big data and cognitive computing

Review

Exploring the Landscape of Explainable Artificial Intelligence (XAI): A Systematic Review of Techniques and Applications

Sayda Umma Hamida [1,†], Mohammad Jabed Morshed Chowdhury [1,2,*,†], Narayan Ranjan Chakraborty [1,†], Kamanashis Biswas [1,3,†] and Shahrab Khan Sami [4,†]

[1] Department of Computer Science and Engineering, Daffodil International University, Birulia, Dhaka 1216, Bangladesh; sayda25-880@diu.edu.bd (S.U.H.); narayan@daffodilvarsity.edu.bd (N.R.C.); kamanashis.biswas@acu.edu.au (K.B.)
[2] Department of Computer Science and IT, La Trobe University, Melbourne, VIC 3086, Australia
[3] Faculty of Law and Business, Australian Catholic University, Brisbane, QLD 4014, Australia
[4] Department of Computer Science and Engineering, Shah Jalal University of Science and Technology, Sylhet 3114, Bangladesh; sami8907456312@gmail.com
* Correspondence: m.chowdhury@latrobe.edu.au
† These authors contributed equally to this work.

Abstract: Artificial intelligence (AI) encompasses the development of systems that perform tasks typically requiring human intelligence, such as reasoning and learning. Despite its widespread use, AI often raises trust issues due to the opacity of its decision-making processes. This challenge has led to the development of explainable artificial intelligence (XAI), which aims to enhance user understanding and trust by providing clear explanations of AI decisions and processes. This paper reviews existing XAI research, focusing on its application in the healthcare sector, particularly in medical and medicinal contexts. Our analysis is organized around key properties of XAI—understandability, comprehensibility, transparency, interpretability, and explainability—providing a comprehensive overview of XAI techniques and their practical implications.

Keywords: artificial intelligence; explainable AI; trust in AI; healthcare AI; AI interpretability; AI transparency; XAI properties

check for updates

Citation: Hamida, S.U.; Chowdhury, M.J.M.; Chakraborty, N.R.; Biswas, K.; Sami, S.K. Exploring the Landscape of Explainable Artificial Intelligence (XAI): A Systematic Review of Techniques and Applications. *Big Data Cogn. Comput.* **2024**, *8*, 149. https://doi.org/10.3390/bdcc8110149

Academic Editor: Alberto Fernandez Hilario

Received: 23 August 2024
Revised: 19 September 2024
Accepted: 27 September 2024
Published: 31 October 2024

1. Introduction

Recently, explainable artificial intelligence (XAI) has become a critical topic of discussion in the artificial intelligence (AI) research and development community. The need for XAI has become more pressing as AI systems are becoming increasingly sophisticated and complex [1]. XAI aims to provide a transparent and interpretable understanding of the decision-making processes of AI systems, particularly in cases where the decisions made by these systems are critical or impactful.

Several factors are hindering the development and adoption of AI in our society. One of the most significant is the growing concern over the ethical implications of AI. As AI systems become more widespread and influential, there is a growing awareness of the potential risks associated with their use. These risks include algorithmic bias, data privacy, and the lack of human oversight in decision making [2]. XAI is crucial for mitigating these risks and ensuring that AI systems are used ethically and responsibly. Another factor driving the development of XAI is the increasing demand for accountability and transparency in AI systems. As AI systems become more integrated into our daily lives, there is a growing need for people to understand how these systems work and how they make decisions. XAI provides a means of achieving this transparency by enabling users to understand AI systems' decision-making processes and identify potential biases or errors. XAI has several applications across a wide range of industries and domains.

For example, XAI can be used in healthcare to explain the decisions made by medical AI systems and to provide clinicians with a more transparent and interpretable understanding of patient data. In finance, XAI can explain the decisions made by algorithmic trading systems and provide investors with greater insight into the factors influencing their investment strategies. In law enforcement, XAI can be used to provide a more transparent and interpretable understanding of the factors influencing criminal profiling and predictive policing.

Despite XAI's growing importance, several challenges and limitations are associated with its development and implementation as illustrated in Figure 1. One of the most significant challenges is the trade-off between interpretability and accuracy. XAI techniques may sacrifice accuracy to provide a more transparent and interpretable understanding of AI systems, which may limit their effectiveness in certain contexts. Additionally, significant technical challenges are associated with developing XAI techniques that can be applied to a wide range of AI systems and domains. To address these challenges, future research in XAI should focus on developing new techniques and tools to provide a more transparent and interpretable understanding of AI systems while maintaining their accuracy and performance. This research should involve interdisciplinary collaborations between computer science, ethics, law, and social sciences to ensure that XAI is developed and deployed responsibly and ethically.

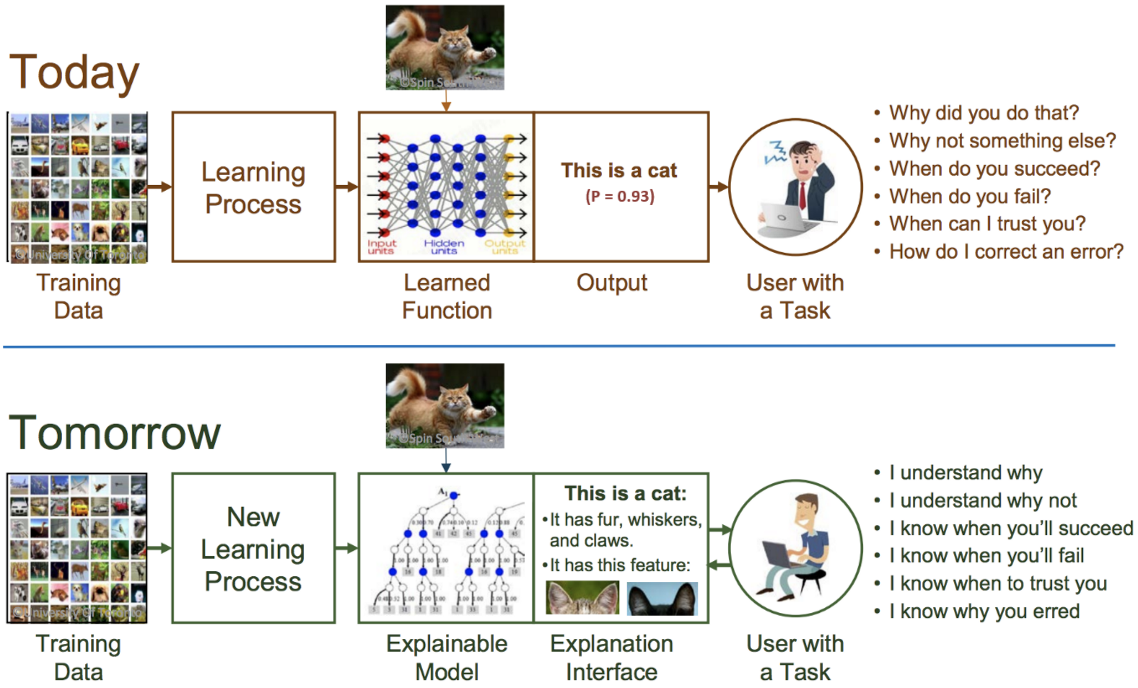

Figure 1. Changes in normal and advance results.

However, several factors, including the growing concern over the ethical implications of AI and the increasing demand for transparency and accountability in AI systems, drive the development of XAI. Despite the challenges and limitations associated with XAI, it represents a significant step towards creating more trustworthy and reliable AI systems that can be used for the benefit of society.

Contributions and Layout

Most of the existing XAI-related reviews talk about the limitations of certain domains, such as ML models and algorithms, and often lack the discussion related to the core principles of XAI. This work provides the following contributions for those having an interest in explainable AI to further their work:

1. We conducted a comprehensive review to analyze and synthesize research, ideas, and considerations in the black-box fields of AI and XAI.
2. We present a meta-analysis of methods to improve the transparency of existing and emerging AI technologies.
3. We create comprehensive instructions and visual aids, such as graphs and tables, to facilitate additional study and application in this field.

The format of this document is as follows: background knowledge on the subject is provided in Section 2. We have discussed our research methodology in Section 3. The finding of our review is presented in Section 4. These results are further analyzed to find more insight and these are presented in Section 5. The study's shortcomings are outlined in Section 6, along with recommendations for further study paths. Lastly, Section 7 brings the study to a close.

2. Background

Since its inception in 1950, artificial intelligence (AI) has contributed significantly to the advancement of emerging technologies for the benefit of humanity. Over time, as the domain gained recognition and familiarity in higher societies and studies, several complications and questions also surfaced. People specifically questioned the stability and dependability of their decisions. Although artificial intelligence was well known, many people were unaware of its workings or reliability Figure 2.

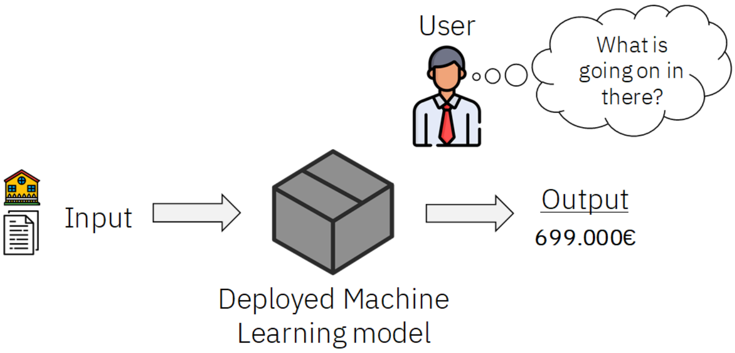

Figure 2. A sample ML model's calculation.

This ambiguity highlights how crucial it is to explain the decisions and actions taken by artificially intelligent systems. Due to this requirement, the need for explainability in AI developed, and explainable AI (XAI) was first presented to the field in 1970. Through revealing decision-making processes, XAI seeks to improve transparency and human comprehension of these systems. This openness in turn encourages trust in the results of artificial intelligence. Furthermore, the relevance of interpretability and transparency in AI systems has been highlighted by the increased complexity of predictive models and their incorporation into crucial systems.

In short, explainable AI is an emerging machine learning method for addressing unclear AI systems decisions. It is a set of approaches and strategies that enable human users to comprehend and trust the choices and outputs generated by machine learning algorithms. It provides the results or produce that we, the people, can see. Moreover, it can define an artificial intelligence model's expected impact and potential biases.

Providing clear explanations fosters trust between users and AI systems. Explainable AI (XAI) not only illustrates how an AI model functions but also highlights its accuracy, fairness, and transparency. This clarity is crucial for organizations to build confidence in AI models as they are integrated into decision-making processes. Trust in expert judgments is particularly vital in critical areas such as healthcare. XAI helps users understand and articulate the workings of an AI system, as illustrated in Figure 3.

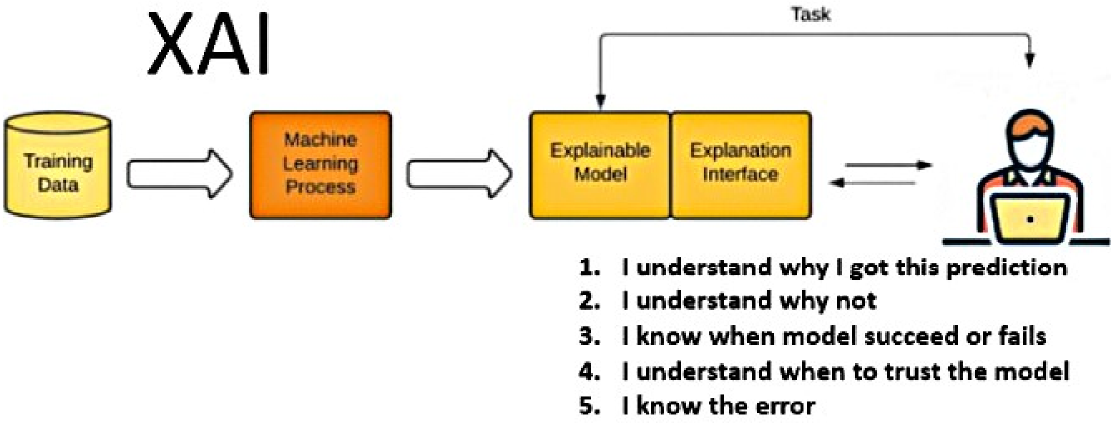

Figure 3. Future of explainable AI.

These explanations of computer-based intelligence models can improve the legitimacy, smoothness, solidity, and dependability of frameworks [3]. Although explainable AI was developed between 1955 and 1960, its study began only in 2004. Several research studies briefly mention two types of model detectors, the glass or white box and the black box. The white box has a straightforward model with clear structures, a manageable number of parameters, and interpretation without the need for additional explanation techniques such as decision trees or linear regression. In contrast, complex models, such as deep neural networks (DNNs) or support vector machines (SVMs), are described as "black boxes" because, despite knowing their structure and weights, their behavior is still difficult to comprehend. These models could have thousands, millions, or even more parameters (weights).

Norbert Wiener first mentioned this black-box idea in 1961 [4]. Norbert categorizes all these unknown systems in a model as a black box. It is a system whose inputs and operations are not visible to the user or other interested parties. More importantly, it is an impenetrable system that is generally difficult for data scientists, programmers, and users to interpret. XAI comes as an emerging field in machine learning to address the black-box decisions. It disclosed the internal functionality of black-box models like linear or logistic regression with great success, Figure 4. Moreover, it raised the number of interpretable models like decision trees, naive Bayes, and others [5].

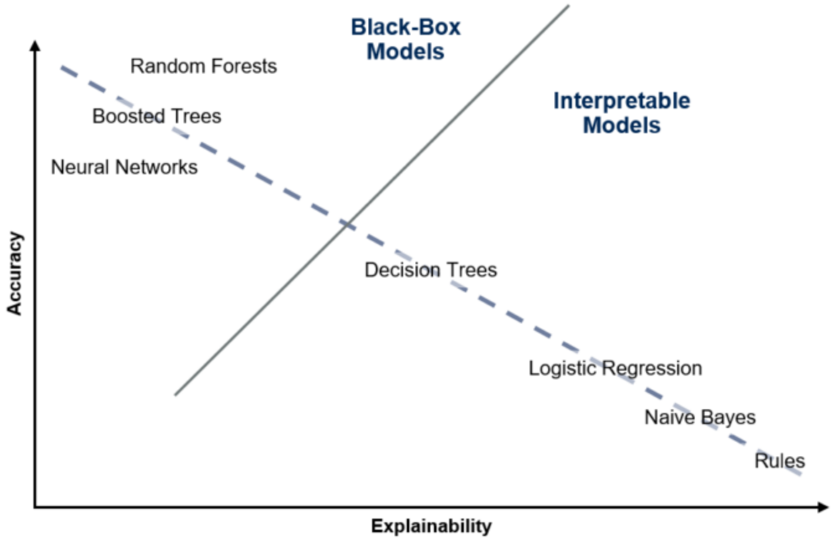

Figure 4. Interpretable model and black-box model in artificial intelligence.

In the journey of making machine learning understandable and trustworthy, the most important area is to disclose the black box section of the model, tree, or algorithm. However, there is no such specific property to convert a black box into a white box. In several studies, interpretability, sometimes transparent and sometimes post hoc denotes the XAI property. In some cases, explainability and interoperability are considered as being the same, although they have subtle differences. When a user lacks access to the system's basic structure, interpretability guarantees the individual can still understand the implications of what it predicts. It aids in bridging the knowledge gap between users and difficult outcomes. Explainability offers more thorough explanations or rationale for particular results, especially in intricate neural network models. It provides a specific decision by employing techniques that shed light on its decision-making procedure.

Specifically, according to [6] XAI algorithms are created with the three concepts of explainability, interpretability, and transparency in mind. While transparency is the strategy of basic openness, such as its framework or parameters, rendering its workings evident, interpretability is the degree to which people can easily grasp a model's judgments. Furthermore, Loyola [7] highlights accountability, openness, and justice as essential standards in the development of machine learning, highlighting the significance of these values.

2.1. Transparency

Transparency shows the clearness of a model, Figure 5. It helps to manage, find, enhance, and defend a model. It shows that the result of an AI model can be appropriately clarified and conveyed. It additionally presents the internal course of a transparent model that extricates model boundaries from preparing information [8]. It even produces marks from testing a piece of information that portrays and is spurred on by the methodology originator. Moreover, transparent AI is reasonable AI that permits people to see whether the models have been entirely tried and bode well so that they can comprehend why specific choices are made [9]. In the AI model, transparency alludes to the capacity to notice the cycles that lead to dynamics inside models. It takes care of this issue by effectively utilizing interpretable models [10].

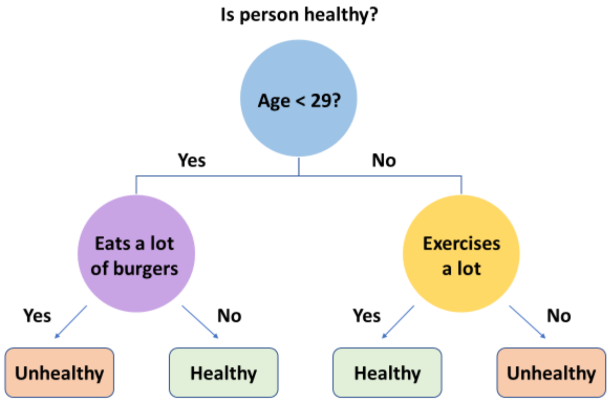

Figure 5. Clearness in XAI model.

2.2. Interpretability

Interpretability is defined as the ability of a model to infer cause and effect, Figure 6. It explains the potential for human-understandable comprehension of the ML model. The capacity to reliably forecast a model's output without knowing the underlying causes is known as interpretability [11]. Machine learning algorithms were notorious for being "black boxes", meaning that it was difficult to understand their inner workings and to share the insights they generated with stakeholders and regulatory authorities [12]. Humans are capable of comprehending the processes and outcomes with ease because of interpretability [13].

However, Lipton et al. distinguish between two categories of interpretability: post hoc and transparent, where transparency is a quality that can be identified before the start of the training, as stated by Lipton et al. Further, they put forth three different but connected queries concerning algorithm transparency, decomposability, and emulation. At the same time, Lipton et al. pose four questions on post hoc interpretability [14]. As they mentioned, text explanation, visualization, local explanations, and explanation by example refer to things that can be learned from the model after training has finished.

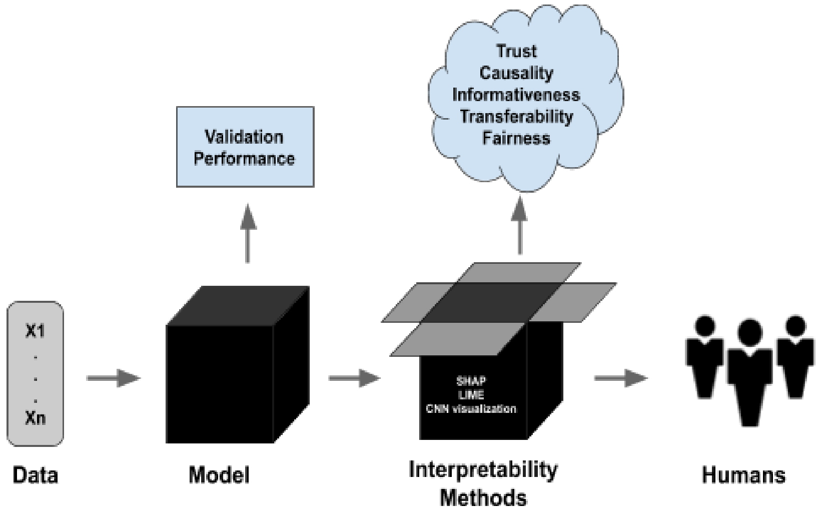

Figure 6. Interpretability in XAI model.

2.3. Explainability

Understanding what a node means and how important it is to the performance of the model is known as explainability. It assists with clarifying information, expectations, and calculations. It is an important concept, yet, any associated definition is not available. It considers an assortment of provisions of the interpretable spaces that have contributed to a guide to delivering a choice (e.g., arrangement or relapse). On the off chance that calculations meet these necessities, they give a premise to defending choices, following and, in this way, confirming them, working on the computation, and investigating new realities. It is an understandable extent where the feature values of an instance are related to its model prediction, Figure 7. In basic terms, it is the understanding of the question of why this is happening [15]. Explainability is very important as it helps analysts understand system outputs simply and quickly while overcoming false positives.

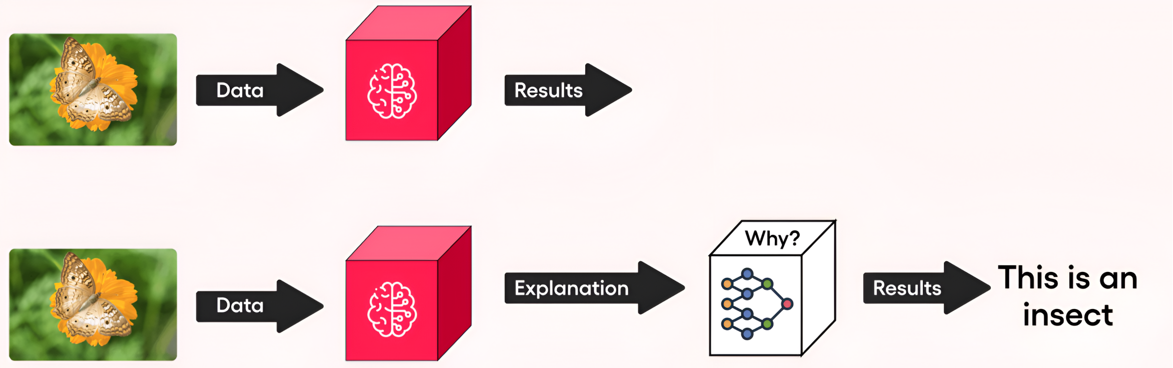

Figure 7. Explainability in XAI model.

Explainability in machine learning (ML) clarifies how a model processes input to produce output, addressing the black-box problem by making models more transparent. For instance, in a healthcare model predicting a patient's likelihood of having a specific disease, ML explainability is categorized into global and local aspects. Global explainability involves providing an overview of the entire learning algorithm, including details about the training data, suitable applications, and potential limitations or misuse of the algorithm. In contrast, local explainability focuses on explaining individual decisions made by the model, helping users understand why a particular prediction or outcome was reached.

XAI is presently a spotlight strategy of making ML models appreciated; it is also mentioned as Interpretable AI or Transparent AI. XAI disclosed several ML solutions, methods, and algorithms through its current status, even though the future objectives of XAI are clear, Figure 8. XAI seeks credibility with a range of stakeholders, such as data scientists who design and implement these technologies, and users who might depend on discoveries produced by XAI.

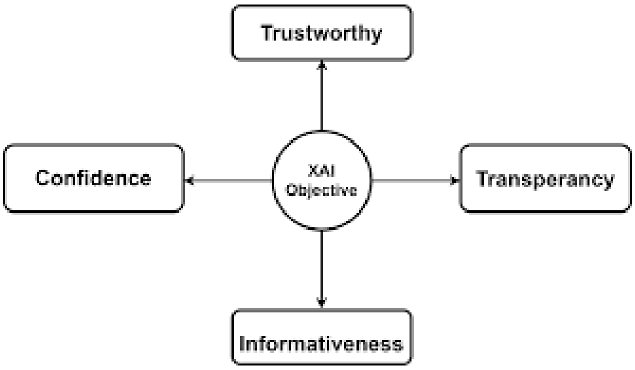

Figure 8. Objectives of XAI.

Moreover, XAI enhances confidence in a domain by providing a clear view of each process step. It offers comprehensive information by explaining each piece of data and ensuring transparency in the reasoning behind specific results. The ultimate goal of this approach is to create a robust model with explainable logic. Additionally, this clarification should be designed to align with human reasoning and dependencies, making the explanations more relatable and understandable.

Presently, explainable AI has its own ethical and explainable principles where the upright section controls inclusiveness and accountability; the understandable division discusses fairness, transparency privacy, and security. Both sections work together for restraining reliability and safety, Figure 9.

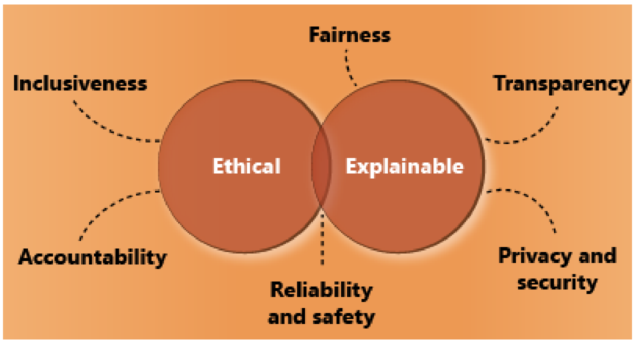

Figure 9. Principles of XAI.

2.4. Research Questions

We conducted our investigation by reviewing prior research on XAI applications. This involved analysing their findings, limitations, and methodologies before summarizing their contributions. The subsequent research questions help refine the focus of our study and enhance its effectiveness.

1. What are the domains that benefited from explainable AI?
2. What are the most essential properties for each domain with respect of explainable AI?

3. What are the available algorithms and frameworks for explainable AI? Are they self-explanatory or not?
4. What are the domain-specific hindrances for explainable AI?

By addressing these research questions, the study aims to explore key aspects of XAI, including its most beneficial applications, core characteristics, prevalent frameworks, methodologies, and significant challenges to its wider adoption. A thorough analysis of recent research will be conducted to identify relevant patterns and insights to address these issues. The following chapter outlines the methodology for compiling and evaluating the body of XAI research. This includes a detailed description of how papers were selected based on specific keywords and criteria, ensuring a focused and comprehensive review of existing work. This systematic approach will enable us to address the research questions and uncover critical findings within the field of XAI.

3. Research Methodology

We addressed our study questions through a systematic literature review (SLR), employing an iterative approach across the planning, conducting, and reporting phases. This iterative process allowed us to rigorously evaluate each stage of the SLR. During the planning phase, we established inclusion and exclusion criteria, search strategies, and study objectives. In the conducting phase, we extracted data, assessed articles using predefined keywords and criteria, and performed searches across multiple databases. Finally, in the reporting phase, we aggregated the data, analysed the results, and provided interpretations and conclusions. This meticulous and iterative approach ensured that we thoroughly addressed our study questions and achieved comprehensive and reliable findings.

3.1. Selection of Primary Studies

A comprehensive search for published systematic reviews and meta-analyses was conducted across several platforms, including IEEE Xplore Digital Library, ScienceDirect, SpringerLink, ACM Digital Library, Google Scholar, ProQuest, and PubMed.

We used the following keywords for the search: Explainable AI, Machine Learning, black box, transparency, interpretability, deep learning, neural networks, reinforcement learning, trustworthiness, malicious AI, systematic review, industry, decision trees, models, and algorithms. Initially, the search was limited to publications up to 2023, but it was later extended to include papers from 2024. Additionally, we traced the references of all retrieved reviews and meta-analyses to identify any potentially overlooked studies

3.2. Inclusion and Exclusion Criteria

The selection process for included studies involved a two-step screening approach. Initially, titles and abstracts were reviewed to conduct a preliminary assessment. Studies that met the initial criteria were then subjected to a full-text evaluation. To be deemed eligible, a review or meta-analysis had to fulfil specific inclusion criteria. This included a focus on explainable AI (XAI), its models, algorithms, decision trees, or systems, and publication in reputable international journals. Both systematic reviews/meta-analyses and surveys addressing these topics were considered for inclusion.

In essence, the article had to include information on XAI technologies, give empirical data about the application and usage of XAI, and be a peer-reviewed work published in a conference proceeding or journal, or part of a book or report. Another point we must mention is that in the case of newly published papers, we skip the highest citation and instead go for the most famous journal. Focusing on six years, we excluded or included articles based on the following characteristics mentioned in Table 1.

To be included in our review, an article must meet the following criteria:

- **Content Relevance:** The article must provide detailed information on XAI technologies and present empirical data regarding the application and usage of XAI.
- **Publication Type:** The work must be peer-reviewed and published in a reputable conference proceeding, journal, book, or report.

For newly published papers, we prioritize publications from well-regarded journals over those with the highest citation counts. Our review focuses on articles published within the last six years. Articles were selected or excluded based on the criteria outlined in Table 1, which details the specific characteristics considered during the inclusion process.

Table 1. Criteria for inclusion and exclusion.

Exclusion	Inclusion
Did not address XAI domain (i.e., model, tree, algorithm, system)	From 2018 to 2024
Non-English articles literature	Peer-reviewed and open access
Duplicate article (i.e., title, author, discussion, and everything)	Must be from XAI domain
It was not a journal article or generic	Relevant, most cited
Not related to explainable ai	Number of citations; although, new papers have fewer citations

3.3. Selection Results

Initially, a total of 298 studies were identified from the selected platforms. After removing duplicates and ineligible studies, 159 papers remained. Applying the inclusion and exclusion criteria further reduced the number to 98. However, due to restrictions, eight additional records were excluded, leaving 90 papers for detailed review. After a thorough examination and a second assessment against the inclusion and exclusion criteria, 27 studies were retained.

Subsequently, an additional 12 articles were identified through snowballing: 5 through forward citation tracking and 7 through backward citation tracking. Although our initial review period covered material from 2018 to mid-2023, we included five journal papers from 2024 to ensure the comprehensiveness of our review.

In total, 44 papers were initially included in our literature overview. Following suggestions from reviewers, we added three additional journals, including three from 2023 and one from 2022. Additionally, five new articles from 2024 were incorporated to ensure the review is up-to-date. As a result, the final number of papers in our comprehensive review is 53.

3.4. Quality Assessment

To ensure the quality of the primary studies included in our review, we employed a quality assurance checklist consisting of nine targeted questions. These questions were designed to evaluate the relevance and rigour of each selected article. The checklist addressed key aspects such as the reliability of experimental results and potential sources of bias within the research.

The quality control questions played a critical role in assessing the relevance and validity of the studies in relation to the research objectives. Each study was evaluated against these criteria to determine its suitability for inclusion. As a result of this thorough evaluation process, no studies were found to be inadequate and excluded from the systematic literature review (SLR).

Quality Assessment Questionnaire (QAQ)

1. Is the paper relevant?
2. Is the research article peer-reviewed?
3. Is it original research or a review?
4. Does it explain any XAI application?

5. Does it use any dataset?
6. Is the quality of the dataset used satisfactorily?
7. Does it follow appropriate methods or approaches?
8. Is the presentation or organization of the paper precise?
9. Does the study add value to a digital library?

3.5. Data Extraction

Using the previously specified quality assurance questions, the data extraction procedure was examined in all the articles. Following the completion of the quality evaluation stage, all of the study's data were extracted, classified, and then kept in an Excel file. A graphical representation of the PRISMA filtering procedure illustrates the number of papers chosen at every phase of the procedure as well as the percentage of papers that were lost from the first keyword searches on every platform to the ultimate primary study selection, Figure 10 [16].

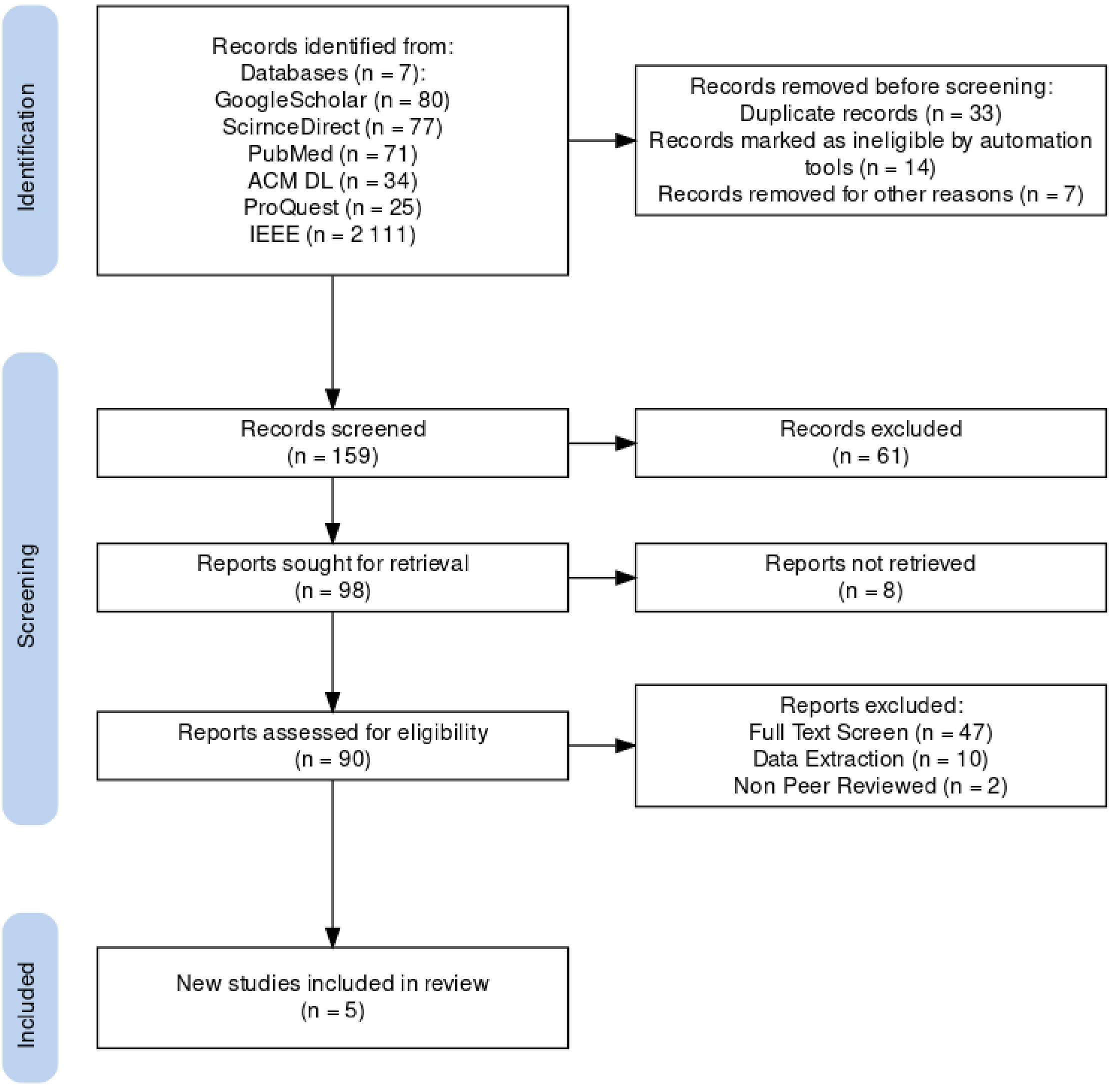

Figure 10. PRISMA diagram of primary paper selection process.

3.6. Data Analysis

We obtained the data from the qualitative and quantitative data categories to answer the research questions. We also carried out a meta-analysis on papers that were going to undergo the last round of data extraction.

3.6.1. Publications over Time

The concept of XAI is mainly related to the black box, which has nonexplainable parts that make researchers interested. Throughout the years, there has been some mentionable work on the basic ideas and limitations of XAI. Figure 11 shows the number of primary studies published each year.

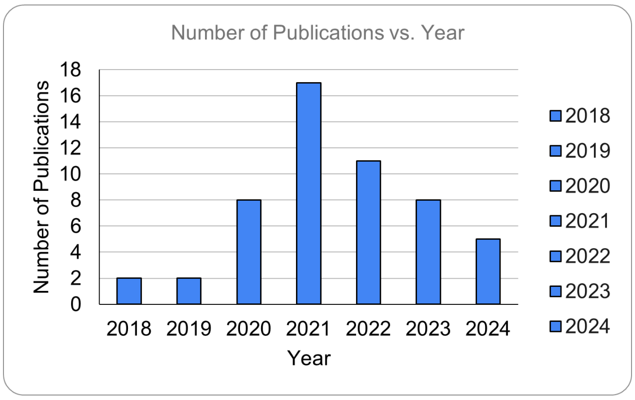

Figure 11. Number of primary studies published over time.

3.6.2. Significant Keyword Counts

Our final 44 studies were checked, which gave some common significant keywords for this review. It was discovered that terms like "black box", "interpretability", "trustworthiness", and "neural network" appeared frequently in the literature. Significant allusions to concepts such as "fuzzy logic", "data fusion", and "graph mining" were also found in several studies.

4. State of the Art

We chose articles with a thorough assessment by QAQ from several domains or research fields. In total, 68% of our final research was journal articles. There were also 18% generic, 5% in progress or conference proceedings, 2% taken from different book sections, and 7% from reports. In total, 58% of the selected articles were systematic reviews that explained the importance of the different XAI systems' principles Figure 12.

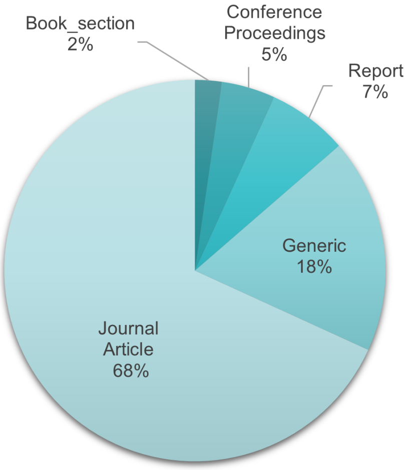

Figure 12. Types of record included in the study.

At the same time, about 42% of the articles only show original research on XAI. Moreover, our study found that 34% of the articles are from the AI domain, as most of the research talks about artificial intelligence and its related areas. Moreover, information fusion, electronics, electrical engineering, sensors, imaging, knowledge engineering, computer science, medical informatics, and decision making benefited from XAI or were affected by XAI. Figure 13 shows all research areas according to the study.

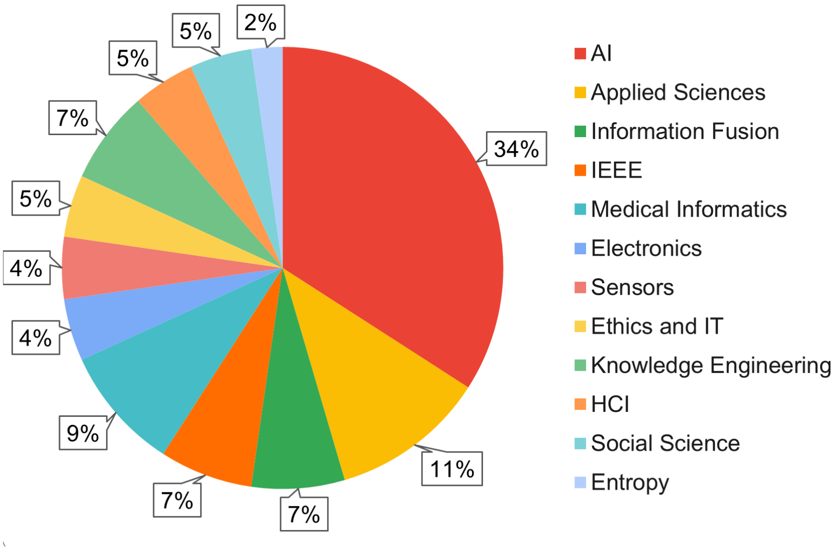

Figure 13. Included research areas.

The study found several available methods, algorithms, and models, and it is essential to use XAI in these. In total, 55% of the reviews focused on neural networks, and DNNs and CNNs were at the top of this line. Moreover, 52% of these reviews discussed machine learning models, algorithms, and classification. Moreover, 13% of the studies mentioned healthcare's CDSSs (clinical decision support systems) or DSSs. We even noticed the use of XAI tools SHAP and LIME in 23% of the systems to increase the acceptability of all these AI systems.

We also found that the healthcare sector is the largest field to give more concern because almost 54% of the research was from this subdomain. In addition, a few articles were from the industrial, informatics science, social science, soil testing and earthquakes, expert system, and bioinformatics sectors, Figure 14.

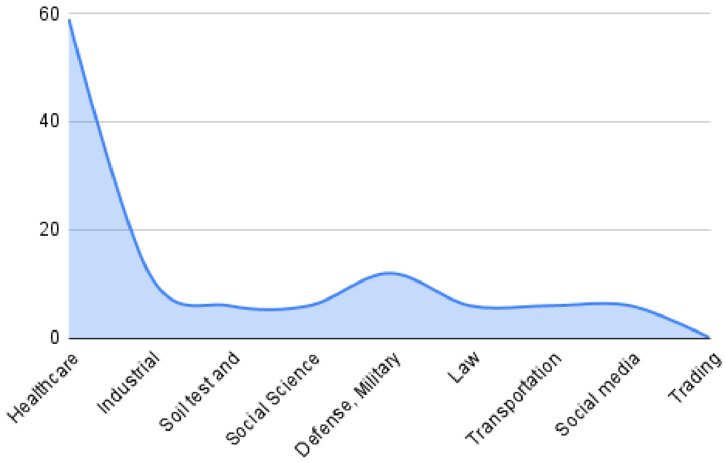

Figure 14. Most affected fields in recent time.

Here, the study covered a diverse range of applications and research areas, providing valuable insights into this rapidly evolving field and discussing various techniques such as Topic Modeling, Knowledge Band, Neighbor Band, Graph Band, and Rule Mining. Among these, clinical decision support systems (CDSSs) stood out, representing approximately

9.1% of the applications mentioned. Another significant area explored was Computer Vision, including CNN-LIME and MLP, constituting around 9.1% of the applications. Further acknowledged were data fusion and image analysis, which together accounted for about 4.5% of the applications. The survey covered a wide range of fields when it came to research disciplines and journals, including computer science, medical informatics and decision making, imaging, and sensors. The above domains stand for the areas of research where XAI applications are common. With its appearance in roughly 6.8% of the cases, medical informatics and decision making had the highest proportion of these. Moreover, imaging and computer science each constituted around 4.5% of the research areas mentioned. Nonetheless, the study shed light on the significance and diverse scope of explainability in the realm of artificial intelligence, encouraging further exploration and advancements in this crucial domain.

Furthermore, this study discovered several influential mentions of accountability, comprehensibility, adaptability, causality, predictability, and accuracy. However, the only principles of XAI we knew were explainability, transparency, and interoperability in the pre-review period. Moreover, these are still the most mentioned principles of XAI. About 65% of our selected papers speak of explanations. At the same time, the necessity for transparency (52%) and interpretability (58%) have been mentioned many times in multiple reviews.

4.1. Discussing the Findings

In our research, we systematically reviewed 53 selected papers on explainable artificial intelligence (XAI) from 2018 to 2024. The papers covered a wide range of topics related to XAI, including surveys, applications, algorithms, fault diagnosis, interpretability, and machine learning, among others. We aimed to synthesize the key findings and insights from these papers to provide a comprehensive overview of the state of XAI research in recent years. In this section, we will discuss the overall findings of our review and highlight the key themes and trends that emerged across the selected papers.

4.1.1. Benefited Domains

After reading the chosen papers, we noticed that XAI was receiving more and more attention across a wide range of areas, with a variety of potential uses. We concentrate on the domains that XAI approaches have been found to benefit in this paper.

The advantages of XAI in the healthcare industry have been underlined in several publications [12,17–19]. To help healthcare professionals make better decisions, XAI approaches have been applied, for instance, to increase the interpretability of CDSSs [19]. Furthermore, XAI has been used to analyse medical imaging data, where it has been demonstrated to increase diagnostic precision and decrease mistakes [20–22]. XAI approaches have the potential to improve healthcare systems' accountability and transparency, which could ultimately lead to better patient outcomes. In the area of geoscience, XAI has also been proven to offer considerable advantages. For instance, XAI approaches have been applied to better predict natural disasters and evaluate the outcomes of analyses of remote sensing data [23]. Additionally, geological data analysis and improved mineral exploration accuracy have been accomplished using XAI [21].

Cybersecurity has also benefited from the application of XAI techniques. XAI can aid in system vulnerability detection, mitigation, and better detection of and retaliation to cyberattacks [24]. Additionally, XAI can contribute to privacy preservation by enabling data analysis while concealing sensitive information [25]. Digital pathology, predictive maintenance, and the Internet of Things (IoT) are other sectors that have been mentioned as possible benefactors of XAI approaches [23,26,27]. For the precise diagnosis and treatment of diseases, XAI can help to increase the accuracy and effectiveness of digital pathology analysis [26]. In the field of predictive maintenance, XAI can assist in the early detection and diagnosis of equipment faults, which can result in significant cost savings and increased safety. Furthermore, through the analysis of intricate patterns of information and the detection of abnormalities that may point to possible cyber threats, XAI has demonstrated promise in boosting the reliability and safety of IoT systems [28].

4.1.2. Algorithms and Frameworks

After a thorough review, we found a significant focus on XAI-related algorithms and frameworks. These frameworks and techniques are designed to improve the machine learning (ML) models' interpretability and transparency, making them simpler for people to understand. Several articles that focused on developing frameworks and algorithms for XAI were discovered throughout our investigation. For example, reference [29] summarises the principles and ideas of XAI, whereas [30] proposes a framework for developing transparent and intelligible models. A full examination of several XAI approaches and algorithms is provided in [31], and the use of fault-tolerant solutions in XAI systems is covered in [32]. Furthermore, we found that the need for XAI framework development is growing. Another paper even proposes a key framework for designing precise and comprehensible recommender systems [12,33] and provides a framework for developing interpretable reinforcement learning (RL) models. Moreover, J. Zhou et al. propose a comparable approach that allows users of the model to engage with and modify XAI representations [34].

4.1.3. Self-Explanatory

We found that explainable artificial intelligence, or XAI, has drawn a lot of interest from the scientific community over time based on our assessment of the selected publications. Our paper intends to present a comprehensive assessment of the numerous uses of XAI in IT research works, focusing on a diversity of approaches and uses.

1. One of XAI's key characteristics is that it is self-explanatory. We assessed a large number of studies to highlight the importance of developing XAI techniques and models since they can provide concise and precise reasons for the decisions made by AI. As an illustration, XAI emphasises the necessity for interpretability in AI systems, which can help increase user confidence and make it simpler to identify biases or defects in the system [34].
2. Another key element of XAI that is frequently brought up in the publications we analysed is transparency and the importance of developing transparent AI models that can provide concise explanations for their decisions [30]. Similarly, reference [35] underlines the requirement for transparency in AI models to ensure their dependability.
3. Numerous publications also discuss different techniques and XAI algorithms. Reference [31] provides a summary of the various XAI algorithms that are employed, including rule-based systems, decision trees, and neural networks. Several describe how neural networks can be used to employ explanations for their decisions, whereas others analyse the use of several modalities for creating explanations [21,29].

4.1.4. Most Essential Properties

Numerous essential XAI traits have been identified in the selected publications, the majority of which concentrate on enhancing user friendliness or providing transparent results or applications. One of the most important qualities of XAI systems is transparency, or the ability to describe their decision-making processes understandably. Transparency is seen as a critical element of XAI in many industries, including healthcare, as it is necessary to make sure that patients and clinicians can understand the reasoning behind a particular diagnostic or treatment decision. Other important qualities of XAI systems include interpretability, which refers to the ability to understand how a system arrived at a particular decision, and accountability, which refers to the ability to track and assign blame for an XAI system's decision [29,36,37]. Other essential XAI system attributes include fairness, robustness, and trust, particularly in sectors where bias and errors may have significant negative effects.

4.1.5. Major Contributions

A comprehensive overview of methodologies and applications in explainable artificial intelligence (XAI) and its critical importance of interpretability and openness in machine learning is presented in [34]. Meanwhile, the XAI privacy-preserving study focuses specifically on evaluating the transparency and interpretability of XAI systems within clinical decision support contexts, with a notable emphasis on healthcare applications [38].

The XAI transparency study introduces a novel paradigm for assessing the transparency of XAI systems, highlighting the significance of the information provided to users [18]. Conversely, the XAI-RL study delves into the challenges associated with integrating XAI into reinforcement learning (RL) systems, offering valuable recommendations for streamlining RL models while preserving interpretability [12].

Moving onto the medical domain, the collection of medical papers delivers a thorough examination of both model-specific and model-independent interpretability methodologies in machine learning, with a strong focus on their practical implications in clinical decision support and mental health analysis [29,39]. Furthermore, the XAI institutional context study proposes a paradigm for evaluating the effectiveness of XAI within clinical decision support systems, shedding light on its institutional implications [40].

Lastly, the XAI quantified explainability research introduces an innovative approach for leveraging XAI technologies to identify anomalies within vast datasets, demonstrating the versatility and potential impact of XAI in data-driven anomaly detection tasks [41]. Overall, these studies collectively contribute significantly to advancing XAI by exploring various dimensions, methodologies, and applications across diverse domains while advocating for transparency, interpretability, and effectiveness in machine learning systems.

4.1.6. Hindrances/Barriers

1. **Lack of interpretability and transparency:** We have found several studies that address the difficulties with or barriers to the adoption and use of XAI methodologies in a variety of disciplines related to real life. One of the main challenges identified is the lack of interpretability and transparency of complex machine learning models. It can be difficult for people to understand how the algorithms produce their predictions because many machine learning models are frequently referred to as “black boxes”. Lack of interpretability may make it difficult for these models to be adopted and accepted in real-world applications, particularly in fields where the stakes are high, like healthcare and finance [18].
2. **Data privacy and security:** Another key obstacle to the adoption of XAI techniques is the need for enhanced data privacy and security. As AI systems are used more frequently and massive amounts of sensitive data are collected, there is growing concern about data breaches and the exploitation of personal information [25]. As a result, XAI algorithms have been developed that prioritize the privacy and security of sensitive data while yet enabling accurate and clear projections.
3. **Contextual knowledge:** Along with the challenges, we have also found many research articles discussing the need for additional information or contextual knowledge to enhance the interpretability and explainability of machine learning models. In this sense, it is essential to consider the patient’s medical history, symptoms, and other relevant criteria in addition to the model’s output when making medical decisions [22]. The interpretability of the models can be improved in the field of natural language processing by including linguistic and domain-specific knowledge in a manner similar to this [35]. The next section includes the benefits and limitations we found in the studies.

4.1.7. Advantages of XAI Techniques

Recent developments in XAI have looked into how to improve model interpretability by using computational models like transformers, GANs, and VAEs. These models are currently being used to produce counterfactual justifications and illustrations, which aid in locating different inputs that might affect a model's forecast [42,43]. This methodology enhances the comprehensibility of artificial intelligence systems and facilitates the visualization of decision-making procedures, particularly in intricate models. Furthermore, the emergence of robust machine learning models such as XGBoost, which are sometimes regarded as "black-box" systems, has prompted the creation of tools like TreeExplainer and SHAP [44,45]. By determining feature importance and providing individual prediction explanations, these techniques efficiently deconstruct intricate tree-based models, increasing XGBoost's transparency without sacrificing accuracy [45].

Thus, a comparison of the recent famous XAI tools is included in Table 2. By providing a transparent and accessible perspective of the decision-making process, XAI techniques boost the trustworthiness and reliability of complicated models [18,30,46]. Domain experts can interpret and assess the models using XAI techniques, as well as identify any biases or defects that may affect how well they perform [32,36,47]. XAI techniques can enhance user experience and hasten decision-making processes in a range of disciplines, such as medical diagnosis, earthquake prediction, and event detection [23,29,31].

Table 2. Comparison of SHAP and LIME based on Key Criteria.

Criteria	**SHAP (Shapley Additive Explanations)**	**LIME (Local Interpretable Model-Agnostic Explanations)**
Model Type	Model-agnostic, works with any model	Model-agnostic, works with any model
Interpretability	High (based on game theory, consistent)	Moderate (approximation model may not be perfect)
Accuracy	High (consistent with predictions)	Moderate (approximate model can sometimes be less accurate)
Computation Time	Moderate to high (especially for large models)	Fast (but less accurate for larger models)
Local/Global	Both (global and local explanations)	Local (only explains individual predictions)
Ease of Implementation	Moderate (requires more computational resources)	Easy to implement and use

4.1.8. Limitations of XAI Techniques

Although XAI approaches have made great progress in providing insights into intricate machine learning models, they are still severely limited in terms of technology, particularly when used in practical applications. Even though they work well, techniques like SHAP and DeepLIFT are computationally intensive and thus hard to scale for massive databases or real-time applications [42,48]. Additionally, although model-agnostic methods like SHAP and LIME might clarify a variety of models, they frequently fall short of model-specific methods in terms of accuracy or generalizability. The emphasis on local interpretability presents another difficulty since many XAI techniques succeed in explaining specific predictions but fall short of offering a more comprehensive, all-encompassing knowledge of the model's behaviour [49,50]. This lack of comprehensive awareness is especially troublesome in vital domains where comprehension of the full decision-making procedure is essential, such as the healthcare industry.

- Since not all models and datasets may be compatible with XAI techniques, it is critical to consider the relative merits of accuracy and interpretability [19,51].
- The definition of an adequate explanation may not be universally agreed upon; therefore, XAI techniques may not provide complete solutions [52,53].
- The accuracy of XAI strategies for identifying and correcting biases and errors may be impacted by the quality and accessibility of the data [54,55].

XAI methods not only present these technological difficulties but also significant ethical issues. One major problem is explanation bias, which occurs when explanations inadvertently reinforce or reflect hidden prejudices in the framework. Furthermore, some XAI methods run the risk of disclosing private information, notably in counterfactual justifications that could disclose variables like race or ethnicity in the data [56,57]. When these techniques are overused, they can also create the appearance of transparency by giving stakeholders excessive confidence in models that have been oversimplified. Additionally, there is a chance of "explanation manipulation", in which unfair or unlawful conduct in simulations might be covered up with carefully chosen explanations [56,57]. Resolving these moral and technological issues will be essential to the appropriate and efficient application of XAI as it develops.

4.1.9. Why the Need for XAI?

We discover that explainable AI (XAI) strategies are becoming more and more necessary in the IT business as we delve deeper into this discussion to find solutions. An AI system's capacity for XAI refers to how well it can communicate the rationale behind its choices and suggestions. By doing this, XAI can improve the usability and trustworthiness of AI systems while also increasing their transparency, accountability, and dependability [19,58]. Several of the studies in our analysis have a particular focus on the use of XAI in the healthcare sector, where the accuracy and dependability of AI systems are critical [22,29,36,59]. Regarding AI applications in the healthcare sector, one problem that XAI can assist in resolving is the need for accurate justifications of medical diagnoses and treatment recommendations. Other publications that are a component of our research cover the use of XAI in many other domains, such as geoscience, predictive maintenance, and privacy-preserving AI [21,38,60]. These studies demonstrate the wide range of applications and the promise of XAI techniques in numerous different disciplines and applications.

5. Analysis

We identified several XAI methods designed to enhance transparency and interpretability in machine learning models [18,30]. These methods include rule-based reasoning, post hoc explanations, and model-specific justifications. Notable frameworks such as Integrated Gradients (IGs), Shapley Additive Explanations (SHAP), and Local Interpretable Model-Agnostic Explanations (LIME) have been developed to support these techniques. Additionally, machine learning platforms like TensorFlow and PyTorch are increasingly incorporating XAI features. These advancements highlight the growing interest in XAI, which has the potential to improve both the clarity and reliability of machine learning models. Here, we briefly discuss the goals of our research and how the results relate to those goals.

5.1. What Are the Domains That Can Benefit from Explainable AI?

AI has substantially enhanced people's lives. It has an impact on every facet of daily life, little and huge. These days, using AI is essential to fulfilling fundamental human needs. Therefore, a system must be transparent and have an explanation mechanism. Our analysis identifies some of the subjects with the most interest, whose advancement might enhance the comfort and quality of human life. The majority of research has been conducted on AI; hence, it is listed first in this row.

According to our findings, XAI is now much more crucial for the expansion of AI sectors. It is clear from this that popular AI subdomains include machine learning (ML), Multi-Level Marketing (MLM), Multi-Layer Perceptron (MLP), algorithms, classification, Deep Learning (DL), neural networks (NNs), DNNs, convolutional neural networks (CNNs), and recurrent neural networks (RNNs) [32,47,54,61–65]. Furthermore, computer vision, fuzzy logic, knowledge engineering, information science, and other fields are impacted by XAI. Therefore, attempting to advance these sectors can result in a variety of opportunities. Even within XAI models and applications, there is enough available for anyone to benefit. The fields of social science, electronics, computer science, and medical informatics should also be mentioned [53]. Our investigation even found other industries that make use of imaging and sensors. Explainable AI has already benefited or assisted several of these subdomains, Figure 15.

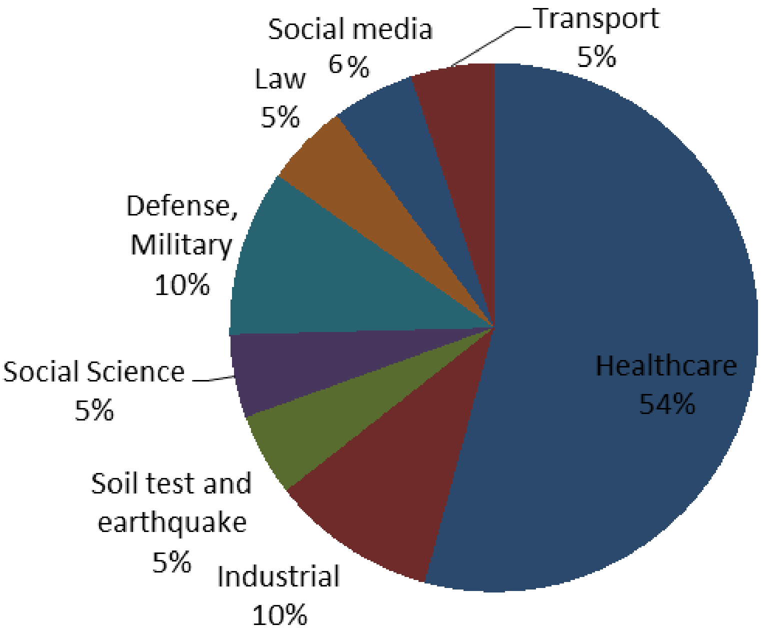

Figure 15. Most affected fields in recent time that can benefit the most.

Through our comprehensive research of XAI-related articles released between 2018 and 2023, we learned that XAI has been utilized in numerous industries to improve the interpretability and transparency of machine learning models. The selection of selected publications makes it obvious that explainable AI (XAI) has generated a lot of interest in a variety of industries, including healthcare, geoscience, finance, and IoT systems. As a result, the proposed domains or subdomains based on our research are shown below:

- **Healthcare:** Our analysis shows that many researchers have encountered difficulties while attempting to understand why explainability is necessary for AI systems. The healthcare sector is among the most prominent where XAI has been shown to have significant promise. XAI has been applied to medical imaging and diagnosis, disease prediction, and medication discovery. The authors proposed a paradigm for developing clinical decision support systems (CDSSs), which assist physicians in making well-informed judgments, in their publication [61]. The framework provided explanations for the decisions the CDSSs made by fusing rule-based reasoning and supervised learning. In a similar vein, a different article on XAI–digital pathology looked at the use of XAI in medical imaging and offered a framework for the use of explainable deep learning in the diagnosis of breast cancer [22]. The framework offers justifications for the choices the model made, which can help doctors choose wisely. The healthcare industry will benefit the most from XAI for medical imaging, biosensors, molecular data, sensing, and electronic medical records (EMRs) as some clinicians rely on intelligence systems for diagnostic results. The overall medical field is noted as having numerous serious difficulties. The need to comprehend decisions and establish trust in some AI systems makes the domains of defence, finance, and law important. Argumentation is a further word worth mentioning. XAI–Ensemble ML asserts that ML and argument theory systems improve in readability and predictability [65]. They even recommend applying it to robotics, the semantic web, security, and medical informatics. This study also mentions several general-purpose systems including process optimisation, investment advice, and business decisions. Therefore, we have listed below the methods that are most frequently employed in this industry [18,27,29,32,36,52,54,58,66]. Tables 3 and 4 shows different application domains and their desired XAI properties.
  - Decision support system: CDSS, MDSS

- Machine learning: MLP, MLA
- Neural network: DNN, RNN, CNN
- Medical Imaging
- Graph Mining
- Image classification
- Data fusion

Table 3. Different application domains and their desired XAI properties

Year	References	Domain	Properties
2018	[67]	Healthcare, Radiology,	Interpretability, Transparency
2019	[46]	Healthcare, Finance, Robotics, Education	Interpretability, Transparency, Fairness
	[17]	Healthcare, Finance, Robotics, Education	Interpretability, Transparency, Trustworthiness
2020	[51]	Healthcare, Finance, Robotics, Education	Transparency, Interpretability, Fairness
	[20]	Healthcare, Finance, Robotics, Education	Interpretability, Transparency, Fairness
	[36]	Healthcare	Interpretability, Transparency, Trustworthiness
	[18]	Healthcare, Finance, Robotics	Interpretability, Trustworthiness, Transparency
	[47]	Healthcare, Finance, Robotics, Education	Transparency, Fairness Interpretability
	[66]	Healthcare, Robotics, Education	Interpretability, Transparency, Trustworthiness
2021	[62]	Healthcare, Finance, Robotics, Education	Interpretability, Transparency, Fairness
	[61]	Healthcare	Interpretability, Transparency,
	[30]	Healthcare, Finance, Robotics, Education	Interpretability, Transparency, Trustworthiness
	[63]	Healthcare, Education, Finance, Robotics	Interpretability, Transparency, Fairness
2022	[29]	Healthcare	Interpretability, Transparency, Fairness
	[21]	Geoscience	Interpretability, Transparency, Fairness
	[38]	Healthcare, Finance, Robotics, Education	Interpretability, Transparency, Privacy
	[40]	Healthcare, Robotics, Finance, Education	Transparency, Interpretability, Fairness
	[54]	Healthcare, Education, Finance, Robotics	Interpretability, Transparency, Fairness

Furthermore, healthcare uses XAI approaches extensively to enhance choices and boost trust with automated options. Research on medical application developments, especially those studied by Wani et al. [42] and Jung et al. [56], reveals various possibilities and other issues, with a focus on enhancing transparency and results for patients. The application of GANs in clinical imaging to enhance diagnostic precision is covered by Islam et al. [43]. Moreover, Berloco et al. [68] apply XAI to the analysis of survival, prioritizing the use of models that are vital to healthcare in situations involving surviving patients.

- **Industrial:** Researchers from the industrial sector intended to demonstrate how CNN and CAM might be applied to real vibration films to characterise the status of the machine by illustrating healthy or harmful circumstances. They deal with a convolutional NN with automated diagnostic methods based on class activation maps for machine monitoring [67] because it either captures or makes it easier to distinguish the area of our image that is flawed and find errors. They also demonstrated the validation of their suggested model using datasets from a water pump and a base-energised cantilever beam. Even in the financial industry, XAI has been applied to risk management, fraud detection, and stock price forecasting. The authors put forth an XAI-based methodology for forecasting daily NFT and DeFi pricing dynamics in the publication XAI–ensemble ML [65]. The approach uses XAI and ensemble machine learning to provide justifications for the model's predictions. Similar to this, the authors of the publication XAI–quantified explainability suggested a paradigm for fraud detection that makes use of quantified explainability. The framework uses XAI and machine learning approaches to explain the outcomes of fraud detection [41]. Additionally, XAI is employed in a wide range of well-known industries for better user experiences, as we list below:
  - Marketing;
  - Insurance;
  - Financial Services.
- **Geoscience:** This is yet another field where XAI has demonstrated significant promise. The authors put forth an XAI-based framework for earthquake prediction in the publication XAI–earthquakes [23]. The structure of the model uses XAI and machine learning techniques to give justifications for the model's predictions. Similar to this, the authors of the study XAI–geoscience presented an XAI-based framework for classifying remote sensing images. The system makes use of XAI and convolutional neural networks (CNNs) to explain the model's choices, which could make it easier to comprehend the categorization outcomes [21].
- **IoT systems:** These are more areas where XAI has demonstrated considerable promise. The authors of the paper XAI-IoT suggested an XAI-based architecture for creating agents for IoT systems. The framework makes use of XAI approaches to offer justifications for the choices made by the agents, which may help us comprehend how the system behaves. Similar to this, the authors of the study XAI–transparency suggested a framework for improving the transparency of IoT systems using XAI approaches [18]. The framework offers justifications for the choices the system makes, which could help us comprehend the system's behaviour and enhance its efficiency [18,28].
- **Soil testing and earthquakes:** A different research investigation on soil testing and earthquakes suggested using SHAP tools to decipher a multi-layer perceptron's (MLP's) output and assess the effects of individual features from the paper XAI–earthquakes. Even using a post-event satellite image from the 2018 Palu earthquake, they could distinguish buildings that had collapsed and those that had not with an overall accuracy of 84% [23]. Furthermore, they provide a transferable, explicable model for increased clarity and precision.
- **Social Science:** The relevance of end-user expectations in developing explanations of AI systems was noted in a social science study [26]. Even outputs from a text classifier system that provided a factual or counterfactual explanation were validated and displayed. However, they discovered systemic mistrust and abuse. They saw the necessity for more open communication with end consumers as a result of that. They advised concentrating on the model's transparency and interoperability as a result.

Table 4. Different application domains and their desired XAI properties

Year	References	Domain	Properties
2022	[57]	Healthcare	Interpretability, Transparency
2023	[28]	IoT	Interpretability, Transparency
	[50]	Autonomous Driving	Transparency, Safety
	[69]	Image and Video Processing	Interpretability, Transparency
	[64]	Anomaly Detection	Interpretability, Transparency
	[65]	Machine Learning	Interpretability, Transparency
	[60]	Predictive Maintenance	Interpretability, Transparency
	[56]	Healthcare	Explainability, Effectiveness, Trust
	[49]	AI Trustworthiness	Interpretability, Trust, Fairness
2024	[43]	Medical Imaging (GANs)	Interpretability, Transparency, Realism
	[42]	Healthcare, IoMT	Trust, Transparency, Scalability
	[68]	Healthcare	Interpretability, Accuracy
	[70]	Cybersecurity	Interpretability, Transparency, Forensic Analysis
	[48]	Machine Learning (CLT)	Explainability, Interpretability, Transparency

XAI's varied applications throughout industries are demonstrated by strategies such as the usage of GANs for medical applications [43], which vary with more conventional XAI methods like neural networks for attention applied to motion prediction [50]. Moreover, several emphasize that easier-to-understand models can promote confidence in AI ([49,70]), supporting the claim made by [48] that explanations should be less cognitively taxing.

Furthermore, there are some notable industries—namely, the military, defence, law, the legal system, transportation, social media, and trading—where XAI products are adopted quickly and have a lot of potential. Various neural network frameworks, deep learning or machine learning classifiers, algorithms, and models are used to improve things. These can gain from explainable AI and its applications even in the future.

5.2. *What Are the Most Essential Properties for Each Domain in the Case of Explainable AI?*

A thorough analysis of the chosen papers revealed that diverse fields use explainable artificial intelligence (XAI) in a variety of ways. The paper that follows focuses on the crucial characteristics of XAI in many domains:

1. XAI in machine learning (ML): Interpretability, transparency, and fairness are three characteristics that make up the core of XAI in ML. Transparency is the capacity to gain access to and examine the data used to train the model, whereas interpretability is the capacity to comprehend how a model makes its conclusions. Fairness relates to making sure the model is not favouring any one group or attribute over another [17,46].
2. XAI in healthcare: Explainability, interpretability, and trustworthiness are three of XAI's key characteristics in the healthcare sector. Explainability is the capacity to comprehend the rationale behind a diagnosis or course of treatment. Understanding

how a model concluded is referred to as interpretability. The reliability and accuracy of the model's predictions are referred to as trustworthiness [29,36].

3. XAI in Natural Language Processing (NLP): Transparency, interpretability, and explainability are three key characteristics of XAI in NLP. Understanding the data and algorithms the model utilized is what is meant by transparency. Understanding the model's internal representations and decision-making process is referred to as interpretability. Explainability is the capacity to comprehend the rationales underlying the model's forecasts [30,63].
4. XAI in robotics and reinforcement learning (RL): Interpretability, transparency, and accountability are three crucial characteristics of XAI in robotics and real-world applications. The ability to comprehend a robot's or agent's decision-making process is referred to as interpretability. Access to and analysis of the data used to train the model are two terms used to describe transparency. Accountability is the capacity to comprehend and justify the actions of the robot or agent [12,33].
5. XAI in anomaly detection: Features that are relevant and interpretable are crucial for XAI in anomaly detection. Understanding which traits or attributes are most crucial for spotting abnormalities is known as feature relevance. Understanding how the model makes predictions is referred to as interpretability [64].
6. XAI in predictive maintenance: The interpretability, explainability, and accuracy of XAI in predictive maintenance are its key characteristics. Understanding how the model generates its predictions is referred to as interpretability. Understanding the rationale behind the predictions made by the model is referred to as explainability. The capacity of the model to precisely forecast equipment failures or maintenance requirements is referred to as accuracy [60].
7. XAI in the Internet of Things (IoT): The interpretability, explainability, and flexibility of XAI in IoT are its key characteristics. Understanding how the model generates its predictions is referred to as interpretability. Understanding the rationale behind the predictions made by the model is referred to as explainability. The term "adaptability" describes a model's capacity to adjust to changing inputs or external factors [28]. Moreover, Limeros et al. [50] investigate the application of explainable artificial intelligence (XAI) in automated driving, with a particular emphasis towards behaviour predictions through graphical models. Their findings demonstrate the growing significance of XAI in situations where safety is paramount.
8. XAI in Financial Data Analysis: In the study of financial data, interpretability, accuracy, and speed are crucial components of XAI. Understanding how the model generates its predictions is referred to as interpretability. The capacity of the model to effectively forecast financial trends or anomalies is referred to as accuracy. Speed describes how rapidly a model can process a lot of financial data [65].
9. Others: Loh et al. [57] along with Ali et al. [49] address more general topics of transparency and confidence, which are important concerns among a variety of industries, and highlight the importance of XAI in fostering confidence and acceptability in artificial intelligence (AI) technologies.

5.3. Mentioning the Major Contribution and Limits We Found in the Above Study

We recognized many important advances and limitations in the area of XAI based on our assessment of the chosen publications.

5.3.1. Major Contributions

1. XAI techniques: Developing and enhancing XAI methodologies for better machine learning model interpretability and transparency was the topic of several articles, such as XAI–concepts and XAI–transparency [18,51].
2. XAI in healthcare: The potential of XAI in enhancing medical diagnosis and treatment choices was investigated by XAI–health and XAI–medical. These studies emphasized the value of transparency and interpretability in the medical field [29,36].

3. XAI in anomaly detection: An essential application in cybersecurity and fraud detection, XAI–anomaly detection examined the methods and difficulties of employing XAI algorithms in anomaly detection [64].
4. XAI in IoT: To increase the interpretability and transparency of decision-making processes, XAI-IoT investigated the application of XAI-based agents in IoT systems. The importance of strong and trustworthy XAI approaches in IoT applications was highlighted in this research [28].
5. XAI in ensemble machine learning: The use of XAI–ensemble ML and ensemble machine learning techniques in forecasting and analysing daily prices of NFT and DeFi was studied. This study demonstrated the potential for improving interpretability and accuracy by merging several machine learning models with XAI approaches [65].
6. XAI in predictive maintenance: XAI–predictive maintenance examined how XAI and machine learning can be used for multi-component systems' predictive maintenance. This study showed how XAI might be used to discover system component relationships and boost the accuracy and effectiveness of preventative maintenance [60].

5.3.2. Major Limitations

1. Lack of standardization: A shortage of consistency in XAI procedures and evaluation measures, which restricts the generalizability and comparability of XAI research, was highlighted by XAI–survey and XAI–interpretability [30,46].
2. Limited scalability: The restricted scalability of XAI approaches, particularly in large-scale and complicated machine learning models, was highlighted by Peeking-BB and XAI–confidence. This restriction limits the use of XAI techniques in real-world situations [52,71].
3. Limited adoption: The insufficient adoption of XAI approaches in real-world applications, notably among end users and decision makers, was examined in XAI–end-users and XAI–argumentation. These publications highlighted the requirement for more user-friendly and intuitive XAI strategies that can successfully explain machine learning models' reasoning and judgment processes to non-technical stakeholders [26,63].
4. Ethical and legal concerns: XAI techniques like privacy-preserving XAI and institutional XAI raise ethical and legal issues, particularly when applied to delicate industries like finance and healthcare. These publications highlighted the requirement for transparent, accountable XAI methods that uphold individual privacy and human rights [38,40].

5.4. What Are the Available Algorithms and Frameworks for Explainable AI? Are They Self-Explanatory or Not?

The Association for Computing Machinery (ACM) issued a declaration on algorithmic responsibility and transparency in January 2017. The ACM emphasizes in this statement that prejudice can be damaging when algorithms are used for automated decision making. After that, DARPA started a program for XAI in May 2017. Their main goal is to present an understandable and incredibly accurate model, mentioned in XAI–supervised ML [62]. Our research shows that XAI systems were widely adopted because they were more respectable, orderly, reliable, and transparent. Additionally, AI models and algorithms are producing data more intelligently. Recent systems used as various XAI approaches in the study include sensing, mapping, imaging, reasoning, vision, CNN, visualization, data fusion, and many more.

There are numerous methods and frameworks available for XAI, according to our thorough analysis of the literature on the topic. Many of these frameworks and algorithms are designed to make AI models more transparent and interpretable so that people can understand how decisions are made.

The XAI–transparency study from 2020 reviews numerous XAI frameworks like LACE, DARTS, and CLEVER, and emphasizes the significance of transparency in XAI [18]. A methodology called LACE (Local Accountability via Conditional Expectation) offers regional justifications for specific predictions. A technology called DARTS (Data Analytics and Reasoning Transparency Service) makes machine learning procedures transparent. Using a paradigm called CLEVER (Cross-Level Explanation via Variable Importance Evaluation and Assessment), explanations are given by assessing the significance of input features. Although these frameworks seek to make AI models transparent and understandable, they may not always be self-explanatory and may require more work to decipher. Figure 16 shows the different tools used in the XAI domain.

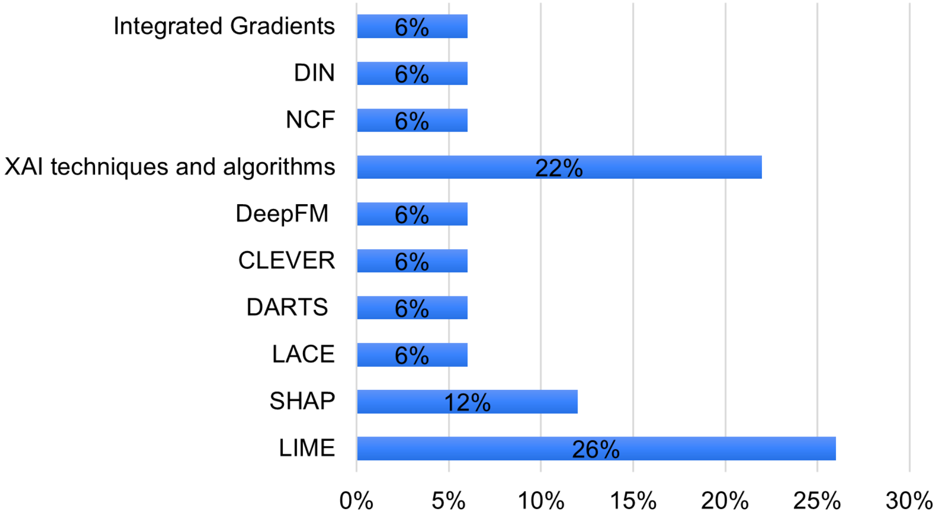

Figure 16. Most mentioned XAI tools.

The XAI-Rec paper examines some XAI algorithms, including DeepFM, NCF, and DIN, and focuses on explainable recommendation systems. DeepFM (Deep Factorization Machine) is a hybrid model that combines factorization machines and deep neural networks to produce a suggestion [20]. The neural network-based approach for the recommendation is called NCF (Neural Collaborative Filtering). A model called DIN (Deep Interest Network) learns user preferences and item characteristics to offer tailored recommendations. These algorithms are good at making recommendations that are easy to understand, but may not always be self-explanatory and may require subject expertise to interpret. A detailed analysis of several XAI algorithms, including tree-based approaches, rule-based methods, and deep learning-based methods, is provided in the work titled XAI–algorithms [47]. XAI methods for multimodal data, including audio, image, and text, are covered in the XAI—-multimodal work [54]. The XAI–predictive maintenance paper (2023) addresses machine learning-based XAI strategies for predictive maintenance, while the XAI–ensemble ML paper (2023) explores XAI techniques for ensemble machine learning [60,65].

5.5. What Are the Domain-Specific Hindrances for Explainable AI?

We can offer a broad response to the question of what the domain-specific barriers to explainable AI are after studying the list of studies. It is significant to note that, depending on the particular domain or application of explainable AI, obstacles may change. According to the chosen articles from 2018 to 2023, explainable AI (XAI) faces several domain-specific challenges as it develops and finds widespread use. These obstacles include social, ethical, and technical issues that limit how transparent, understandable, and reliable XAI systems can be.

The requirement for defined and proven evaluation techniques and criteria for judging the interpretability and transparency of XAI models is one of the technical challenges facing the technology. There is a need for a systematic and thorough evaluation methodology that can objectively measure the effectiveness of XAI techniques in many application domains, as detailed in the 2019 XAI survey paper [46]. Furthermore, particularly in multimodal and multitask environments, the complexity and heterogeneity of XAI models might make it difficult to describe how they function internally and how they make decisions [54].

Making a trade-off between model complexity and interpretability is another technological problem for XAI. Complex models, like deep neural networks, can achieve high accuracy but are frequently challenging to understand because of their black-box nature, as explained in XAI concepts from 2020 [51]. However, while simpler models like decision trees may be easier to understand, accuracy may suffer. As a result, hybrid XAI models that can balance accuracy and interpretability based on the application domain XAI interpretability are required [30].

The implementation of XAI systems can be hampered by social and ethical issues in addition to technical ones. According to XAI–health from 2020, there are worries regarding the potential bias and prejudice of XAI models in healthcare, especially in impoverished and marginalized communities [36]. Furthermore, the adoption and usage of XAI systems in clinical practice may be constrained by the end user's desire for greater trust in and comprehension of XAI models, such as doctors and patients [26].

Data protection, security, and accountability concerns are some additional ethical considerations for XAI. As mentioned in XAI–privacy preserving from 2022, XAI models frequently call for substantial volumes of private data, including financial and medical records, which raises worries about data breaches and unauthorized access [38]. Additionally, the XAI models' lack of accountability and transparency may have unforeseen consequences including worsening socioeconomic inequities and continuing preexisting biases [55].

In the geosciences field, [21] highlighted the difficulties in deciphering complex geographic data as well as the demand for understandable models that may help in decision making around natural catastrophes and climate change. The difficulties of deciphering the output of machine learning models and the significance of outlining the variables influencing the projections of financial assets like NFTs and DeFi were examined in the finance domain by [65].

The necessity for XAI systems to account for interdependencies between components in multi-component systems and to provide understandable explanations for predictions linked to predictive maintenance was highlighted in the industrial realm by the paper [60]. XAI–anomaly detection explored the difficulties in locating pertinent features and making sure that the explanations given by XAI systems are pertinent and meaningful to end users in the area of anomaly detection [64].

6. Discussion

To identify and assess the current status of research in the field of explainable artificial intelligence (XAI), the study conducted a comprehensive review of the literature. The research examined 53 publications that were published in numerous respected journals and conference proceedings over six years, from 2018 to 2024. The analysis focused on finding similar themes, research topics, methodologies, difficulties, and possibilities related to XAI. The papers were chosen based on their relevance to XAI.

According to the review, XAI is a new and quickly expanding discipline that aims to provide strategies and tactics for making AI systems more understandable, comprehensible, and transparent to humans. XAI seeks to solve issues such as lack of accountability, transparency, and trust, as well as the possibility of biased or unjust decision making, which are brought on by the growing use of AI systems [72]. Model-agnostic explanations, interpretability metrics, transparency and trust, and post hoc explanations are just a few of the research areas of XAI that were highlighted by the study [73].

Several methodologies and techniques used in XAI were also recognized in the review, including rule-based explanations, feature relevance methods, visualization techniques, and natural language explanations. According to the report, XAI offers a number of potentials, including increasing public confidence in and understanding of AI systems, identifying and reducing biases, strengthening decision making, and increasing the standard of human–AI interaction. Each source research paper was thoroughly examined and pertinent qualitative and quantitative information was taken.

The study also noted many obstacles that XAI must overcome, including the need for interdisciplinary cooperation, the requirement of a trade-off between interpretability and performance, a lack of standards, and the complexity of AI systems. The evaluation concluded that XAI is still in its infancy and that additional study is required to address these difficulties and fully achieve XAI's promise. As a result, the report offers a comprehensive and insightful picture of the current research landscape of XAI. The paper outlines numerous potential problems in the field and emphasizes the importance of XAI in enhancing the trust, transparency, and interpretability of AI systems. The report also underlines the requirement for standardization and interdisciplinary cooperation to improve the XAI area. The results of this study can be used by practitioners and researchers in the field of AI to direct future work on XAI research and development.

6.1. Limitations

Here, several notable limitations of our paper are listed in the part below:

1. The selection of papers is limited to those published between 2018 and early 2024, and it is possible that important papers on the topic were published after this time frame.
2. The search for papers may have been incomplete, as it was limited to specific databases and keywords.
3. The evaluation of papers may have seen some subjectivity in the analysis and evaluation process.
4. The analysis only considers papers published in English, potentially missing out on important contributions published in other languages.
5. The study has not covered all forms of explainable AI but rather the selected papers with the terms "interpretability, transparency, and explainability".

6.2. Recommendations

Based on the analysis of the 44 papers on explainable artificial intelligence, we propose the following potential future research directions that could benefit the IT industry:

- **Impact on User Trust and Acceptance:** Investigate how XAI influences user trust and acceptance of AI systems [35,69].
- **Standardized Metrics and Evaluation:** Develop standardized metrics and evaluation methods for XAI to enhance comparability and effectiveness [26].
- **Ethical Implications and Responsible AI:** Explore the ethical implications of XAI and create frameworks for responsible AI development [23,33,51].
- **Enhancing Cybersecurity and Privacy:** Study how XAI can be used to improve the cybersecurity and privacy of AI systems [38,40].
- **Real-World Applications:** Develop XAI solutions tailored to practical problems in industries such as healthcare, finance, and IoT, including detailed evaluations and predictions in these domains [28,41,63].

6.3. Future Direction

Based on our findings and analysis, we identify the following key directions for future research:

- **Comparative Studies of XAI Techniques:** Conduct comparative studies to identify the most effective and efficient XAI techniques and methods.
- **Ethical and Social Impact:** Investigate the ethical implications of XAI, including its impact on data privacy, algorithmic bias, and human agency.
- **Development of New Techniques:** Develop innovative XAI techniques and tools to improve the interpretability, transparency, and trustworthiness of AI systems.
- **User Experience:** Study user experiences with XAI, focusing on how users perceive, interpret, and utilize the explanations provided by AI systems.

- **Decision-Making Enhancement:** Explore how XAI can enhance decision-making processes in various domains, including criminal justice, climate change, and social services.
- **Robustness and Security:** Investigate how XAI can contribute to the robustness and security of AI systems, including protection against adversarial attacks and cybersecurity threats.
- **Adoption and Acceptance:** Study the impact of XAI on the adoption and acceptance of AI systems, identifying factors that influence user trust and confidence.

7. Conclusions

Understanding explainable AI (XAI) is essential in today's AI-driven landscape. XAI enhances transparency by elucidating AI decision-making processes, which is crucial in high-stakes domains like healthcare and finance. This transparency fosters user trust and facilitates the detection and mitigation of biases, ensuring fairness and ethical AI use. Despite these benefits, current XAI techniques face significant challenges, including the trade-off between accuracy and interpretability, scalability issues, high computing costs, and domain-specific limitations [74]

Our review highlights that while XAI can improve model performance and compliance with regulatory standards, it also reveals areas needing further exploration. The trade-off between model accuracy and interpretability remains a critical challenge, as more interpretable models often sacrifice some degree of accuracy. Additionally, the scalability of XAI techniques and their applicability to diverse domains require further research.

Future research should focus on developing scalable XAI solutions that balance accuracy with interpretability, reducing computing costs, and addressing domain-specific issues. Exploring these areas will enhance XAI's effectiveness, support its integration into various applications, and advance the development of responsible AI practices. By addressing these challenges, XAI can better serve its role in risk assessment, debugging, and educational initiatives, ultimately contributing to the advancement of fair and ethical AI technologies.

Author Contributions: S.U.H.: conceptualization of the review topic, literature review, synthesis of research findings, writing—original draft, visualization, and references. M.J.M.C.: analysis of primary research papers, writing, review, editing, and supervision. N.R.C.: critical review and interpretation of findings, review, editing, and supervision. S.K.S.:compilation of references and formatting of the review article. K.B.: compilation of references, and formatting of the review article. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: We gratefully acknowledge all our colleagues who generously provided their valuable feedback and insights on our paper. Their constructive comments and suggestions have significantly contributed to improving the quality and clarity of our work.

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

References

1. Knight, W. *AI's Language Problem*; MIT Technology Review: Cambridge, MA, USA, 2020. Available online: https://www.technologyreview.com/2016/08/09/158125/ais-language-problem/ (accessed on 15 July 2023).
2. Innovation, Science and Economic Development Canada. Government of Canada Announces Next Steps in Safeguarding Research. *Canada.ca*. March 2021. Available online: https://www.canada.ca/en/innovation-science-economic-development/news/2021/03/government-of-canada-announces-next-steps-in-safeguarding-research.html (accessed on 22 March 2021).
3. IBM. Explainable AI. Ibm.com. 2021. Available online: https://www.ibm.com/watson/explainable-ai (accessed on 6 September 2021).
4. Wiener, N. Cybernetics: Or the Control and Communication in the Animal and the Machine. In *Analog VLSI: Signal and Information Processing*; MIT Press: Cambridge, CA, USA, 1961. Available online: https://direct.mit.edu/books/oa-monograph/4581/Cybernetics-or-Control-and-Communication-in-the (accessed on 15 November 2021).

5. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Res. Net* **2015**, *521*, 436–444. [CrossRef] [PubMed]
6. Patil, R. Industry Trends in AI—Topic 3—Explainable AI. Linkedin. Available online: https://www.linkedin.com/pulse/industry-trends-ai-topic-3-explainable-rajesh-patil (accessed on 5 May 2022).
7. Loyola-González, O. Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View. *IEEE Access* **2019**, *7*, 154096–154113. [CrossRef]
8. Transparency and Responsibility in Artificial Intelligence. Available online: https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/innovatie/deloitte-nl-innovation-bringing-transparency-and-ethics-into-ai.pdf (accessed on 13 September 2021).
9. Haasdijk, E. A Call for Transparency and Responsibility in Artificial Intelligence. Deloitte Netherlands. Available online: https://www2.deloitte.com/nl/nl/pages/innovatie/artikelen/a-call-for-transparency-and-responsibility-in-artificial-intelligence.html (accessed on 13 September 2021).
10. Model Transparency and Explainability. *Ople.Ai.* 24 March 2020. Available online: https://ople.ai/ai-blog/model-transparency-and-explainability/ (accessed on 13 September 2021).
11. Choudhury, A. Explainability vs. Interpretability In Artificial Intelligence and Machine Learning. Analyticsindiamag.com. 14 January 2019. Available online: https://analyticsindiamag.com/explainability-vs-interpretability-in-artificial-intelligence-and-machine-learning/ (accessed on 16 September 2021).
12. Sequeira, P.; Gervasio, M. Interestingness Elements for Explainable Reinforcement Learning: Understanding Agents' Capabilities and Limitations. *Artif. Intell.* **2020**, *288*, 103367. [CrossRef]
13. Model Interpretability. Datarobot.com. Available online: https://www.datarobot.com/wiki/interpretability/ (accessed on 16 September 2021).
14. Shen, O. Interpretability in Machine Learning: An Overview. The Gradient. 21 November 2020. Available online: https://thegradient.pub/interpretability-in-ml-a-broad-overview/ (accessed on 13 September 2021).
15. Model Interpretability. Darpa Mil. Available online: https://www.darpa.mil/attachments/XAIProgramUpdate.pdf (accessed on 13 September 2021).
16. Systematic Reviews—Subject and Research Guides at Macquarie University. Macquarie University. 2024. Available online: https://libguides.mq.edu.au/systematic_reviews/prisma_screen#:~:text=The%20PRISMA%20flow%20diagram%20visually,recorded%20at%20the%20different%20stages (accessed on 26 September 2024).
17. Carvalho, D.V.; Pereira, E.M.; Cardoso, J.S. Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics* **2019**, *8*, 832. [CrossRef]
18. Meske, C.; Bunde, E. Transparency and Trust in Human-AI-Interaction: The Role of Model-Agnostic Explanations in Computer Vision-Based Decision Support. In *Artificial Intelligence in HCI*; Springer: Cham, Switzerland, 2020; pp. 54–69. [CrossRef]
19. Moradi, M.; Samwald, M. Post-Hoc Explanation of Black-Box Classifiers Using Confident Itemsets. *Expert Syst. Appl.* **2021**, *165*, 113941. [CrossRef]
20. Zhang, Y.; Chen, X. Explainable Recommendation: A Survey and New Perspectives. *Found. Trends® Inf. Retr.* **2020**, *14*, 1–101; ISBN 978-1-68083-658-5. [CrossRef]
21. Mamalakis, A.; Ebert-Uphoff, I.; Barnes, E. Neural Network Attribution Methods for Problems in Geoscience: A Novel Synthetic Benchmark Dataset. *Environ. Data Sci.* **2022**, *1*, 7. [CrossRef]
22. Evans, T.; Retzlaff, C.; Geißler, C.; Kargl, M.; Plass, M.; Müller, H.; Kiehl, T.-R.; Zerbe, N.; Holzinger, A. The Explainability Paradox: Challenges for XAI in Digital Pathology. *Future Gener. Comput. Syst.* **2022**, *133*, 281–296. [CrossRef]
23. Matin, S.S.; Pradhan, B. Earthquake-Induced Building-Damage Mapping Using Explainable AI (XAI). *Sensors* **2021**, *21*, 4489. [CrossRef]
24. Femi, P.S.; Ashwini, K.; Kala, A.; Rajalakshmi, V. Explainable Artificial Intelligence for Cybersecurity. *Wirel. Commun. Cybersecur.* **2023**, *103*, 149–174. [CrossRef]
25. Confalonieri, R.; Weyde, T.; Besold, T.R.; Moscoso del Prado Martín, F. Using Ontologies to Enhance Human Understandability of Global Post-Hoc Explanations of Black-Box Models. *Artif. Intell.* **2021**, *296*, 103471. [CrossRef]
26. Riveiro, M.; Thill, S. "That's (Not) the Output I Expected!" On the Role of End User Expectations in Creating Explanations of AI Systems. *Artif. Intell.* **2021**, *298*, 103507. [CrossRef]
27. van der Waa, J.; Nieuwburg, E.; Cremers, A.; Neerincx, M. Evaluating XAI: A Comparison of Rule-Based and Example-Based Explanations. *Artif. Intell.* **2021**, *291*, 103404. [CrossRef]
28. Dobrovolskis, A.; Kazanavičius, E.; Kižauskienė, L. Building XAI-Based Agents for IoT Systems. *Appl. Sci.* **2023**, *13*, 4040. [CrossRef]
29. Singh, A.; Sengupta, S.; Lakshminarayanan, V. Explainable Deep Learning Models in Medical Image Analysis. *J. Imaging* **2020**, *6*, 52. [CrossRef]
30. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **2021**, *23*, 18. [CrossRef]
31. Sahoh, B.; Choksuriwong, A. Beyond Deep Event Prediction: Deep Event Understanding Based on Explainable Artificial Intelligence. In *Explainable AI and Other Applications of Intelligent Computing*; Springer: Cham, Switzerland, 2021; pp. 91–117. [CrossRef]
32. Hudec, M.; Mináriková, E.; Mesiar, R.; Saranti, A.; Holzinger, A. Classification by Ordinal Sums of Conjunctive and Disjunctive Functions for Explainable AI and Interpretable Machine Learning Solutions. *Knowl.-Based Syst.* **2021**, *220*, 106916. [CrossRef]

33. Wells, L.; Bednarz, T. Explainable AI and Reinforcement Learning—A Systematic Review of Current Approaches and Trends. *Front. Artif. Intell.* **2021**, *4*, 550030. [CrossRef]
34. Zhou, J.; Gandomi, A.H.; Chen, F.; Holzinger, A. Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics. *Electronics* **2021**, *10*, 593. [CrossRef]
35. Setzu, M.; Guidotti, R.; Monreale, A.; Turini, F.; Pedreschi, D.; Giannotti, F. GLocalX—From Local to Global Explanations of Black Box AI Models. *Artif. Intell.* **2021**, *294*, 103457. [CrossRef]
36. Amann, J.; Blasimme, A.; Vayena, E.; Frey, D.; Madai, V. Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective. *BMC Med. Inf. Decis. Mak.* **2020**, *20*, 1–9. [CrossRef] [PubMed]
37. de Sousa, I.P.; Vellasco, M.M.B.R.; Costa da Silva, E. Explainable Artificial Intelligence for Bias Detection in COVID CT-Scan Classifiers. *Sensors* **2021**, *21*, 5657. [CrossRef] [PubMed]
38. Ogrezeanu, I.; Vizitu, A.; Ciușdel, C.; Puiu, A.; Coman, S.; Boldișor, C.; Itu, A.; Robert, D.; Moldoveanu, F.; Suciu, C.; et al. Privacy-Preserving and Explainable AI in Industrial Applications. *Appl. Sci.* **2022**, *12*, 6395. [CrossRef]
39. Merry, M.; Riddle, P.; Warren, J. A Mental Models Approach for Defining Explainable Artificial Intelligence. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 7. [CrossRef]
40. Theunissen, M.; Browning, J. Putting Explainable AI in Context: Institutional Explanations for Medical AI. *Ethics Inf. Technol.* **2022**, *24*. [CrossRef]
41. Varandas, R.; Gonçalves, B.; Gamboa, H.; Vieira, P. Quantified Explainability: Convolutional Neural Network Focus Assessment in Arrhythmia Detection. *BioMedInformatics* **2022**, *2*, 124–138. [CrossRef]
42. Wani, N.A.; Kumar, R.; Mamta; Bedi, J.; Rida, I. Explainable AI-Driven IoMT Fusion: Unravelling Techniques, Opportunities, and Challenges with Explainable AI in Healthcare. *Inf. Fusion* **2024**, *110*, 102472. [CrossRef]
43. Islam, S.; Aziz, M.T.; Nabil, H.R.; Jim, J.R.; Mridha, M.F.; Kabir, M.M.; Asai, N.; Shin, J. Generative Adversarial Networks (GANs) in Medical Imaging: Advancements, Applications, and Challenges. *IEEE Access* **2024**, *12*, 35728–35753. [CrossRef]
44. Manesh, A. Interpretable Machine Learning (IML) with XGBoost and Additive Tools. *Medium*, 2022. Available online: https://medium.com/@anicomanesh/interpretable-machine-learning-iml-with-xgboost-and-additive-tools-42258fb1f14 (accessed on 17 September 2024).
45. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. *arXiv* **2016**, *13*. [CrossRef]
46. Xu, F.; Uszkoreit, H.; Du, Y.; Fan, W.; Zhao, D.; Zhu, J. Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. In *Natural Language Processing and Chinese Computing*; Zhang, H., Ed.; Springer: Cham, Switzerland, 2019; pp. 563–574. [CrossRef]
47. Janssen, M.; Hartog, M.; Matheus, R.; Ding, A.; Kuk, G. Will Algorithms Blind People? The Effect of Explainable AI and Decision-Makers' Experience on AI-Supported Decision-Making in Government. *Soc. Sci. Comput. Rev.* **2020**, *40*, 089443932098011. [CrossRef]
48. Fox, S.; Rey, V.F. A Cognitive Load Theory (CLT) Analysis of Machine Learning Explainability, Transparency, Interpretability, and Shared Interpretability. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1494–1509. [CrossRef]
49. Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J.M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Díaz-Rodríguez, N.; Herrera, F. Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence. *Inf. Fusion* **2023**, *99*, 101805. [CrossRef]
50. Limeros, S.C.; Majchrowska, S.; Johnander, J.; Petersson, C.; Llorca, D.F. Towards Explainable Motion Prediction Using Heterogeneous Graph Representations. *Transp. Res. Part C Emerg. Technol.* **2023**, *157*, 104405. [CrossRef]
51. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, V.R.; et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
52. van der Waa, J.; Schoonderwoerd, T.; van Diggelen, J.; Neerincx, M. Interpretable Confidence Measures for Decision Support Systems. *Int. J. Hum.-Comput. Stud.* **2020**, *144*, 102493. [CrossRef]
53. Marton, S.; Lüdtke, S.; Bartelt, C. Explanations for Neural Networks by Neural Networks. *Appl. Sci.* **2022**, *12*, 30980. [CrossRef]
54. Yang, G.; Ye, Q.; Xia, J. Unbox the Black-Box for the Medical Explainable AI via Multi-Modal and Multi-Centre Data Fusion: A Mini-Review, Two Showcases and Beyond. *Inf. Fusion* **2022**, *77*, 29–52. [CrossRef]
55. Bruijn, H.; Warnier, M.; Janssen, M. The Perils and Pitfalls of Explainable AI: Strategies for Explaining Algorithmic Decision-Making. *Gov. Inf. Q.* **2021**, *39*, 101666. [CrossRef]
56. 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]
57. Loh, H.W.; Ooi, C.P.; Seoni, S.; Barua, P.D.; Molinari, F.; Acharya, U.R. Application of Explainable Artificial Intelligence for Healthcare: A Systematic Review of the Last Decade (2011–2022). *Comput. Methods Programs Biomed.* **2022**, *226*, 107161. [CrossRef]
58. Bresso, E.; Monnin, P.; Bousquet, C.; Calvier, F.E.; Ndiaye, N.C.; Petitpain, N.; Smaïl-Tabbone, M.; Coulet, A. Investigating ADR Mechanisms with Explainable AI: A Feasibility Study with Knowledge Graph Mining. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 18. [CrossRef] [PubMed]
59. Nyrup, R.; Robinson, D. Explanatory Pragmatism: A Context-Sensitive Framework for Explainable Medical AI. *Ethics Inf. Technol.* **2022**, *24*, 3. [CrossRef] [PubMed]

60. Gashi, M.; Mutlu, B.; Thalmann, S. Impact of Interdependencies: Multi-Component System Perspective Toward Predictive Maintenance Based on Machine Learning and XAI. *Appl. Sci.* **2023**, *13*, 53088. [CrossRef]
61. 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]
62. Burkart, N.; Huber, M.F. A Survey on the Explainability of Supervised Machine Learning. *J. Artif. Intell. Res.* **2021**, *70*, 245–317. [CrossRef]
63. Vassiliades, A.; Bassiliades, N.; Patkos, T. Argumentation and Explainable Artificial Intelligence: A Survey. *Knowl. Eng. Rev.* **2021**, *36*, e5. [CrossRef]
64. Tritscher, J.; Krause, A.; Hotho, A. Feature Relevance XAI in Anomaly Detection: Reviewing Approaches and Challenges. *Front. Artif. Intell.* **2023**, *6*. [CrossRef]
65. Ghosh, I.; Alfaro-Cortés, E.; Gámez, M.; García-Rubio, N. Prediction and Interpretation of Daily NFT and DeFi Prices Dynamics: Inspection Through Ensemble Machine Learning & XAI. *Int. Rev. Financ. Anal.* **2023**, *87*, 102558. [CrossRef]
66. Sun, K.H.; Huh, H.; Adhi Tama, B.; Lee, S.Y.; Jung, J.H.; Lee, S. Vision-Based Fault Diagnostics Using Explainable Deep Learning With Class Activation Maps. *IEEE Access* **2020**, *8*, 129169–129179. [CrossRef]
67. Yamashita, R.; Nishio, M.; Do, R.K.; Togashi, K. Convolutional Neural Networks: An Overview and Application in Radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef]
68. Berloco, F.; Marvulli, P.M.; Suglia, V.; Colucci, S.; Pagano, G.; Palazzo, L.; Aliani, M.; Castellana, G.; Guido, P.; D'Addio, G.; et al. Enhancing Survival Analysis Model Selection through XAI(t) in Healthcare. *Appl. Sci.* **2024**, *14*, 6084. [CrossRef]
69. Sheu, R.-K.; Pardeshi, M.S.; Pai, K.-C.; Chen, L.-C.; Wu, C.-L.; Chen, W.-C. Interpretable Classification of Pneumonia Infection Using Explainable AI (XAI-ICP). *IEEE Access* **2023**, *11*, 28896–28919. [CrossRef]
70. Jinad, R.; Islam, A.; Shashidhar, N. Interpretability and Transparency of Machine Learning in File Fragment Analysis with Explainable Artificial Intelligence. *Electronics* **2024**, *13*, 2438. [CrossRef]
71. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [CrossRef]
72. Chiu, T.K.F.; Xia, Q.; Zhou, X.; Chai, C.; Cheng, M. Systematic Literature Review on Opportunities, Challenges, and Future Research Recommendations of Artificial Intelligence in Education. *CAEAI* **2023**, *4*, 100118. [CrossRef]
73. Vilone, G.; Longo, L. Explainable Artificial Intelligence: A Systematic Review. *arXiv* **2020**, arXiv:2006.00093. [CrossRef]
74. AI Tools. Latest AI Tools and Technologies Every Mobile App Developer Should Know. *Appmysite* **2024**. Available online: https://www.appmysite.com/blog/top-ai-tools-for-mobile-app-developers/ (accessed on 17 July 2024).

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.