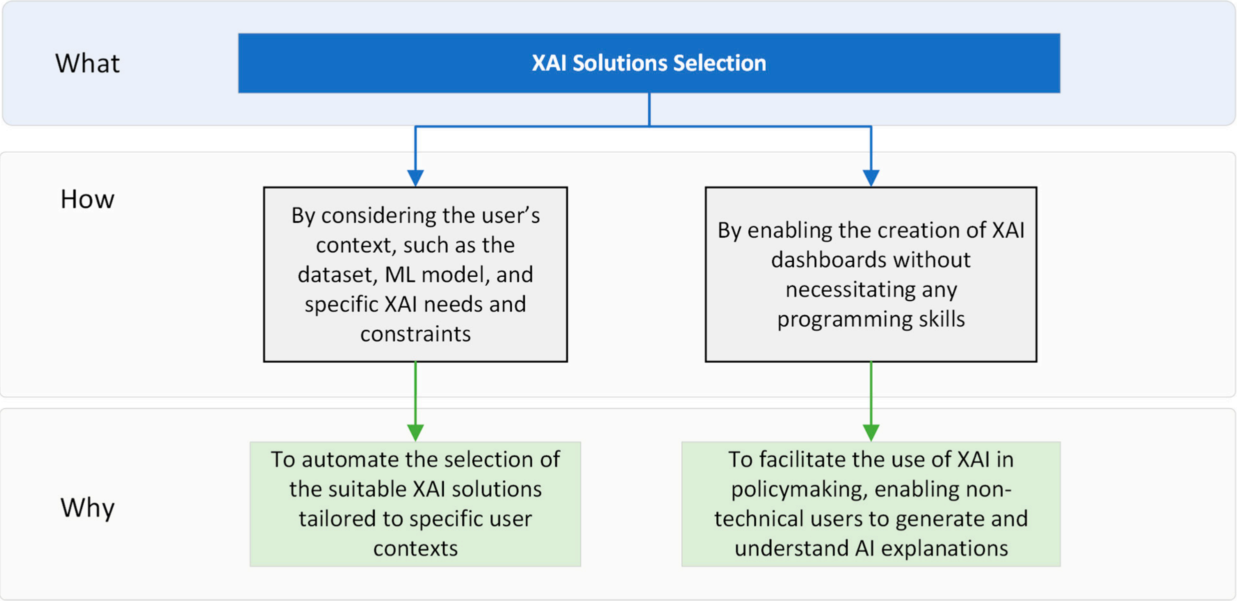
What
XAI Solutions Selection
How
By considering the user's context, such as the dataset, ML model, and specific XAI needs and constraints
By enabling the creation of XAI dashboards without necessitating any programming skills
Why
To automate the selection of the suitable XAI solutions tailored to specific user contexts
To facilitate the use of XAI in policymaking, enabling non-technical users to generate and understand AI explanations

**Figure 13.** Flowchart of XAI solutions (What) that consider users and enable technical features (How) to expand the accessibility of solutions for a wider range of users (Why).

The benefits of multimodal data fusion present a compelling case for the development of XAI frameworks that inherently support and optimise such integrations. This could lead to more accurate XAI models, particularly in complex domains like education and healthcare, where data diversity is the norm. As the XAI field progresses, addressing these gaps will be vital for creating universally applicable and robust explainable AI solutions.

## 5. Conclusions

This systematic review examined recent advancements in the field of XAI from January 2014 to January 2024. The objective was to identify the primary research directions of XAI development and the reasons behind these advancements. Two research questions were formulated to guide this research.

The first research question aimed to understand the focal directions of the research community. A thematic analysis was conducted on the selected papers, categorising them into three key areas: model development, evaluation metrics and methods, and user-centred and XAI system design.

The review found that most advancements focused on enhancing the interpretability and transparency of ML models. The second category addressed the various approaches developed to evaluate XAI models, focusing on explanation consistency, effectiveness, and overall model evaluation. The final category placed greater emphasis on user-centred design in XAI systems, aiming to understand stakeholders' needs, develop conceptual frameworks and establish design guidelines to make XAI systems more accessible to end-users.

Furthermore, the review revealed that some approaches are designed for specific data types. We found that the most reviewed approaches targeted tabular and textual data, with less emphasis on image data.

The second research question investigated 'What', 'How', and 'Why' these advancements were developed. The discussion section elaborates on the methods used and the ultimate goals of these innovations. It was found that these advancements aimed to bridge the gap between technical model outputs and user understanding. Methods such as contrastive and counterfactual explanations, natural language explanations, and multimodal data fusion were employed to achieve this. These approaches aim to enhance user trust, improve model performance, and ensure AI systems' ethical and responsible deployment.

## 6. Future Work

The field of XAI continues to evolve, providing several ways for future research to explore aspects not covered in this review. While this review presents a general overview of advancements in XAI, we noticed that a few publications provided advancements in the context of education. Researchers can apply the methodologies discussed in this paper to examine advancements in other domains, such as the medical, energy, and financial sectors, to name a few.

Furthermore, the findings presented in this review can aid researchers in addressing the challenges and limitations within the XAI field. Despite ongoing efforts, numerous limitations persist, necessitating continuous problem-solving by researchers. Investigating how these challenges are addressed and the efficacy of various solutions could constitute a valuable direction for future research, ultimately helping to mitigate the limitations faced by the XAI sector.

**Author Contributions:** Conceptualisation, Z.M.A. and S.P.; methodology, Z.M.A.; validation, Z.M.A., S.P., and N.A.; resources, Z.M.A.; writing—original draft preparation, Z.M.A.; writing—review and editing, N.A.; visualisation, Z.M.A.; supervision, S.P. All authors have read and agreed to the published version of the manuscript.

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

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

## References

1. Antoniadi, A.M.; Du, Y.; Guendouz, Y.; Wei, L.; Mazo, C.; Becker, B.A.; Mooney, C. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review. *Appl. Sci.* **2021**, *11*, 5088. [CrossRef]
2. Rachha, A.; Seyam, M. Explainable AI in education: Current trends, challenges, and opportunities. In Proceedings of the SoutheastCon 2023, Orlando, FL, USA, 1–16 April 2023; pp. 232–239. [CrossRef]
3. Naiseh, M. C-XAI: Design Method for Explainable AI Interfaces to Enhance Trust Calibration. Ph.D. Thesis, Bournemouth University, Poole, UK, 2021.
4. Das, A.; Rad, P. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv* **2020**, arXiv:2006.11371. [CrossRef]
5. Chen, Z. Algorithms and Applications of Explainable Machine Learning. Ph.D. Thesis, State University of New York at Stony Brook, Stony Brook, NY, USA, 2023.
6. Loyola-Gonzalez, O. Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access* **2019**, *7*, 154096–154113. [CrossRef]
7. Lin, Z.; Trivedi, S.; Sun, J. Generating with confidence: Uncertainty quantification for black-box large language models. *arXiv* **2023**, arXiv:2305.19187. [CrossRef]
8. Jang, Y.; Choi, S.; Kim, H. Development and validation of an instrument to measure undergraduate students' attitudes toward the ethics of artificial intelligence (AT-EAI) and analysis of its difference by gender and experience of AI education. *Educ. Inf. Technol.* **2022**, *27*, 11635–11667. [CrossRef]
9. Yeo, G.F.A.; Hudson, I.; Akman, D.; Chan, J. A Simple Framework for XAI Comparisons with a Case Study. In Proceedings of the 2022 5th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 27–30 May 2022; pp. 501–508.
10. Ma, K. Integrated hybrid modeling and SHAP (SHapley Additive exPlanations) to predict and explain the adsorption properties of thermoplastic polyurethane (TPU) porous materials. *RSC Adv.* **2024**, *14*, 10348–10357. [CrossRef]
11. Hamilton, N.; Webb, A.; Wilder, M.; Hendrickson, B.; Blanck, M.; Nelson, E.; Roemer, W.; Havens, T.C. Enhancing visualization and explainability of computer vision models with local interpretable model-agnostic explanations (LIME). In Proceedings of the 2022 IEEE Symposium Series on Computational Intelligence (SSCI), Singapore, 4–7 December 2022; pp. 604–611.
12. Cao, S.; Hu, Y. Creating machine learning models that interpretably link systemic inflammatory index, sex steroid hormones, and dietary antioxidants to identify gout using the SHAP (SHapley Additive exPlanations) method. *Front. Immunol.* **2024**, *15*, 1367340. [CrossRef] [PubMed]
13. Biecek, P.; Burzykowski, T. Local interpretable model-agnostic explanations (LIME). *Explan. Model Anal. Explor. Explain Examine Predict. Models* **2021**, *1*, 107–124. [CrossRef]

14. Qin, A.; Boicu, M. EduBoost: An Interpretable Grey-Box Model Approach to Identify and Prevent Student Failure and Dropout. In Proceedings of the 2023 IEEE Frontiers in Education Conference (FIE), College Station, TX, USA, 18–21 October 2023; pp. 01–07.
15. Karpagam, G.; Varma, A.; Samrddhi, M. Understanding, Visualizing and Explaining XAI Through Case Studies. In Proceedings of the 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 25–26 March 2022; pp. 647–654.
16. Jung, J.; Lee, H.; Jung, H.; Kim, H. Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review. *Heliyon* **2023**, *9*, e16110. [CrossRef] [PubMed]
17. Sipos, L.; Schäfer, U.; Glinka, K.; Müller-Birn, C. Identifying explanation needs of end-users: Applying and extending the XAI question bank. In Proceedings of the Mensch und Computer 2023, Rapperswil, Switzerland, 3–6 September 2023; pp. 492–497.
18. Jiang, H.; Senge, E. On two XAI cultures: A case study of non-technical explanations in deployed AI system. *arXiv* **2021**. [CrossRef]
19. Venkatesh, S.; Narasimhan, K.; Adalarasu, K. An overview of interpretability techniques for explainable artificial intelligence (xai) in deep learning-based medical image analysis. In Proceedings of the 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 17–18 March 2023; pp. 175–182.
20. Weitz, K.; Schlagowski, R.; André, E.; Männiste, M.; George, C. Explaining It Your Way-Findings from a Co-Creative Design Workshop on Designing XAI Applications with AI End-Users from the Public Sector. In Proceedings of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024; pp. 1–14.
21. Singh, A.; Ahlawat, N. A review article:-the Growing Role of Data Science and AI in Banking and Finance. *Int. Res. J. Mod. Eng. Technol. Sci.* **2023**, *8*, 1047–1051. [CrossRef]
22. Mittelstadt, B.; Russell, C.; Wachter, S. Explaining explanations in AI. In Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 29–31 January 2019; pp. 279–288.
23. Keane, M.T.; Smyth, B. Good counterfactuals and where to find them: A case-based technique for generating counterfactuals for explainable AI (XAI). In Proceedings of the Case-Based Reasoning Research and Development: 28th International Conference, ICCBR 2020, Salamanca, Spain, June 8–12 2020; Proceedings 28. pp. 163–178.
24. Cambria, E.; Malandri, L.; Mercorio, F.; Mezzanzanica, M.; Nobani, N. A survey on XAI and natural language explanations. *Inf. Process. Manag.* **2023**, *60*, 103111. [CrossRef]
25. Page, M.J.; Moher, D.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ* **2021**, *372*, n160. [CrossRef]
26. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; PRISMA Group*, t. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann. Intern. Med.* **2009**, *151*, 264–269. [CrossRef] [PubMed]
27. Haddaway, N.R.; Page, M.J.; Pritchard, C.C.; McGuinness, L.A. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Syst. Rev.* **2022**, *18*, e1230. [CrossRef] [PubMed]
28. Braun, V.; Clarke, V. Thematic analysis. In *APA Handbook of Research Methods in Psychology*; Cooper, H., Camic, P.M., Long, D.L., Panter, A.T., Rindskopf, D., Sher, K.J., Eds.; American Psychological Association: Washington, DC, USA, 2012; Volume 2, pp. 57–71.
29. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
30. Bistarelli, S.; Mancinelli, A.; Santini, F.; Taticchi, C. Arg-xai: A tool for explaining machine learning results. In Proceedings of the 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI), Macao, China, 31 October–2 November 2022; pp. 205–212.
31. Bardos, A.; Mollas, I.; Bassiliades, N.; Tsoumakas, G. Local explanation of dimensionality reduction. In Proceedings of the 12th Hellenic Conference on Artificial Intelligence, Corfu, Greece, 7–9 September 2022; pp. 1–9.
32. Kauffmann, J.; Esders, M.; Ruff, L.; Montavon, G.; Samek, W.; Müller, K.-R. From clustering to cluster explanations via neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 1926–1940. [CrossRef]
33. Kenny, E.M.; Ford, C.; Quinn, M.; Keane, M.T. Explaining black-box classifiers using post-hoc explanations-by-example: The effect of explanations and error-rates in XAI user studies. *Artif. Intell.* **2021**, *294*, 103459. [CrossRef]
34. Malandri, L.; Mercorio, F.; Mezzanzanica, M.; Seveso, A. Model-contrastive explanations through symbolic reasoning. *Decis. Support Syst.* **2024**, *176*, 114040. [CrossRef]
35. Vo, V.; Le, T.; Nguyen, V.; Zhao, H.; Bonilla, E.V.; Haffari, G.; Phung, D. Feature-based learning for diverse and privacy-preserving counterfactual explanations. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Long Beach, CA, USA, 6–10 August 2023; pp. 2211–2222.
36. Galhotra, S.; Pradhan, R.; Salimi, B. Explaining black-box algorithms using probabilistic contrastive counterfactuals. In Proceedings of the 2021 International Conference on Management of Data, Virtual Event, China, 20–25 June 2021; pp. 577–590.

37. Yu, J.; Cristea, A.I.; Harit, A.; Sun, Z.; Aduragba, O.T.; Shi, L.; Al Moubayed, N. INTERACTION: A generative XAI framework for natural language inference explanations. In Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 18–23 July 2022; pp. 1–8.
38. Qu, Y.; Li, F.; Li, L.; Dou, X.; Wang, H. Can we predict student performance based on tabular and textual data? *IEEE Access* **2022**, *10*, 86008–86019. [CrossRef]
39. Han, L.; Zhou, Q.; Li, T. Improving Requirements Classification Models Based on Explainable Requirements Concerns. In Proceedings of the 2023 IEEE 31st International Requirements Engineering Conference Workshops (REW), Hannover, Germany, 4–5 September 2023; pp. 95–101.
40. Antwarg, L.; Galed, C.; Shimoni, N.; Rokach, L.; Shapira, B. Shapley-based feature augmentation. *Inf. Fusion* **2023**, *96*, 92–102. [CrossRef]
41. Nimmy, S.F.; Hussain, O.K.; Chakrabortty, R.K.; Hussain, F.K.; Saberi, M. An optimized Belief-Rule-Based (BRB) approach to ensure the trustworthiness of interpreted time-series decisions. *Knowl. -Based Syst.* **2023**, *271*, 110552. [CrossRef]
42. Ratul, Q.E.A.; Serra, E.; Cuzzocrea, A. Evaluating attribution methods in machine learning interpretability. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; pp. 5239–5245.
43. Huang, J.; Wang, Z.; Li, D.; Liu, Y. The Analysis and Development of an XAI Process on Feature Contribution Explanation. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022; pp. 5039–5048.
44. Rokade, P.; Alluri, B.K.R. Towards quantification of explainability algorithms. In Proceedings of the 5th International Conference on Advances in Artificial Intelligence, Virtual Event, UK, 20–22 November 2021; pp. 31–37.
45. Speith, T.; Langer, M. A new perspective on evaluation methods for explainable artificial intelligence (xai). In Proceedings of the 2023 IEEE 31st International Requirements Engineering Conference Workshops (REW), Hannover, Germany, 4–5 September 2023; pp. 325–331.
46. Zhang, Y.; Xu, F.; Zou, J.; Petrosian, O.L.; Krinkin, K.V. XAI evaluation: Evaluating black-box model explanations for prediction. In Proceedings of the 2021 II International Conference on Neural Networks and Neurotechnologies (NeuroNT), Saint Petersburg, Russia, 16 June 2021; pp. 13–16.
47. Sovrano, F.; Vitali, F. An objective metric for Explainable AI: How and why to estimate the degree of explainability. *Knowl. -Based Syst.* **2023**, *278*, 110866. [CrossRef]
48. Aslam, M.; Segura-Velandia, D.; Goh, Y.M. A conceptual model framework for XAI requirement elicitation of application domain system. *IEEE Access* **2023**, *11*, 108080–108091. [CrossRef]
49. Adnan, M.; Uddin, M.I.; Khan, E.; Alharithi, F.S.; Amin, S.; Alzahrani, A.A. Earliest possible global and local interpretation of students' performance in virtual learning environment by leveraging explainable AI. *IEEE Access* **2022**, *10*, 129843–129864. [CrossRef]
50. Miller, T. Explainable ai is dead, long live explainable ai! hypothesis-driven decision support using evaluative ai. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 12–15 June 2023; pp. 333–342.
51. Severes, B.; Carreira, C.; Vieira, A.B.; Gomes, E.; Aparício, J.T.; Pereira, I. The Human Side of XAI: Bridging the Gap between AI and Non-expert Audiences. In Proceedings of the 41st ACM International Conference on Design of Communication, Orlando, FL, USA, 26–28 October 2023; pp. 126–132.
52. Martinez, M.P.; Azqueta-Alzuaz, A. A No Code XAI Framework for Policy Making. In Proceedings of the 2023 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), Pafos, Cyprus, 19–21 June 2023; pp. 556–561.
53. Cugny, R.; Aligon, J.; Chevalier, M.; Roman Jimenez, G.; Teste, O. Autoxai: A framework to automatically select the most adapted xai solution. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, 17–21 October 2022; pp. 315–324.
54. Liao, Q.V.; Gruen, D.; Miller, S. Questioning the AI: Informing design practices for explainable AI user experiences. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–15.
55. Palacio, S.; Lucieri, A.; Munir, M.; Ahmed, S.; Hees, J.; Dengel, A. Xai handbook: Towards a unified framework for explainable AI. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 3766–3775.
56. Mohammadi, B.; Malik, N.; Derdenger, T.; Srinivasan, K. Regulating Explainable AI (XAI) May Harm Consumers. *Available at SSRN 4602571*. 2020. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4602571 (accessed on 17 February 2025).
57. Mohammadi, B.; Malik, N.; Derdenger, T.; Srinivasan, K. Sell Me the Blackbox! Regulating eXplainable Artificial Intelligence (XAI) May Harm Consumers. *arXiv* **2022**, arXiv:2209.03499. [CrossRef]
58. Thunki, P.; Reddy, S.R.B.; Raparthi, M.; Maruthi, S.; Dodda, S.B.; Ravichandran, P. Explainable AI in Data Science-Enhancing Model Interpretability and Transparency. *Afr. J. Artif. Intell. Sustain. Dev.* **2021**, *1*, 1–8.

59. Wanner, J.; Herm, L.-V.; Heinrich, K.; Janiesch, C.; Zschech, P. White, Grey, Black: Effects of XAI Augmentation on the Confidence in AI-based Decision Support Systems. In Proceedings of the ICIS 2020, Hyderabad, India, 13–16 December 2020.
60. Love, P.E.; Fang, W.; Matthews, J.; Porter, S.; Luo, H.; Ding, L. Explainable artificial intelligence (XAI): Precepts, models, and opportunities for research in construction. *Adv. Eng. Inform.* **2023**, *57*, 102024. [CrossRef]
61. Kaplan, S.; Uusitalo, H.; Lensu, L. A unified and practical user-centric framework for explainable artificial intelligence. *Knowl.-Based Syst.* **2024**, *283*, 111107. [CrossRef]
62. Rahaman, S.; Hossain, M.S. A belief rule based (BRB) system to assess asthma suspicion. In Proceedings of the 16th International Conference Computer and Information Technology, Bangladesh, Khulna, 8–10 March 2014; pp. 432–437.
63. Kook, L.; Saengkyongam, S.; Lundborg, A.R.; Hothorn, T.; Peters, J. Model-based causal feature selection for general response types. *arXiv* **2023**, arXiv:2309.12833. [CrossRef]
64. Du, D.; Cui, Z. An Evolutionary Model for Earthquake Prediction Considering Time-Series Evolution and Feature Extraction and Its Application. *J. Phys. Conf. Ser.* **2022**, *2333*, 012013. [CrossRef]
65. Alarcon, G.; Willis, S. Explaining Explainable Artificial Intelligence: An integrative model of objective and subjective influences on XAI. In Proceedings of the 56th Hawaii International Conference on System Sciences, Maui, HI, USA, 3–6 January 2023.
66. Gerlach, J.; Hoppe, P.; Jagels, S.; Licker, L.; Breitner, M.H. Decision support for efficient XAI services-A morphological analysis, business model archetypes, and a decision tree. *Electron. Mark.* **2022**, *32*, 2139–2158. [CrossRef]
67. Farrow, R. The possibilities and limits of XAI in education: A socio-technical perspective. *Learn. Media Technol.* **2023**, *48*, 266–279. [CrossRef]

**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.