

# Literature Review of Explainable Tabular Data Analysis

Helen O'Brien Quinn <sup>1,\*</sup>, Mohamed Sedky <sup>1</sup>, Janet Francis <sup>1</sup>  and Michael Streeton <sup>2</sup>

<sup>1</sup> Department of Digital, Tech, Innovation & Business Staffordshire University, College Road, Stoke on Trent ST4 2DE, UK

<sup>2</sup> Connexica Ltd., Unit D, Dyson Court, Dyson Way, Stafford Technology Park, Stafford ST18 0LQ, UK

\* Correspondence: helen.o'brienquinn@research.staffs.ac.uk

**Abstract:** Explainable artificial intelligence (XAI) is crucial for enhancing transparency and trust in machine learning models, especially for tabular data used in finance, healthcare, and marketing. This paper surveys XAI techniques for tabular data, building on previous work done, specifically a survey of explainable artificial intelligence for tabular data, and analyzes recent advancements. It categorizes and describes XAI methods relevant to tabular data, identifies domain-specific challenges and gaps, and examines potential applications and trends. Future research directions emphasize clarifying terminology, ensuring data security, creating user-centered explanations, improving interaction, developing robust evaluation metrics, and advancing adversarial example analysis. This contribution aims to bolster effective, trustworthy, and transparent decision making in the field of XAI.

**Keywords:** explainable artificial intelligence (XAI); tabular data; interpretable; machine learning (ML)

## 1. Introduction

Artificial intelligence (AI) and machine learning (ML) algorithms are used to build models capable of achieving impressive performance with regards to prediction or classification accuracy in a wide range of domains. However, they typically have a complex, black-box structure that prevents users from gaining a better understanding of the data or the task. As a result, the field of explainable artificial intelligence (XAI) has developed and grown, generating a lot of research in recent years. It is a fast-moving research field, and one in which XAI tries to show the workings of black box algorithms in a more transparent and more easily understood way for users who have a variety of diverse levels of knowledge. This is especially important when the data and models are used to make safety-critical decisions, such as in medical diagnostics or in financial risk assessments, where incorrect decisions could be made due to bias and false correlations in the data and the model [1]. Another driver for the development of XAI is the right to an explanation for users, which is enshrined in law in the European Union General Data Protection Regulations (GDPR).

Understanding how and why ML models make the decisions they do allows for deeper trust to be fostered in AI and better ethical AI solutions because not only is there trust but also accountability [2]. Making explanations easier to understand can function as a driver for the adoption of further AI [3].

Tabular data are the main type of data that companies and businesses deal with, it can be structured or semi-structured, and much of it is in tables, spreadsheets, databases, and data warehouses. In the case of tabular data, XAI becomes particularly important because many industries and applications rely on this type of data for decision making, such as in finance, healthcare, marketing, and many other areas [4,5]. These are shown in Table 1, which gives the domains and the areas in which explainable tabular data are used. The areas of usage are discussed in more detail in Section 5. Figure 1 illustrates keywords and terms used in explainable tabular data and shows how they interrelate. XAI is the interface between ML models and human decision makers; it allows for ethical, effective,



**Citation:** O'Brien Quinn, H.; Sedky, M.; Francis, J.; Streeton, M. Literature Review of Explainable Tabular Data Analysis. *Electronics* **2024**, *13*, 3806. <https://doi.org/10.3390/electronics13193806>

Academic Editor: Ping-Feng Pai

Received: 29 July 2024

Revised: 11 September 2024

Accepted: 16 September 2024

Published: 26 September 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



**Table 1.** *Cont.*

Domains	Examples of Applications of Explainable Tabular Data
Retail Sector	Customer churn prediction
	Improve product suggestions to a customer
	Customer segmentation
Human resources	Employee churn prediction
	Evaluate employee performance
Manufacturing Sector	Logistics and supply chain management
	Order fulfilment
	Quality control
	Process control
	Planning and scheduling
	Predictive maintenance
Utility Sector	Smart grid load balancing
	Forecast energy consumption for customers
Education	Predict student adaptability
	Predict student exam grades
	Course recommendations

Explainable AI is an important field of research in which several surveys have been carried out, with some examining its specific role with regards to tabular data [5], a survey of explainable supervised machine learning [2], a survey on XAI and natural language explanations [7], revealing black box logic [8], a methods-centric overview with concrete examples [4], a systematic survey of surveys on methods and concepts [9], from approaches, limitations, and applications aspects [10]. Sahakyan et al., compiled a survey of XAI used with tabular data, claiming that, to their knowledge, it was the first survey of this type [5]. Tabular data are used in a myriad of different disciplines, which makes it surprising that there are not more specific XAI methods for it.

This research article seeks to build on the work done by Sahakyan et al., by providing an up-to-date survey of XAI techniques relevant to tabular data by thoroughly analyzing previous studies. The method used in this literature review is shown in Figure 2. Initially, the research topic was defined to understand the scope of the review. Key words were chosen, and several digital library databases, Google Scholar, IEEE Xplore, ACM Digital Library, PubMed, Scopus, and Science Direct, were searched using the key words ‘Explainable artificial intelligence’, ‘XAI’, ‘survey’, and ‘tabular data’, and limited to the years 2021 to 2024, these being the years following Sahakyan et al.’s survey on XAI for tabular data [5]. Table 2 details the reasons for choice of databases, and Figure 2 shows the way in which the articles were selected.

**Table 2.** Showing the databases chosen and the reasons for their choice.

Database	Reasons
Google Scholar	Comprehensive Coverage: Accesses a wide range of disciplines and sources, including articles, theses, books, and conference papers, providing a broad view of available literature. User-Friendly Interface: Easy to use, making it accessible Citation Tracking: Shows how often articles have been cited and helps to gauge their influence and relevance.

Table 2. Cont.

Database	Reasons
IEEE Xplore	Specialised Focus: On electrical engineering, computer science, and electronics. High-Quality Publications: Includes peer-reviewed journals and conference proceedings from reputable organizations. Cutting-Edge Research: Provides access to the latest research published in technology and engineering.
ACM Digital Library	Focus on Computing and Information Technology: Resources specifically related to computing, software engineering, and information systems. Peer-Reviewed Content: High academic quality through rigorous peer review. Conference Proceedings: Important conferences in computing, giving the latest research developments.
PubMed	Biomedical Focus: A vast collection of literature in medicine, life sciences, and health, often innovative computing solutions. Free Access: Many articles are available for free. High-Quality Research: Peer-reviewed journals and is a trusted source for medical and clinical research.
Scopus	Extensive Database: A wide range of disciplines Citation Analysis Tools: Provides metrics for authors and journals. Quality Control: Peer-reviewed literature, reliability of the sources.
ScienceDirect	Multidisciplinary Coverage: A vast collection of scientific and technical research. Quality Journals: High-impact journals. Full-Text Access: Access to a large number of full-text articles, facilitating in-depth research.

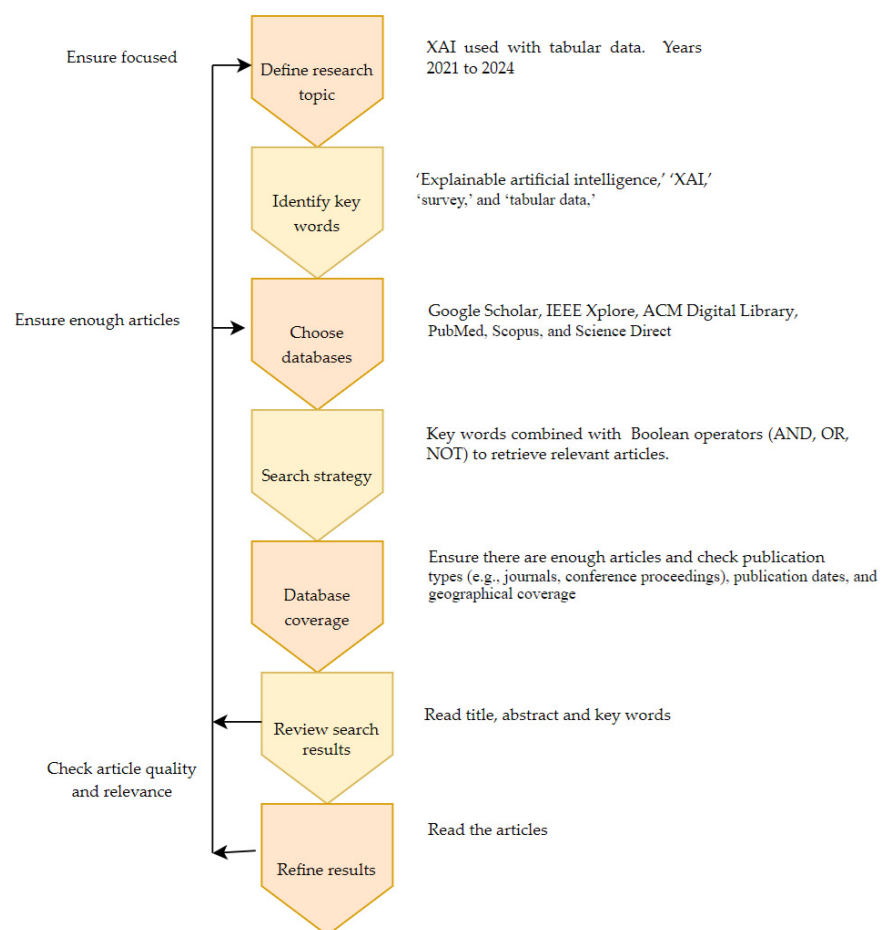


Figure 2. System used to select articles.

The coverage of the databases was evaluated, and the range of articles used is shown in Figure 3. To concentrate the focus of the review, the initial selection was made from the articles obtained from the search terms detailed in Table 3. The list of references of the selected studies was examined to include other papers that might not have been retrieved from the selected electronic databases. The one hundred and twenty-eight articles found were then whittled down using the search string ‘tabular’, giving a total of twelve articles. A further search of the twelve articles using the search string ‘Sahakyan’ left two articles. When the field was further narrowed to articles published in 2023, the total number of articles published was fifty-seven; five of those mention tabular data, and two of those articles mention Sahakyan et al., Table 4 shows the search results by year, and from this it can be seen that the interest in explainable tabular data is slowly growing. Tabular data are found in a range of different disciplines, and XAI research is taking place, in finance, health, management, statistics, environment, energy, law, engineering, and computing.

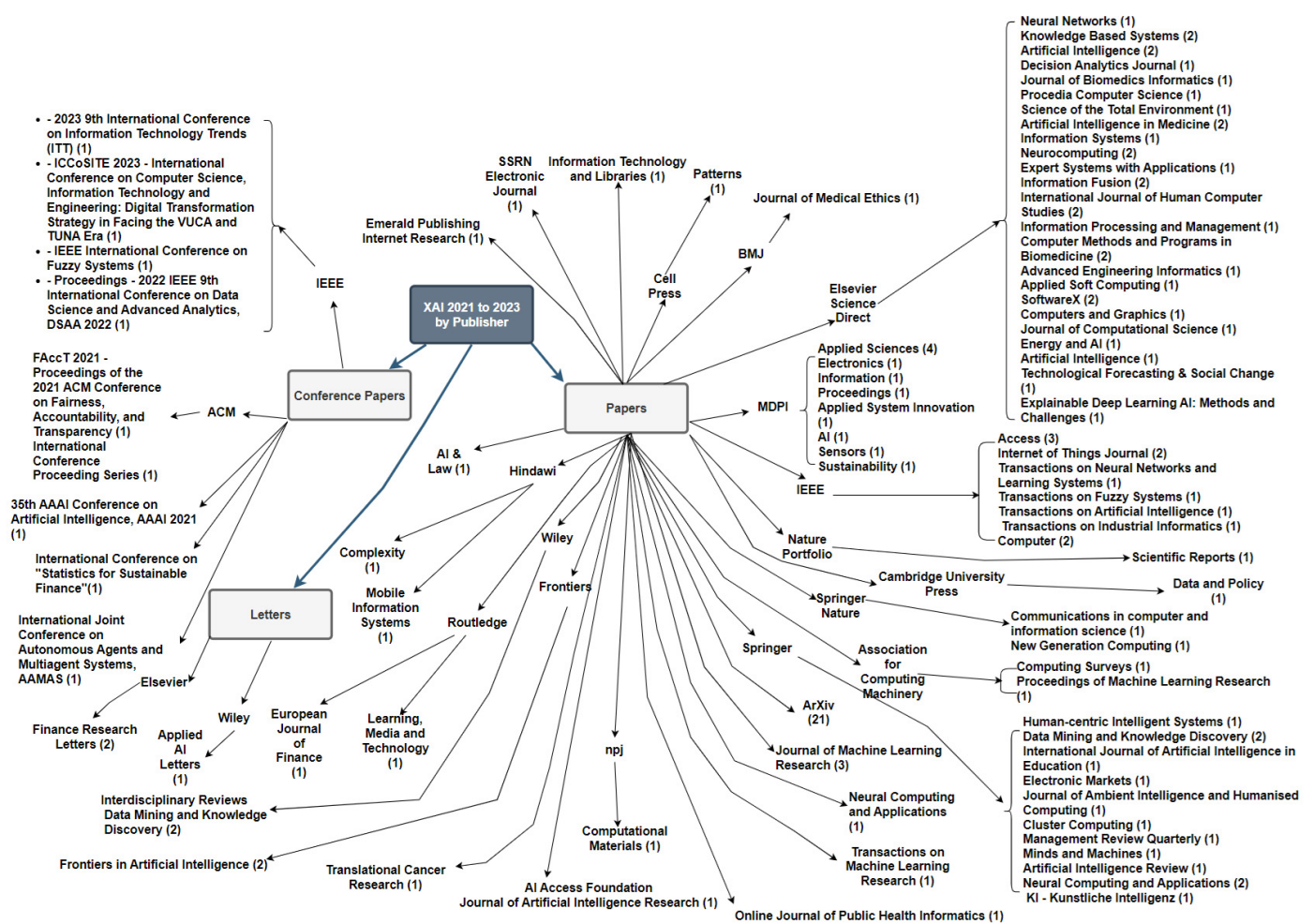


Figure 3. Chosen articles by Journal and Publisher.

Table 3. Shows the breakdown of the papers by the search terms used.

Search Terms	Number of Papers
XAI AND explainable artificial intelligence	128
XAI AND explainable artificial intelligence AND 2021	28
XAI AND explainable artificial intelligence AND 2022	43

Table 3. Cont.

Search Terms	Number of Papers
XAI AND explainable artificial intelligence AND 2023	57
2021 AND tabular	2
2022 AND tabular	5
2023 AND tabular	5
2021 AND survey (in title)	5
2022 AND survey (in title)	1
2023 AND survey (in title)	8
2021 AND survey AND tabular	1
2022 AND survey AND tabular	6
2023 AND survey AND tabular	11
2021 AND survey AND tabular AND Sahakyan (Sahakyan's article)	1
2022 AND survey AND tabular AND Sahakyan	0
2023 AND survey AND tabular AND Sahakyan	2

Figure 3 shows the breakdown of all the types of articles used in this literature review, divided by journal and publisher. This illustrates how widespread the use of tabular data has become.

#### *The Objectives of This Survey*

To analyze the various techniques, inputs, and methods used to build XAI models since 2021 to determine if any newly proposed model is better for tabular data than the already described models. This research article builds on the work done by Sahakyan et al. by providing an up-to-date survey of XAI techniques pertinent to tabular data and meticulously analyzing recent studies. An extensive literature review was conducted using digital library databases, focusing on the phrases 'Explainable artificial intelligence', 'XAI', 'survey', and 'tabular data' from 2021 to 2024. The objectives of this survey are threefold:

1. To analyze the various techniques, inputs, and methods used to build XAI models since 2021, aiming to identify any superior models for tabular data that have been created since Sahakyan et al.'s, paper.
2. To identify and expand upon Sahakyan et al.'s description of need, challenges, gaps, and opportunities in XAI for tabular data.
3. To explore evaluation methods and metrics used to assess the effectiveness of XAI models specifically concerning tabular data and to see if any new metrics have been developed.

The document is structured as follows: Section 2 introduces the theoretical foundations of explainability, providing an overview of fundamental concepts underpinning XAI. Section 3 delves into existing techniques for explainable tabular data analysis, categorizing and describing different XAI techniques suitable for tabular data analysis. Section 4 highlights challenges and gaps in explainable tabular data analysis. Section 5 explores applications of explainable tabular data analysis and its impact on decision making and user trust in various domains. Section 6 looks ahead to future directions and emerging trends, outlining potential research avenues in explainable tabular data analysis. Finally, Section 7 concludes by summarizing the key findings from the literature review.

## 2. Background

XAI is a way of making it easier for users to understand how models obtain their results. The main aim is to make the workings of a black box model transparent [11]. This has meant that XAI research has burgeoned in recent years, and with it many ways to

explain AI systems and their predictions. These methods, when integrated with an AI system, may, but not always, meet the legal demands of legislation and guidelines such as the GDPR's right-to-explanation [12] and DARPA's policies [13].

Conducting XAI research surveys is not a simple task because, not only are there different definitions for the same word, but the taxonomies used are also different, meaning that it is difficult to compare like with like [14–16]. Barbiero et al., have tried to remedy this by devising a theoretical basis for XAI taxonomies [17], and Ali et al., have proposed their own taxonomy, as have Graziani et al. [1,15]. The problem of multiple different taxonomies continues.

The meanings of the terminology around XAI vary, with many terms having no one agreed meaning [5,16]. Unfortunately, the various methods created have been used in isolation, with the result that all the methods have not been brought together to form one universally agreed way to explain all AI systems [17].

The X in XAI is referred to as 'explainable;' however the ML community often appears to use the words explainability and interpretability interchangeably to explain transparency and trust. These words, whilst not having agreed formal definitions, are slightly different. Vilone and Longo define interpretability as "The capacity to provide or bring out the meaning of an abstract concept" [18]. Rudin described interpretability as being a domain-specific idea and so maintained that it could not have an all-purpose definition [19]. Barredo Arrieta et al. define interpretability as being the ability to explain or to provide the meaning in a human understandable way [20]. Sahakyan et al. viewed explainability as a way of showing the features that have contributed to the result of a specific instance [5]. Barredo Arrieta et al. also agreed with this definition, describing it as any procedure or action conducted by the model to make clear its internal workings, meaning it provided a rationale for its decision [20]. They also viewed explainability as being an active process, while interpretability has been considered as being passive. Barbiero et al. argued that without agreed explainable AI definitions based in mathematics, there are likely to be a plethora of taxonomies containing unnecessary information, the wrong types of questions and a narrowing of the field for future research directions, [17].

Transparency is another term that has no one agreed definition. Often transparency is seen as an umbrella term frequently used with XAI and can be taken to mean the ability to understand an algorithm and its decision making through such things as simulatability, decomposability, and algorithmic transparency [5,16,20,21]. Haresamudram et al. argue that this is all algorithm-based and should have a much broader definition, which includes the whole life cycle. The definition would consist of algorithmic transparency, interaction transparency, and social transparency [21]. The different aspects of transparency are explained in Table 4.

How deep learning systems make their decisions is often known as the 'black-box problem' because although it is not clear how they make decisions, they give more accurate results. Wadden explains that the 'black-box problem' is not unique to AI; it occurs in other fields. Agency is a black box in the field of youth sociology; in race and social issues, fertility and childbirth are black-boxes [22]. The use of this term in other fields can lead to misunderstandings, particularly when looking to use XAI in, for example, medical settings. Burrell split the opacity problem into three parts, the first they referred to as self-protection and concealment, where algorithms are deliberately opaque to keep proprietary information from others [23]. The second opacity was due to the code being indecipherable to most members of the public, and the third opacity was that which comes from increasingly complex systems, where learning alters the original algorithm.

Sahakyan et al. also break this problem into three distinct parts: model explanation problems, model inspection problems, and outcome explanation problems [5]. They posit that a model explanation can be obtained by using a more transparent model to simulate the black-box's behavior. A model inspection explanation can be generated in visual or textual format of some properties of the model or its output. And an outcome explanation

can be the way an explanation of a particular instance has been generated or how it can be altered.

Markus et al. proposed, like Sahakyan et al., three separate sorts of explanations. The first is model-based explanations, where a simpler model is used to explain the more complex model. The second, attribution-based explanations, where the model is explained in terms of its input features. The third is example-based explanations, where individual instances are examined and used to explain how the model works [24].

**Table 4.** Showing definitions of different aspects of transparency.

Aspects of Transparency	Definitions	Reference
Transparency	Transparency does not ensure that a user will fully understand the system, but it does provide access to all relevant information regarding the training data, data preprocessing, system performance, and more.	[16]
Algorithmic transparency	Refers to the user's capacity to comprehend the process the model uses to generate a specific output based on its input data. The main limitation for algorithmically transparent models is that they must be entirely accessible for exploration through mathematical analysis and techniques.	[20]
Decomposability	Decomposability is the capacity to explain each component of a model, including its inputs, parameters, and calculations. This enhances the understanding and interpretation of the model's behavior. However, similar to algorithmic transparency, not all models can achieve this. For a model to be decomposable, each input must be easily interpretable, meaning complex features may hinder this ability. Additionally, for an algorithmically transparent model to be decomposable, all its parts must be comprehensible to a human without needing external tools.	[20]
Simulatability	This is a model's capacity to be understood and conceptualized by a human, with complexity being a main factor. Simple models like single perceptron neural networks fit this criterion, more complex rule-based systems with excessive rules do not. An interpretable model should be easily explained through text and visualizations. The model must be sufficiently self-contained for a person to consider and reason on it as a whole.	[20]
Interaction transparency	Is the clarity and openness in the interactions between users and AI systems? It involves giving users feedback they understand about the system's actions, decisions, and processes, allowing them to understand how their inputs influence outcomes. This transparency fosters trust and enables users to engage more effectively with technology, as they can see and understand the rationale behind the AI's behavior.	[21]
Social transparency	This is the openness and clarity of an AI system's impact on social dynamics and user interactions. It involves making the system's intentions, decision-making processes, and effects on individuals and communities clear to users and stakeholders. Social transparency helps users understand how AI influences relationships, societal norms, and behaviors, fostering trust and the responsible use of technology.	[21]

Brożek et al. approach this problem from a legal and psychological perspective, breaking it down into four distinct parts: the opacity problem, the strangeness problem, the unpredictability problem, and the justification problem [25]. They argue that 'the human mind is much more of a black box than the most sophisticated machine learning algorithm' [25]. In the case of the algorithms, how they work is known, even if how they arrived at their decision is not. How the human mind works is still largely unknown. They argue that the opacity problem is not a significant problem, as the workings of algorithms are much better known than the workings of the human mind. They contend that algorithms have been designed to uncover patterns beyond the grasp of the human mind. From their perspective, while explainability is valuable, it should not hold a higher priority than human decision making. They stated that it is the unpredictability of algorithmic decisions that leads to human skepticism. They refer to this phenomenon as 'strangeness' and suggest that explanations should be approached from a psychological perspective rather than a purely technical one.



Comprehensibility is another term often used when explaining AI. This is often understood as the ability of a learning algorithm to show its learned knowledge in a way that is understandable to humans. Others refer to the quality of the language used by an explainability method [18]. Allgaier et al. opted for a more technical approach and proposed measures to facilitate the examination of code, like providing percentage breakdowns for training, validation, and testing datasets. They argued that the comprehensibility of the XAI system becomes increasingly reliant on domain knowledge as the AI application becomes more specialized. Additionally, they recommended explaining variable names rather than relying on abbreviations, particularly in the context of tabular data, and cautioned against excessive length in explanations, such as overly complex decision trees [14]. Ali et al. found that for comprehensibility, explanations need to be short and readable [1]. Table 5 shows a summary of definitions and aspects of comprehensibility.

**Table 5.** Showing definitions of different aspects of comprehensibility.

Comprehensibility	Definitions	Reference
Comprehensibility	The clarity of the language employed by a method for providing explanations.	[18]
Comprehensible systems	Understandable systems produce symbols, allowing users to generate explanations for how a conclusion is derived.	[18]
Degree of comprehensibility	This is a subjective evaluation, as the potential for understanding relies on the viewer's background knowledge. The more specialized the AI application, the greater the reliance on domain knowledge for the comprehensibility of the XAI system.	[14]
Comprehensibility of individual explanations	The length of explanations and how readable they are.	[1]

Just as there are many ways of generating XAI explanations, there are also many ways of measuring the quality of the explanations. Usually, different XAI methods yield differing explanations on the same dataset and model; the 'best' metric is then selected from all the metrics. If this is unachievable, a new XAI method might be written [26]. What do you measure to ensure the quality of explanations? Various aspects of explanations are checked; Li et al. devised a method of assessing consistency and run-time efficiency of explanations [26]. Nauta et al. viewed explanation assessment as a twelve-faceted problem and developed a twelve-property quality evaluation [27]. Their twelve properties were correctness, consistency, completeness, continuity, contrastively, covariate completeness, compactness, composition, confidence, context, coherence, and controllability. Lopes et al. looked at evaluation methods as being either human centered or computer-centered and devised their own taxonomy for evaluations [28]. They looked at trust, explanation usefulness and satisfaction, understandability, performance, interpretability, and fidelity. Baptista et al. evaluated SHapley Additive explanations (SHAP) explanations via the established metrics of monotonicity, 'trendability', and 'prognosability', finding that SHAP did trend the metrics but that model complexity could be a problem [29]. Fouladgar et al. examined the sensitivities of time series XAI models; they found that the sensitivities varied according to how the hyperparameters were set [30]. Oblizanov et al., used synthetic data with local interpretable model-agnostic explanations (LIME) and SHAP and devised a metrics calculation tool. The tool used a modified faithfulness metric, a monotonicity metric, and an incompleteness metric. SHAP and LIME were comparable in terms of accuracy of explanation; however, the explanations were not as accurate when used with a decision tree compared to when they were used with linear regression [31].

### 3. Existing Techniques for Explainable Tabular Data Analysis

Sahakyan et al. explained that explanations could be divided into two groups: those that are model-specific and those that are model-agnostic. The model-specific explanations are obtained by exploiting the model's features and parameters; they also use the model's structure and its weights. Their main disadvantage is that they cannot usually be generalized and used on other types of models [5]. The model-agnostic explanations are generated after a model has been trained. They can be used to explain a variety of diverse types of models, but their main disadvantage is that the explanations may not be faithful to the underlying AI as they are produced by a different model.

Models are often categorized into two main types of explanations: local and global. Local explanations pertain to elucidating a single prediction, whereas global explanations aim to expound upon the entire model's behavior [32]. Taxonomies are established based on the distinct approaches adopted for providing explanations, which encompass functioning-based, result-based, conceptual, and mixed methods [32].

The functioning-based approach employs perturbation techniques, involving the manipulation of a model's inputs to identify the most influential features, primarily focusing on local explanations. Additionally, it can leverage the model's structural components, such as gradients, to determine feature importance. This approach also encompasses meta-explanations, architectural modifications, and illustrative examples. The result-based approach also relies on feature importance, surrogate models, and examples. The surrogate model is a simplified rendition of the original model. The conceptual approach classifies explanations according to overarching concepts, such as applicability, scope, problem type, and output type. The mixed approach represents a hybrid, incorporating elements from various other explanation approaches [32].

The explanations can be further divided into five techniques or methods [33]. The first is explanation by simplification, where the explanation is generated through distillation and rule extraction. The second is explanation by feature relevance, where the explanation is generated by measuring the amount of influence a feature has on prediction. The third is visual explanation, where explanations are in the form of pictures or plots providing information about the model's predictions. The fourth is explanation by concept, where a human-relatable concept is used to generate explanations [34]. The fifth is an explanation by example, where an explanation is generated by a factual example that justifies the outcome predicted [35].

There are many ways of generating explanations and many different taxonomies [32]. Figure 4 shows a taxonomy based on the variety of types of explanation and is the most like Sahakyan et al.'s, approach [33].

Several surveys on XAI and several aspects of XAI have been conducted recently [1,4,7,8,17,36–47], to name a few. Surveys on evaluating XAI have been conducted by [27], who looked at quantitative XAI evaluation methods, and [26], who defined two metrics, consistency and efficiency, to evaluate XAI methods.

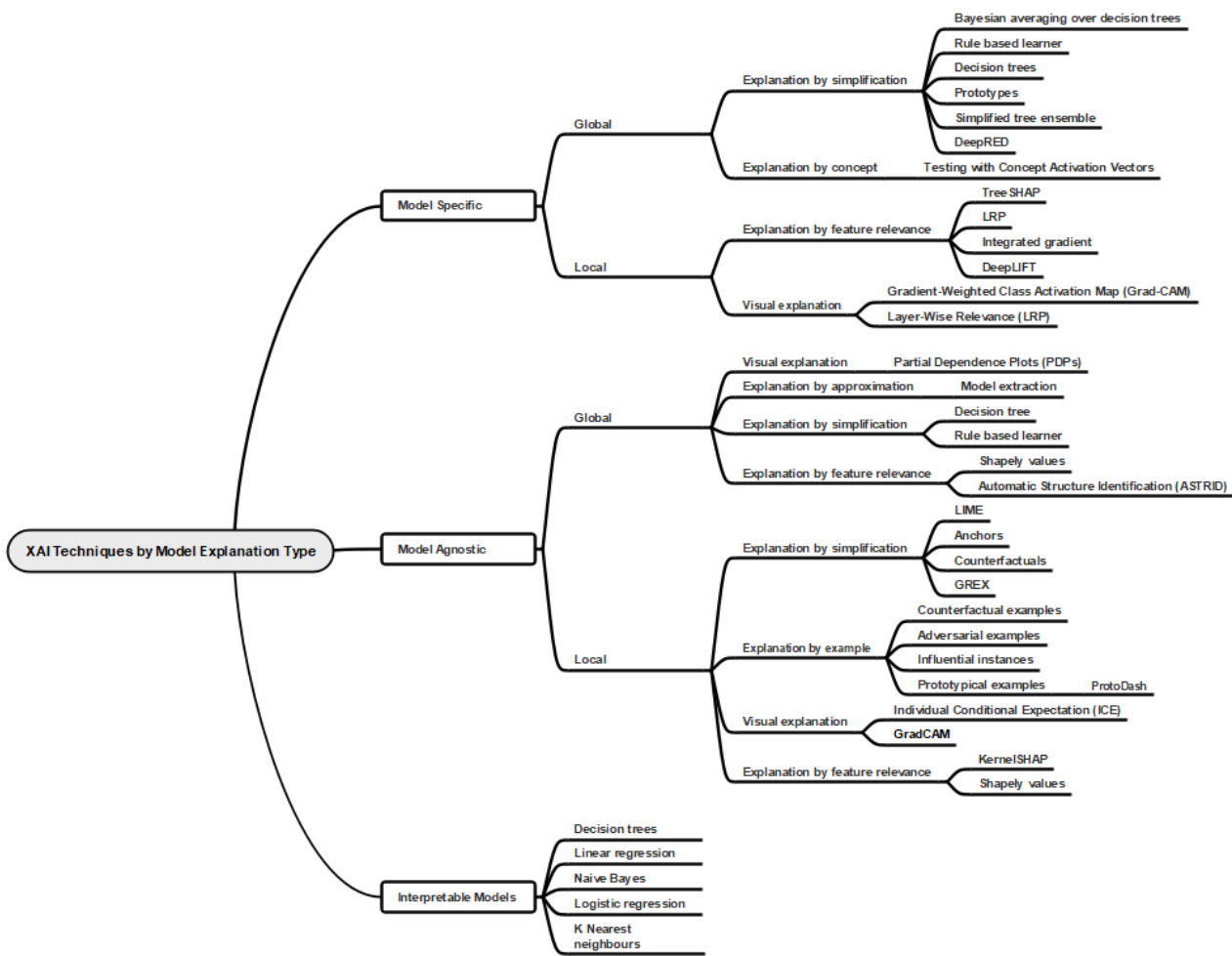
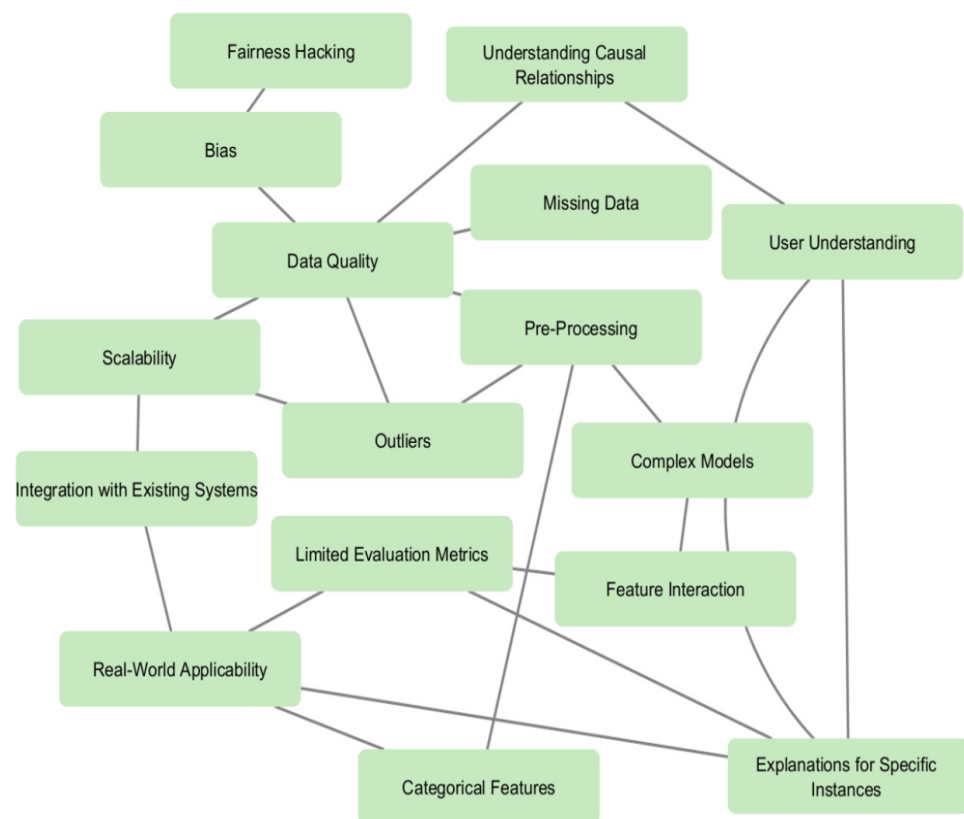


Figure 4. A taxonomy of XAI techniques by explanation type.

#### 4. Challenges and Gaps in Explainable Tabular Data Analysis

Explainable tabular data analysis plays a critical role in enhancing the transparency and interpretability of machine learning models applied to structured data. Despite its importance, several challenges and gaps persist in this domain, hindering its widespread adoption and effectiveness.

In their survey, Sahakyan et al. delved into the varied types of explanation and their adaptability to different data types. Their findings revealed that many explanation techniques were primarily developed for image and audio data or tailored to specific AI models, rendering them inadequately suited for tabular data types [5]. XAI techniques are tailored for model types or data forms, and many of these techniques can be complex and challenging to use. Moreover, these techniques face issues related to scalability and coverage. This disparity stems from the inherent attributes of tabular data, where the structure is defined by individual cells serving as fundamental units within the table. Each row in a tabular dataset encapsulates unique attributes, offering a row-centric perspective, while the columns consist of consistent data types, be it numeric or categorical values. Unlike image data, which exhibit spatial and geometric structures, the arrangement of columns in tabular data does not influence the integrity of the underlying information, leading to distinct challenges in constructing adaptable and effective explanation methods for this specific data format. Figure 5 illustrates the challenges and gaps in explainable tabular data analysis, and how interrelated they are.



**Figure 5.** A network view of the challenges and gaps in explainable tabular data analysis.

#### 4.1. Challenges of Tabular Data

The challenges inherent in tabular data are multifaceted, arising from issues such as data quality, outliers, missing values, imbalanced classes, and high dimensionality. Despite the abundance of data in tabular datasets, a sizable portion may be of inferior quality, introducing challenges for accurate analysis and modeling. Outliers, representing extreme values within the data, can skew results and impact the performance of machine learning models. Additionally, the presence of missing values further complicates the analysis process, requiring robust imputation techniques to manage the gaps in the data effectively. Imbalanced class distributions pose challenges in classification tasks, where certain classes are underrepresented, potentially leading to biased model outcomes. Moreover, the high dimensionality of tabular data, characterized by many features, can overwhelm traditional machine learning algorithms and hinder model performance.

#### 4.2. Bias, Incomplete and Inaccurate Data

Tabular data, by their very nature, reflect the real-world phenomena it aims to capture. If the data are incomplete, biased, or contain inaccuracies, then the insights and explanations derived from that data are inherently limited. For example, if a dataset is missing key variables that are important predictors of the target outcome, the XAI techniques may end up highlighting less relevant features as being important, leading to incomplete or potentially misleading explanations.

Similarly, if the data contain systematic biases, such as under representation of certain demographic groups, the explanations generated may overlook or fail to account for these biases, failing to provide a comprehensive and fair understanding of the model's decision-making process. This can occur deliberately and is known as fairness hacking. It is the unethical practice of adding or removing sensitive attributes from the testing to lead outsiders to believe that the results are fair [48]. This practice also facilitates the illegitimate exploitation of the many definitions of fairness. This happens when tests are performed

using many different fairness metrics, but only the metrics that produce positive results are reported. During the process of fairness hacking, the fairness shortcomings in ML algorithms are hidden by reporting that these algorithms are fair by certain metrics, while the metrics that have produced negative results are suppressed. This practice undermines transparency and accountability of explanations, as it allows organizations to misrepresent the true fairness of their AI systems. It is important that fairness testing is conducted comprehensively and rigorously, with all results being reported.

Noisy, or erroneous data can introduce confounding factors, making it challenging for XAI methods to disentangle the true underlying relationships and provide accurate explanations.

#### 4.3. Explanation Quality

The quality of explanations can be heavily dependent on the quality and completeness of the data used to train the machine learning models and generate the explanations. Careful data curation, validation, and feature engineering become essential prerequisites for deploying effective and trustworthy XAI systems, particularly in high-stakes applications where the explanations play a crucial role in decision making. Addressing data quality concerns should be a key consideration in the development and deployment of XAI techniques for tabular data analysis.

#### 4.4. Scalability of Techniques

The scalability of explainable techniques poses a significant challenge in the context of tabular data, especially as modern datasets continue to increase in both size and complexity. For example, in healthcare, tabular patient records encompass a wide range of medical information, including diagnoses, treatment histories, laboratory results, and imaging data. As these datasets expand, existing explainable methods may struggle to efficiently process and interpret the burgeoning volume of tabular healthcare data [49]. Moreover, large-scale tabular datasets, containing thousands or even millions of records, are increasingly prevalent in various domains such as financial transactions, e-commerce sales, and scientific research. This necessitates the development of scalable explainable techniques capable of delivering timely and resource-efficient explanations without compromising accuracy, particularly when analyzing vast tabular datasets to detect anomalies in financial transactions or uncover potential disease markers in extensive genomic data. An example of existing XAI used with tabular data is local interpretable model-agnostic explanations. It generates explanations for specific instances, not for the whole dataset, and it can be resource-intensive with complex models or vast datasets. Therefore, addressing the scalability gap is crucial to ensure that explainable AI techniques remain practical and effective in navigating and extracting meaningful insights from complex tabular datasets across diverse domains, especially as the volume and complexity of tabular data continue to grow at an unprecedented rate.

#### 4.5. Neural Networks

Neural networks, known for their ability to manage complex patterns in data, face limitations when dealing with such data quality issues as those found in tabular datasets. Consequently, extensive pre-processing steps are essential to address outliers, missing values, and other quality-related issues before leveraging neural networks for analysis. Another critical challenge lies in the inherent complexity of learning the intricate relationships between features in each dataset iteration, necessitating the relearning of structures and patterns in each training cycle. This constant relearning process renders traditional inductive biases used in deep learning less effective when applied to tabular data.

Furthermore, the presence of categorical features in tabular data adds another layer of complexity, as deep learning algorithms traditionally struggle with processing this type of data efficiently. Ensuring accurate feature importance and stability in predictions becomes imperative, as small fluctuations in feature values can significantly impact the

overall model predictions and result in unreliable outcomes [50]. The interplay of these challenges underscores the need for advanced pre-processing techniques, specialized model architectures, and robust interpretability methods to effectively navigate and extract meaningful insights from complex tabular datasets [51]. Determining how each feature contributes to the model predictions in the context of tabular data can be intricate, especially when using complex machine learning algorithms. Explaining the significance of each feature accurately for decision making is not straightforward.

#### 4.6. XAI Methods

It is common for machine learning models that are used in tabular data analysis to include random forests and boosting techniques, as documented by Tjoa & Guan [52]. However, Borisov et al.'s observations indicated that the most prevalent types of explainable artificial intelligence explanations for tabular data encompass feature-highlighting explanations and counterfactual explanations [50]. Despite their effectiveness it is worth noting that these models may not offer the same level of interpretability as simpler models such as single decision trees or logistic regression. Moreover, as the capabilities of deep learning algorithms continue to advance, working with tabular data remains a persistent challenge.

Table 6 and Figure 6 show some of the XAI methods used with tabular data, along with their pros and cons. The types of XAI have been classified into five groups: counterfactual explanations, feature importance, feature interactions, decision rules, and simplified models. Counterfactual explanations are a sophisticated tool for analyzing how minimal perturbations in input features can influence model predictions. For example, they can illustrate scenarios such as, "If the income were £5000 higher, the loan would be approved". These explanations typically use advanced optimization techniques to identify data points in proximity to the original input that result in altered outcomes, utilizing methodologies such as perturbation, generative models, or local interpretable models. While counterfactuals provide clear and actionable insights, they often involve significant computational resources and must ensure that changes remain plausible to avoid user confusion. Moreover, they may yield inconsistent results when subjected to minor variations in input data [53]. In contrast, feature importance methods assess the quantitative impact of each feature on the model's predictions, utilizing techniques like permutation importance and Shapley additive explanations. These approaches identify key features that significantly influence outcomes and can guide model enhancements. However, they often assume that features are independent, which can result in misleading conclusions, particularly when features are highly correlated. Moreover, inconsistencies may occur between local and global importance evaluations, and even minor changes to the dataset can cause significant variations in importance rankings [54]. Feature interactions explore the combined effects of multiple features on predictions, offering a more detailed understanding through statistical analyses and visualizations like interaction plots. However, as the number of features grows, complexity can complicate both interpretation and visualization. Additionally, models that emphasize feature interactions may be prone to overfitting, especially in datasets with a limited number of samples [54]. Decision rules are logical statements that typically take the form 'If [Condition], then [Outcome]'. It specifies a conditional relationship where the outcome or decision is determined if the conditions specified in the rule are satisfied [55]. The final group are simplified models. Hassija et al. refer to transferring 'dark knowledge', complex and hidden insights from sophisticated black-box models to simpler ones like decision trees. This transfer allows these streamlined models to match the predictive capabilities of more complex systems while enhancing interpretability [54]. Simplified models, such as linear approximations or decision trees, aim to approximate the decision boundaries of complex models while preserving interpretability. While these models generally exhibit faster computation times and enhanced understandability, they often incur substantial accuracy losses and may introduce biases due to their restricted representation of feature relationships. Their effectiveness can also diminish outside the specific context of the evaluated instance, leading to a false sense of model reliability.

**Table 6.** Types of XAI for tabular data and their pros and cons.

Summary of XAI Types					
Type of XAI	Description	Examples	Pros	Cons	Evaluation
Counterfactual explanations	Counterfactual explanations illustrate how minimal changes in input features can change the model's prediction, e.g., "If income increases by £5000, the loan is approved".	DiCE	Causal insight—understand the causal relationship between input features and predictions.	Complexity—generating counterfactuals is computationally intensive, particularly for complex models and high-dimensional data.	Alignment with predicted outcome—ensuring the generated counterfactual instances closely reflect the intended predicted outcome.
		WatcherCF	Personalized explanations—tailors individualized insights for better insights.	Complexity—generating counterfactuals is computationally intensive, particularly for complex models and high-dimensional data.	Alignment with predicted outcome—ensuring the generated counterfactual instances closely reflect the intended predicted outcome.
		GrowingSpheresCF	Decision support—aids decision making with actionable outcome-focused changes	Model specificity—effectiveness is influenced by the underlying model's characteristics.	Proximity to original instance—maintaining similarity to the original instance whilst altering the fewest features possible.
				Interpretation—conveying implications can necessitate domain expertise.	Diverse outputs—capable of producing multiple diverse counterfactual explanations.
					Feasible feature values—the counterfactual features should be practical and adhere to the data distribution.
Feature importance	Feature importance methods assess how much each feature contributes to the model's predictions.	Permutation Importance	Helps in feature selection and model interpretability.	May not capture complex feature interactions.	Relative importance—rank features based on their contribution to the model's prediction.
		Gain Importance.	Provides insight into the most influential features driving the model's decisions.	Can be sensitive to data noise and model assumptions.	Stability—ensure consistency of feature importance over different subsets of the data or re-trainings of the model.

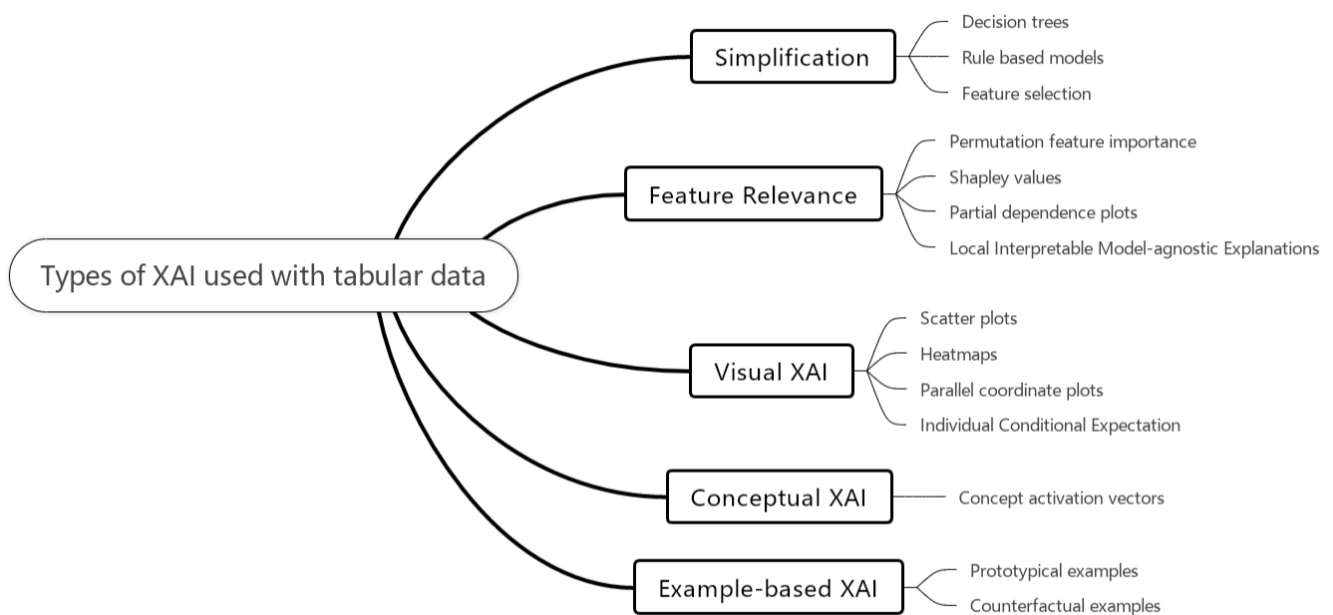
Table 6. Cont.

Summary of XAI Types					
Type of XAI	Description	Examples	Pros	Cons	Evaluation
		SHAP			Model impact—Assessing the influence of individual features on the model's predictive performance
		LIME			
Feature interactions	Feature interaction analysis looks at how the combined effect of multiple input features influences the model's predictions.	Partial Dependence plots	Reveals intricate and synergistic connections among features.	Visualizing and interpreting features can be difficult, especially when dealing with high-dimensional data.	Non-linear relationships—uncovers and visualizes complex, non-linear interactions among the features.
		Accumulated Local Effects plots.	Enhances insight into the model's decision-making mechanism.	The computational complexity grows as the number of interacting features increases.	Holistic insight—provides a comprehensive understanding of how features collectively impact the model's predictions.
		Individual Conditional Expectation Plots			Predictive power—evaluates the combined effects of interacting features on the model's performance.
		Interaction Values			
Decision rules	Decision rules provide clear, human-readable guidelines derived from the model, such as "If age > 30 and income > 50k, then approve loan".	Decision Trees	Provides clear and intuitive insights into the model's predictions.	Might struggle to capture complex relationships in the data, leading to oversimplification.	Transparency—offers clear and interpretable explanations of the conditions and criteria used for decision making.
		Rule-Based Models	Easily understood by non-technical stakeholders.	Can be prone to overfitting, reducing generalization performance	Understandability—ensures ease of understanding by non-technical stakeholders and experts alike.
		Anchors			Model adherence—check that decision rules capture accurately the model's decision logic without oversimplification.



Table 6. Cont.

Summary of XAI Types					
Type of XAI	Description	Examples	Pros	Cons	Evaluation
Simplified models	Simplified models are interpretable machine learning models that approximate the behavior of a more complex black-box model	Generalized Additive Models.	Gives a balance between model interpretability and model complexity.	Might not capture the total complexity of the underlying data generating process.	Balance of complexity—achieves an optimal compromise between model simplicity and predictive performance.
		Interpretable Tree Ensembles.	Offers global insights into the model's decision-making process	Needs careful model choice and tuning to maintain a good trade-off between interpretability and accuracy.	Interpretable representation—ensures that the offers transparent and intuitive insights into the original complex model's behavior.
					Fidelity to original model—Assesses the extent to which the simplified model captures the key characteristics and patterns of the original complex model.



**Figure 6.** The main types of XAI used with tabular data.

Borisov et al., further pinpointed three fundamental challenges pertaining to tabular data and deep learning, namely inference, data generation, and interpretability [50]. Traditional machine learning models such as decision trees or logistic regression are inherently more interpretable compared to complex models like neural networks commonly used with tabular data. Therefore, explaining the predictions of these more advanced models becomes a challenge. Their studies revealed that decision tree ensembles outperformed deep learning models in the context of tabular data analysis. Additionally, their research included assessing the interpretability of these models, leading to the conclusion that improved benchmarks for the claimed interpretability characteristics and their pragmatic applicability are necessary requirements. This research underscores the imperative for comprehensive assessments and benchmarks to validate the interpretability of models and their suitability for real-world applications. These findings encourage further research and development to address the challenges associated with deep learning algorithms and their application to tabular data [50].

#### 4.7. Benchmark Datasets

In the realm of tabular data, there is a significant lack of universally accepted benchmark datasets and evaluation frameworks. Unlike the fields of computer vision and natural language processing, which have well-established standard benchmark datasets and evaluation protocols, the landscape of tabular data remains unstandardized. Two benchmarks used on XAI for tabular data are synthetic tabular datasets and intrusion detection datasets. Synthetic datasets are datasets that mimic the structure and complexity of real-world tabular data, providing standardized testing environments for evaluating XAI techniques [5]. Whereas the intrusion detection datasets are real-world datasets from domains like network security, such as flow-based network traffic data, used to assess the performance of XAI approaches in practical applications [5].

Without clear benchmarks, comparing the performance and effectiveness of various XAI techniques for tabular data becomes challenging. Researchers and practitioners typically use their own custom datasets or evaluation metrics, complicating meaningful comparisons and assessments of the generalizability of proposed methods. This lack of standardization also impedes the development of best practices, making it difficult to identify the most promising XAI approaches for specific tabular data problems.

#### 4.8. Scalability

Tabular data often consist of multiple features or attributes with varying types and scales, requiring XAI techniques to address the complexity and dimensionality of the data in a distinct manner [56]. Non-tabular data, such as images or text, have different properties and may need specialized approaches for effective interpretation. The scalability of XAI techniques varies notably between tabular and non-tabular data. Tabular data present a particularly pronounced scalability challenge due to the growing volume and intricacy of datasets across diverse domains.

#### 4.9. Data Structure

The XAI used for tabular data has similarities but also differences when compared to the XAI used with other forms of data. The key differences discussed are interpretability techniques, where feature importance and local data instances are used more commonly with tabular data due to the different nature of the data [56]. The inherent structure of tabular data, with rows and columns, introduces unique challenges and opportunities for XAI techniques compared to the non-tabular formats of other data types. For example, the interaction between features in tabular data may require specific methods to interpret and explain the model's decision process [42].

For tabular data, XAI techniques often focus on interpreting the importance and interactions of the various features or variables within the dataset. Methods such as Shapley additive explanations (SHAP), local interpretable model-agnostic explanations (LIME), and feature importance analysis are commonly employed to identify the key drivers behind the model's outputs. These techniques leverage the structured and quantitative nature of tabular data to provide insights into the model's decision-making process.

In contrast, XAI for image or text data may involve different approaches, such as visualizing the regions of an image that are most salient to the model's prediction or highlighting the specific words or phrases that contribute the most to a text classification outcome. These modality-specific techniques take advantage of the unique properties of the data, such as spatial relationships in images or semantic dependencies in text.

Despite these differences, there are some common themes and approaches that span multiple data types. For example, the concept of counterfactual explanations, which explore "what-if" scenarios to understand how changes in the input would affect the model's output, can be applied to both tabular and unstructured data.

#### 4.10. Model Evaluation and Benchmarks

To address the challenge of evaluating the practical usefulness of interpretability methods, there is a growing emphasis on creating improved benchmarks for assessing the interpretability characteristics of models used with tabular data [57]. These benchmarks help validate the effectiveness of interpretability methods in real-world applications, ensuring their reliability and utility.

Additionally, when dealing with tabular data, the black-box nature of certain machine learning models, such as deep neural networks, presents a fundamental gap in transparency [54]. In such cases, post-hoc explainable techniques like SHAP and LIME may offer insights into model predictions but may not fully capture the underlying mechanisms of these complex models, particularly when applied to extensive tabular datasets [58].

Moreover, the lack of consensus and standardization in assessing these explanations adds another layer of complexity. The absence of consistent evaluation methods and terminology in explainable AI (XAI) further complicates the comprehensive assessment of explanations, especially in the context of diverse disciplines dealing with extensive tabular datasets. Therefore, addressing scalability and standardization gaps becomes crucial to ensure the practicality and effectiveness of explainable AI techniques when analyzing large and complex tabular datasets, such as those found in healthcare, finance, and scientific research [59].

#### 4.11. Review

Addressing future challenges in explainable artificial intelligence requires a multi-faceted strategy that encompasses technological advancements, regulatory frameworks, and ethical considerations. Enhancing interpretability could be achieved by developing user-centric explanations tailored to various audiences and establishing standardized formats for clarity. Investing in advanced visualization tools could enable users to intuitively grasp model behavior. Tackling model complexity would involve promoting inherently interpretable models and adopting hybrid approaches that merge complex models with interpretable surrogates. Robust data governance is vital for maintaining data quality and fairness, alongside regular fairness audits to identify and mitigate biases. Building trust and accountability calls for transparency in model training and embedding explainability as a fundamental principle from the design phase. Collaboration among AI researchers, ethicists, and domain experts is essential for exploring new methodologies, supported by open-source initiatives and shared benchmarks for evaluating XAI methods. Ongoing education and training for stakeholders would enhance their understanding of XAI concepts, while incorporating ethics into AI curricula would prepare future practitioners to address the moral implications of their work. By implementing these strategies, the field of XAI could evolve to be more effective, transparent, and trustworthy, fostering greater acceptance and application of AI technologies in critical domains.

While XAI methods for tabular data share some similarities with those for other data formats, the unique challenges and characteristics of tabular data necessitate the development of specialized techniques to provide comprehensive and actionable explanations. Understanding these nuances is crucial for selecting and applying the most appropriate XAI approaches for a given problem and data context.

In summary, while explainable techniques hold promise for enhancing the transparency and trustworthiness of machine learning models applied to tabular data, several challenges and gaps must be addressed to realize their full potential. Overcoming these obstacles will require interdisciplinary collaboration among researchers, practitioners, and policymakers to develop innovative solutions and establish best practices.

### 5. Applications of Explainable Tabular Data Analysis

Explainable tabular data analysis has various uses across different domains where understanding and explaining the results of machine learning models used on tabular data are crucial. XAI is an interdisciplinary research field, and its applications can be found in finance, healthcare, autonomous vehicles, and many more areas [11].

#### 5.1. Financial Sector

Examples of explainable tabular data analysis use cases in the financial sector are identity verification in client onboarding, transaction data analysis [60], fraud detection in claims management [60,61], anti-money laundering monitoring [62], financial trading [63], risk management [62], and the processing of loan applications, to name a few [60,61]. Černevičienė and Kabašinskas identified that credit management was an important research area [60].

In the finance industry, banks and lending institutions use explainable tabular data analysis to explain credit scoring models. Customers can receive explanations for not only why their credit applications were accepted or rejected but also to help them understand what factors influenced the decision [64]. Counterfactual explanations describe what changes in the input data would result in a different outcome from a predictive model. These explanations provide insights into the “what-if” scenarios by highlighting the key features or variables that, if different, could have led to an alternate prediction or outcome. Counterfactual explanations are particularly valuable in machine learning interpretability as they help users understand the sensitivity of a model’s predictions to changes in input variables and provide actionable insights for improving outcomes or understanding the

model's decision-making process. These explanations can be used to give customers an idea of what they need to do to improve their credit scores.

An important area is risk management, which encompasses fraud detection, loan default, and bankruptcy prediction [62]. XAI can identify patterns in transaction data and explain why certain transactions are flagged as fraudulent. Post-hoc explanations such as visual explanations, explanations by simplification, and feature relevance are the types of XAI used.

Retail banking customer churn is another area in which XAI can be used in order to predict which customers are likely to leave and why [65]. The types of explanations used in this case are SHAP and LIME. The data can give information to plan a retention scheme.

### 5.2. Healthcare Sector

In healthcare, explainable tabular data analysis can be applied to explain the model predictions used for patient risk assessment, diagnosis, or treatment recommendations [63]. The most commonly used type of explanation is post-hoc explanations such as LIME and SHAP. Doctors and patients can benefit from the explanations in making personalized healthcare decisions.

Patient diagnosis can be achieved by using tabular data from patient records. XAI can help explain predictions made by models regarding potential diagnoses or treatment plans [66,67].

Drug efficacy can be achieved by analyzing patient responses to treatments across various demographics. The role of XAI is in helping to understand which factors influence drug effectiveness [68].

Context-sensitive explanations in the medical field have been researched [69]. This is useful when trying to provide personalized healthcare because there are multiple departments with their own types of data and with their own terminology; explanations will be needed to provide explanations to people with differing levels of understanding. By using context-sensitive explanations, the explanations can be tailored to each different setting. Multimodal explanations are also used to integrate diverse forms of information to provide more comprehensive and meaningful insights into the reasoning behind AI model predictions or decisions [70].

Cloud-based explainability services are cloud-hosted platforms or tools that offer explanations and interpretability for machine learning models and AI systems. These services are designed to help organizations and developers analyze, understand, and validate the decisions made by AI models, especially when deployed in cloud environments. Research has been undertaken in the healthcare sector to use XAI on cloud-based and edge-based electronic health records (EHRs); post-hoc explanations such as LIME, SHAP, and counter-factual explanations were used [71].

### 5.3. Fraud

Explainable tabular data analysis can be used in the field of fraud detection to provide explanations for why a transaction or an anomaly was flagged as potentially fraudulent [60,61,64]. This helps fraud analysts understand the model's decision and take appropriate actions in a timely fashion. Anomaly detection methods, like isolation forest and local outlier factor (LOF), are frequently employed for fraud detection. These techniques pinpoint instances that deviate significantly from the majority of the data, and the identified anomalies can be analyzed to uncover the reasons for their classification as fraudulent.

Research has been done on real-time interpretability, which is where explanations are provided in real time [72]. This means that a system is able to generate explanations or justifications for its outputs promptly and continuously as new data are processed and predictions are made.

#### 5.4. Retail Sector

It is used in business to explain why a customer is predicted to churn [73]. This can include highlighting the key features contributing to the churn risk and allowing companies to take targeted actions to retain their customers. In human resources (HR) departments, it can be employed to understand why employees are likely to leave a company. This can help with retention strategies and the improvement of workplace conditions. Insurance companies use this type of analysis to explain how premiums are calculated and why certain people receive higher or lower rates. This transparency can build and increase trust with policyholders.

In car dealerships, XAI can be used, with regression models, to regularly evaluate the performance of their service advisors [74]. The service advisors are the public face of the dealerships and need to retain and recruit customers.

Customer segmentation is the categorization of current and prospective clients by common attributes. XAI can help explain the reasoning behind segmenting customers based on purchasing behavior and demographics [75].

Context awareness explanations have been used in recommender systems to improve suggestions made to the customer [76].

#### 5.5. Manufacturing Sector

Manufacturing companies can use XAI to explain why a product failed a quality control check. This can lead to work and process improvements and reduce the number of defects. In logistics and supply chain management, this sort of analysis can explain decisions related to inventory management, shipping routes, and order fulfillment, assisting with better decision making [77,78]. E-commerce platforms use this type of analysis to explain product recommendations to customers. It can highlight the features or historical behavior that influenced the recommendations.

Manufacturers make use of decision support systems (DSSs) to assist with quality control, process control, planning, and scheduling [79]. DSSs allow for the integration of transparent and interpretable machine learning models and methodologies within DSSs to provide human users with understandable explanations for AI-generated recommendations, predictions, or decisions. The goal of incorporating XAI in DSS is to enhance human trust, comprehension, and acceptance of AI suggestions by offering insights into the rationale behind the system's outputs. XAI can help by using sensor data and historical maintenance records to explain predictions about equipment failures [80].

Integration of business intelligence (BI) refers to the process of combining XAI with traditional business intelligence tools and platforms to enhance data analysis, decision-making, and insight generation within an organization. By integrating XAI with BI systems, companies can leverage advanced analytics, machine learning models, and interpretability features to gain deeper and more actionable insights from their data. An example of this is where XAI workflows are tailored to support a manufacturer's daily operations and decision making by taking into account the interconnected value of individual data components, models, and services. This type of collaborative platform can use a variety of AI explainability techniques for a wide range of manufacturing applications [81].

#### 5.6. Utility Sector

Explainability as a service (EaaS) is a cloud-based solution that offers explainable artificial intelligence (XAI) capabilities through a subscription or pay-per-use model. It helps organizations to use advanced explainability tools for their AI applications without needing to develop these capabilities internally. Many sectors, such as healthcare, finance, human resources, and insurance, use EaaS.

It can be applied by utility companies to explain how they forecast energy consumption for customers. XAI is used in smart grid load balancing, where different types of data are used from several sources; some of the data are tabular data [82]. This transparency can help consumers make informed decisions about their energy usage. Explainable tabular data

analysis can also be used to explain why certain legal decisions or compliance assessments have been made. This can be critical where transparency and fairness are essential.

### 5.7. Education

In the education sector, XAI can show why students receive the grades they do or recommendations for courses. This can help teachers tailor their teaching methods. Student adaptation involves adjusting study habits, learning strategies, and time management skills to meet the demands of different subjects or educational levels. XAI has been used to predict levels of adaptability [83]. Nnadi et al. used SHAP, LIME, Anchor, and accumulated local effects (ALE), all of which are post-hoc explanations.

### 5.8. Summary

In these applications, explainable tabular data analysis not only improves model transparency but also enhances decision making, fosters trust, and assists in compliance with regulations, making it a valuable tool to use across a wide range of industries and use cases.

## 6. Future Directions and Emerging Trends

A significant amount of earlier research has concentrated on improving the quality of explanations and offering suggestions for future research. Nevertheless, there is still uncertainty surrounding the terminology employed in the field of XAI, highlighting the need for future research to establish standardized definitions to encourage a broader acceptance of XAI.

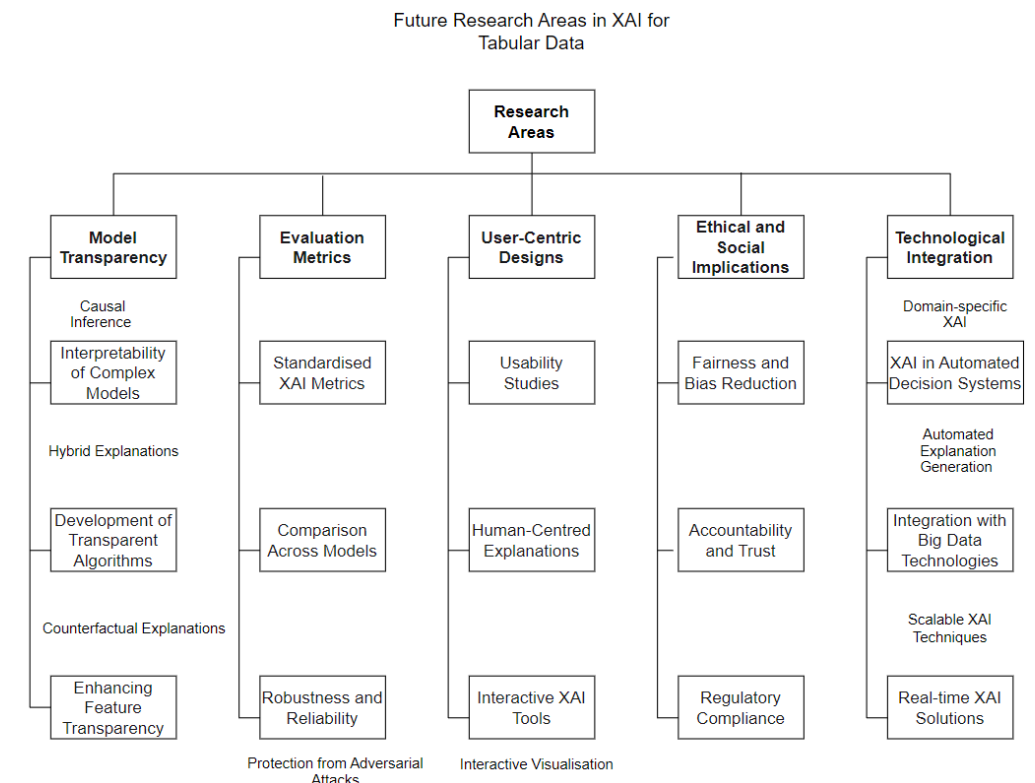
The work done by Vellido and Martín-Guerrero as far back as 2012 highlighted an important challenge in the field of explainable artificial intelligence (XAI)—the lack of publicly available tabular datasets that possess both annotated labels and concept-related attributes [84]. Alkhatib et al. noted that this gap in accessible data has hindered progress in specific research efforts focused on XAI techniques for tabular data. In 2023, they encountered obstacles in locating publicly accessible tabular datasets that possessed both annotated labels and concept-related attributes [85]. The creation of such datasets would encourage specific research into XAI and tabular data.

Currently, an area of research focus is on methods of adversarial example-based analysis; this is being done on natural language processing tasks and on images. Examination of these methods on tabular data could be a useful area of research [42]. Explanations can always be improved and created to be more user-focused; this would involve the development and use of intelligent interfaces capable of interacting with the user and generating relevant explanations [75]. In critical decision areas such as medicine, for example, the expert user would need to be able to interact with the system and would need to know that any mapping of ‘causability’ and explainability is extremely reliable [42]. This would give two areas of further research, relevant explanations, and user interaction with the model or system. Table 7 and Figure 7 give further suggestions for research ideas that have been obtained from the articles examined in this research.

Security has a significant role to play in promoting trust in the use of XAI. People want to know that their data is safely stored and that results from models using their data are only available to a restricted set of people. This is also an obligation under GDPR, where personal data should not be stored longer than necessary for the purposes for which it was collected; this is the principle of storage limitation. The research and development of XAI is particularly important in the context of the GDPR, which requires the implementation of appropriate technical and organizational safeguards to ensure the security of personal data against unauthorized or unlawful processing, accidental loss, destruction, or damage. Under GDPR fairness, hacking would come under unlawful processing.

**Table 7.** Suggested areas of further research.

Possible Research Areas	Suggestions
Hybrid Explanations [86]	Combining multiple XAI techniques to provide more comprehensive and robust explanations for tabular data models [86]. Integrating global and local interpretability methods to offer both high-level and instance-specific insights.
Counterfactual Explanations	Generating counterfactual examples that show how the model’s predictions would change if certain feature values were altered [87]. Helping users understand the sensitivity of the model to different feature inputs and how to achieve desired outcomes.
Causal Inference [85]	Incorporating causal reasoning into XAI methods to better understand the underlying relationships and dependencies in tabular data [88]. Identifying causal features that drive the model’s predictions, beyond just correlational relationships.
Interactive Visualizations	Developing interactive visualization tools that allow users to explore and interpret the model’s behavior on tabular data [89]. Enabling users to interactively adjust feature values and observe the corresponding changes in model outputs [89].
Scalable XAI Techniques [4]	Designing XAI methods that can handle the growing volume and complexity of tabular datasets across various domains [90]. Improving the computational efficiency and scalability of XAI techniques to support real-world applications.
Domain-specific XAI	Tailoring XAI approaches to the specific needs and requirements of different industries and applications that rely on tabular data, such as finance, healthcare, and manufacturing. Incorporating domain knowledge and constraints to enhance the relevance and interpretability of explanations [91].
Automated Explanation Generation [92]	Developing AI-powered systems that can automatically generate natural language explanations for the model’s decisions on tabular data [93]. Bridging the gap between the technical aspects of the model and the end-user’s understanding [93].



**Figure 7.** Future research areas in XAI for tabular data.



Another aspect of data security under GDPR is safety from adversarial attacks; as these are constantly changing, it is an area of constant research [8]. LIME and SHAP explanations for tabular data can be manipulated by exploiting their reliance on perturbing input data for estimation. The proposed attack substitutes a biased black-box model with a model surrogate to effectively conceal bias, for instance, from auditors. An out-of-distribution detector is trained to divide the input data so that the black-box model's predictions remain biased within the in-distribution, but its behavior on the perturbed data is controlled. This allows the explanations to seem fair, even though the underlying model may be biased. The aim is to use the fundamental mechanisms of LIME and SHAP, which rely on perturbing the input data to estimate the feature importance and give explanations. By carefully crafting a model surrogate that exhibits the desired behavior on the perturbed data, the real nature of the black-box model's biases can effectively be hidden. This technique presents a real and concerning threat, as it could enable the masking of harmful biases in high-stakes AI systems from regulatory oversight and public scrutiny. It emphasizes the need for robust and comprehensive testing of XAI methods to ensure they are not susceptible to such manipulative attacks [94]. SHAP is the XAI technique most susceptible to adversarial attack [94].

The newer EU AI Act takes a risk-based approach and classifies AI into four categories, with each being deemed less risky than the previous one. Depending upon the level of risk, different requirements apply. The AI Act places particular focus on high-risk AI systems, which are required to meet stringent requirements, including transparency and explainability. This means that providers of high-risk AI systems must ensure that their models can provide understandable explanations for their decisions and predictions. High-risk AI systems, such as those used in healthcare and credit scoring, are expected to be transparent and have human oversight; however, many such systems use black box algorithms and tabular data. This aspect of the requirements, as described in Articles 13 and 14, can be met using XAI. Users of high-risk AI systems have the right to receive clear information about how the AI system works, which necessitates the development and integration of XAI techniques to fulfill these obligations.

For intermediate systems like recommender systems and chatbots, users should be informed of the capabilities and limitations of the system, and in particular users should be told that they are interacting with a bot [95].

To facilitate meaningful human oversight, the AI Act requires that explanations be provided in a manner that humans can comprehend. This enhances the importance of XAI in making AI models interpretable and understandable. There is a requirement that AI systems, particularly high-risk ones, are designed in a way that allows for the traceability of decisions. XAI techniques help by providing detailed insights into how different inputs affect outputs, making it easier to audit and trace decisions.

The AI Act emphasizes the importance of ensuring that AI systems do not perpetuate biases or discrimination. For tabular data, XAI methods can help detect and explain biases in the system, allowing steps to promote fairness to be taken. Further research areas will no doubt become apparent as the Act is used more widely [96].

Existing XAI has few evaluation metrics to demonstrate how effective the explanations are. This could be a vast area of research, particularly when explanations are generated for a wide range of users, all with their own specific needs. There needs to be an accepted way of measuring not only the quality but also the satisfaction of explanations. This would benefit users and researchers as it would allow for a better comparison between models.

Future research in XAI should focus on establishing clear definitions, enhancing user-focused explanations, incorporating user interaction, ensuring data security, and devising robust evaluation metrics. Adversarial example-based analysis, user-specific explanations, and addressing security concerns are emerging areas of interest in XAI research [42,90]. Another key area is that of intra-metric measurements to avoid the selection of different favorable metrics and omitting the unfavorable metrics.

## 7. Conclusions

Explainable artificial intelligence has emerged as an important field of research in recent years, driven by the need to understand the complex, black-box nature of AI and machine learning models. XAI aims to provide more transparent and easily understandable explanations of how these models make decisions. This is crucial for building trust and enabling ethical decision-making, especially in safety-critical domains such as healthcare and finance. Given that tabular data are the predominant data format used in many industries, XAI becomes particularly important for these types of datasets. While several surveys have been conducted on XAI approaches, methods, and applications, there is a surprising lack of XAI techniques specifically tailored for tabular data, despite its widespread use across diverse disciplines. This was highlighted in Sahakyan et al.'s paper, and it remains the case. Overall, XAI is seen as a main driver for the broader adoption and trust in AI-powered decision-making systems.

Machine learning and XAI communities often use the terms “explainability” and “interpretability” interchangeably when referring to the goal of making AI systems more transparent and trustworthy. There is still a lack of agreed-upon formal definitions, which has led to the proliferation of different taxonomies and frameworks in the XAI field. Establishing agreed-upon mathematical definitions for these concepts could help provide a more unified foundation for XAI research and development.

The “black-box problem” is a multifaceted challenge, and researchers have proposed various frameworks for categorizing and addressing the various aspects of model opacity. The “black-box problem” also has legal, psychological, and technical dimensions, and researchers are exploring various approaches to improve the comprehensibility and explainability of AI systems beyond just the technical aspects.

Assessing the quality of XAI explanations is a multi-faceted challenge, with researchers exploring diverse technical and human-centric metrics and frameworks to tackle this problem. There are model-specific explanations that exploit internal model structures and model-agnostic explanations that are more generally applicable but may be less faithful to the underlying model. Taxonomies of explanation approaches further categorize the different explanations into five key techniques for generating XAI explanations. These are simplification, feature relevance, visual, conceptual, and example-based, which can be further organized into different taxonomical frameworks.

Despite the critical importance of explainable tabular data analysis, there are significant challenges and gaps that persist in this domain, stemming from the inherent properties of tabular data and the limitations of current XAI techniques. Addressing these challenges is crucial for enhancing the transparency and interpretability of machine learning models applied to structured data.

There are significant scalability challenges of explainable techniques for tabular data, as well as the limitations of neural networks in effectively managing the unique characteristics and quality issues present in tabular datasets. The complex challenges posed by categorical features, feature importance, and stability in tabular data require the development of advanced techniques, specialized architectures, and robust interpretability methods to effectively analyze and extract insights from complex tabular datasets.

There is also a critical need for improved benchmarks and assessment frameworks to validate the interpretability and practical usefulness of machine learning models, especially when applied to complex tabular datasets, which have unique challenges and scalability issues associated with applying XAI techniques, and there is a need for greater standardization and interdisciplinary efforts to address these gaps.

There is an urgent need for further research into XAI techniques that are robust against adversarial attacks. If malicious actors can exploit the vulnerabilities of prominent explanation methods to obscure harmful biases, it poses a grave threat to the integrity and trustworthiness of AI systems, particularly in high-stake domains. Developing XAI approaches that are impervious to such manipulative tactics is paramount, as it would ensure that the explanations provided are a faithful and accurate reflection of the model's

true decision-making process. This, in turn, would bolster transparency, accountability, and public trust in the deployment of AI, allowing for more rigorous auditing and oversight. The pursuit of XAI techniques resistant to adversarial attacks is a crucial frontier in using the full potential of AI while safeguarding against its misuse.

The key research challenges and future directions include standardizing terminology, addressing data availability, adapting adversarial analysis to tabular data, and developing more user-focused and reliable explanations, particularly for critical applications. The exploration and advancement of XAI must consider the security and regulatory requirements, particularly under the GDPR and the EU AI Act, while also addressing the need for robust evaluation metrics and user-focused explanations.

This survey has examined the language and terminology used in XAI generally so as to clarify XAI and make it clearer to those using it. Sahakyan et al.'s paper explained models that could be used with tabular data along with their pros and cons. This paper has sought to build on their paper by discussing the terminology used, the types of XAI applications used in different domains, and has sought to highlight challenges and gaps that are still present and to suggest further areas that could be carried out in this field.

Explainable AI has emerged as a critical field of research, driven by the need to build trust and enable ethical decision making in the face of complex, black-box AI models. As tabular data remains the predominant format used across many industries, the development of XAI techniques tailored for structured data is of paramount importance. However, the research community still faces significant challenges in establishing agreed-upon definitions, addressing the unique characteristics of tabular data, and devising robust evaluation frameworks. Addressing these gaps is crucial not only for enhancing the transparency and interpretability of AI systems but also for ensuring compliance with evolving security and regulatory requirements, such as those outlined in the GDPR and the EU AI Act. By focusing on user-centric explanations, robust evaluation metrics, and the incorporation of security best practices, the XAI research community can pave the way for the broader adoption and trust in AI-powered decision-making systems across a wide range of safety-critical domains.

**Author Contributions:** Conceptualization, H.O.Q.; methodology, H.O.Q.; software, H.O.Q.; investigation, H.O.Q.; writing—original draft preparation, H.O.Q.; writing—review and editing, M.S. (Mohamed Sedky) and H.O.Q.; supervision, M.S. (Michael Streeton) and J.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project was supported by Connexica Ltd. and Innovate UK [Grant Reference: KTP10021376].

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** Michael Streeton was employed by Connexica Ltd., Unit D, Dyson Court, Dyson Way, Stafford Technology Park, Staffordshire ST18 0LQ, UK. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. 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](#)]
2. Burkart, N.; Huber, M.F. A Survey on the Explainability of Supervised Machine Learning. *J. Artif. Intell. Res.* **2021**, *70*, 245–317. [[CrossRef](#)]
3. Weber, L.; Lapuschkin, S.; Binder, A.; Samek, W. Beyond explaining: Opportunities and challenges of XAI-based model improvement. *Inf. Fusion* **2023**, *92*, 154–176. [[CrossRef](#)]
4. Marcinkevičs, R.; Vogt, J.E. Interpretable and explainable machine learning: A methods-centric overview with concrete examples. *WIREs Data Min. Knowl. Discov.* **2023**, *13*, e1493. [[CrossRef](#)]
5. Sahakyan, M.; Aung, Z.; Rahwan, T. Explainable Artificial Intelligence for Tabular Data: A Survey. *IEEE Access* **2021**, *9*, 135392–135422. [[CrossRef](#)]

6. Alicioglu, G.; Sun, B. A survey of visual analytics for Explainable Artificial Intelligence methods. *Comput. Graph.* **2021**, *102*, 502–520. [CrossRef]
7. Cambria, E.; Malandri, L.; Mercurio, F.; Mezzanzanica, M.; Nobani, N. A survey on XAI and natural language explanations. *Inf. Process. Manag.* **2023**, *60*, 103111. [CrossRef]
8. Chinu, U.; Bansal, U. Explainable AI: To Reveal the Logic of Black-Box Models. *New Gener. Comput.* **2023**, *42*, 53–87. [CrossRef]
9. Schwalbe, G.; Finzel, B. A comprehensive taxonomy for explainable artificial intelligence: A systematic survey of surveys on methods and concepts. *Data Min. Knowl. Discov.* **2023**, *38*, 3043–3101. [CrossRef]
10. Yang, W.; Wei, Y.; Wei, H.; Chen, Y.; Huang, G.; Li, X.; Li, R.; Yao, N.; Wang, X.; Gu, X.; et al. Survey on Explainable AI: From Approaches, Limitations and Applications Aspects. *Hum.-Centric Intell. Syst.* **2023**, *3*, 161–188. [CrossRef]
11. Hamm, P.; Klesel, M.; Coberger, P.; Wittmann, H.F. Explanation matters: An experimental study on explainable AI. *Electron. Mark.* **2023**, *33*, 17. [CrossRef]
12. Lance, E. *Ways That the GDPR Encompasses Stipulations for Explainable AI or XAI*; SSRN, Stanford Center for Legal Informatics: Stanford, CA, USA, 2022; pp. 1–7. Available online: <https://ssrn.com/abstract=4085089> (accessed on 30 August 2023).
13. Gunning, D.; Vorm, E.; Wang, J.Y.; Turek, M. DARPA’s explainable AI (XAI) program: A retrospective. *Appl. AI Lett.* **2021**, *2*, e61. [CrossRef]
14. Allgaier, J.; Mulansky, L.; Draelos, R.L.; Pryss, R. How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare. *Artif. Intell. Med.* **2023**, *143*, 102616. [CrossRef] [PubMed]
15. Graziani, M.; Dutkiewicz, L.; Calvaresi, D.; Amorim, J.P.; Yordanova, K.; Vered, M.; Nair, R.; Abreu, P.H.; Blanke, T.; Pulignano, V.; et al. A global taxonomy of interpretable AI: Unifying the terminology for the technical and social sciences. *Artif. Intell. Rev.* **2022**, *56*, 3473–3504. [CrossRef]
16. Bellucci, M.; Delestre, N.; Malandain, N.; Zanni-Merk, C. Towards a terminology for a fully contextualized XAI. *Procedia Comput. Sci.* **2021**, *192*, 241–250. [CrossRef]
17. Barbiero, P.; Fioravanti, S.; Giannini, F.; Tonda, A.; Lio, P.; Di Lavore, E. Categorical Foundations of Explainable AI: A Unifying Formalism of Structures and Semantics. In *Explainable Artificial Intelligence. xAI, Proceedings of the Communications in Computer and Information Science, Delhi, India, 21–24 May 2024*; Springer: Cham, Switzerland, 2024; Volume 2155, pp. 185–206. [CrossRef]
18. Vilone, G.; Longo, L. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Inf. Fusion* **2021**, *76*, 89–106. [CrossRef]
19. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [CrossRef]
20. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Benetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
21. Haresamudram, K.; Larsson, S.; Heintz, F. Three Levels of AI Transparency. *Computer* **2023**, *56*, 93–100. [CrossRef]
22. Wadden, J.J. Defining the undefinable: The black box problem in healthcare artificial intelligence. *J. Med Ethic* **2021**, *48*, 764–768. [CrossRef]
23. Burrell, J. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data Soc.* **2016**, *3*, 1–12. [CrossRef]
24. Markus, A.F.; Kors, J.A.; Rijnbeek, P.R. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *J. Biomed. Inform.* **2021**, *113*, 103655. [CrossRef] [PubMed]
25. Brożek, B.; Furman, M.; Jakubiec, M.; Kucharzyk, B. The black box problem revisited. Real and imaginary challenges for automated legal decision making. *Artif. Intell. Law* **2023**, *32*, 427–440. [CrossRef]
26. Li, D.; Liu, Y.; Huang, J.; Wang, Z. A Trustworthy View on Explainable Artificial Intelligence Method Evaluation. *Computer* **2023**, *56*, 50–60. [CrossRef]
27. Nauta, M.; Trienes, J.; Pathak, S.; Nguyen, E.; Peters, M.; Schmitt, Y.; Schlötterer, J.; van Keulen, M.; Seifert, C. From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI. *ACM Comput. Surv.* **2023**, *55*, 295. [CrossRef]
28. Lopes, P.; Silva, E.; Braga, C.; Oliveira, T.; Rosado, L. XAI Systems Evaluation: A Review of Human and Computer-Centred Methods. *Appl. Sci.* **2022**, *12*, 9423. [CrossRef]
29. Baptista, M.L.; Goebel, K.; Henriques, E.M. Relation between prognostics predictor evaluation metrics and local interpretability SHAP values. *Artif. Intell.* **2022**, *306*, 103667. [CrossRef]
30. Fouladgar, N.; Alirezaie, M.; Framling, K. Metrics and Evaluations of Time Series Explanations: An Application in Affect Computing. *IEEE Access* **2022**, *10*, 23995–24009. [CrossRef]
31. Oblizanov, A.; Shevskaya, N.; Kazak, A.; Rudenko, M.; Dorofeeva, A. Evaluation Metrics Research for Explainable Artificial Intelligence Global Methods Using Synthetic Data. *Appl. Syst. Innov.* **2023**, *6*, 26. [CrossRef]
32. Speith, T. A Review of Taxonomies of Explainable Artificial Intelligence (XAI) Methods. In Proceedings of the FAccT ‘22: 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, 21–24 June 2022; pp. 2239–2250.
33. Kurdziolek, M. Explaining the Unexplainable: Explainable AI (XAI) for UX. *User Experience Magazine*. 2022. Available online: <https://uxpamagazine.org/explaining-the-unexplainable-explainable-ai-xai-for-ux/> (accessed on 20 August 2023).

34. Kim, B.; Wattenberg, M.; Gilmer, J.; Cai, C.; Wexler, J.; Viegas, F.; Sayres, R. Interpretability beyond feature attribution: Quantitative Testing with Concept Activation Vectors (TCAV). In Proceedings of the 35th International Conference on Machine Learning, ICML, Stockholm, Sweden, 10–15 July 2018; Volume 6, pp. 4186–4195. Available online: <https://proceedings.mlr.press/v80/kim18d/kim18d.pdf> (accessed on 30 July 2024).
35. Kenny, E.M.; Keane, M.T. Explaining Deep Learning using examples: Optimal feature weighting methods for twin systems using post-hoc, explanation-by-example in XAI. *Knowl. Based Syst.* **2021**, *233*, 107530. [[CrossRef](#)]
36. Alfeo, A.L.; Zippo, A.G.; Catrambone, V.; Cimino, M.G.; Toschi, N.; Valenza, G. From local counterfactuals to global feature importance: Efficient, robust, and model-agnostic explanations for brain connectivity networks. *Comput. Methods Programs Biomed.* **2023**, *236*, 107550. [[CrossRef](#)] [[PubMed](#)]
37. An, J.; Zhang, Y.; Joe, I. Specific-Input LIME Explanations for Tabular Data Based on Deep Learning Models. *Appl. Sci.* **2023**, *13*, 8782. [[CrossRef](#)]
38. Bharati, S.; Mondal, M.R.H.; Podder, P. A Review on Explainable Artificial Intelligence for Healthcare: Why, How, and When? *IEEE Trans. Artif. Intell.* **2023**, *5*, 1429–1442. [[CrossRef](#)]
39. Chaddad, A.; Peng, J.; Xu, J.; Bouridane, A. Survey of Explainable AI Techniques in Healthcare. *Sensors* **2023**, *23*, 634. [[CrossRef](#)]
40. Chamola, V.; Hassija, V.; Sulthana, A.R.; Ghosh, D.; Dhingra, D.; Sikdar, B. A Review of Trustworthy and Explainable Artificial Intelligence (XAI). *IEEE Access* **2023**, *11*, 78994–79015. [[CrossRef](#)]
41. Chen, X.-Q.; Ma, C.-Q.; Ren, Y.-S.; Lei, Y.-T.; Huynh, N.Q.A.; Narayan, S. Explainable artificial intelligence in finance: A bibliometric review. *Financ. Res. Lett.* **2023**, *56*, 104145. [[CrossRef](#)]
42. Di Martino, F.; Delmastro, F. Explainable AI for clinical and remote health applications: A survey on tabular and time series data. *Artif. Intell. Rev.* **2022**, *56*, 5261–5315. [[CrossRef](#)]
43. Kök, I.; Okay, F.Y.; Muyanli, O.; Özdemir, S. Explainable Artificial Intelligence (XAI) for Internet of Things: A Survey. *IEEE Internet Things J.* **2023**, *10*, 14764–14779. [[CrossRef](#)]
44. Haque, A.B.; Islam, A.N.; Mikalef, P. Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technol. Forecast. Soc. Chang.* **2023**, *186*, 122120. [[CrossRef](#)]
45. Sahoh, B.; Choksuriwong, A. The role of explainable Artificial Intelligence in high-stakes decision-making systems: A systematic review. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 7827–7843. [[CrossRef](#)]
46. Saranya, A.; Subhashini, R. A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. *Decis. Anal. J.* **2023**, *7*, 100230. [[CrossRef](#)]
47. Sosa-Espadas, C.E.; Orozco-Del-Castillo, M.G.; Cuevas-Cuevas, N.; Recio-Garcia, J.A. IREX: Iterative Refinement and Explanation of classification models for tabular datasets. *SoftwareX* **2023**, *23*, 101420. [[CrossRef](#)]
48. Meding, K.; Hagendorff, T. Fairness Hacking: The Malicious Practice of Shrouding Unfairness in Algorithms. *Philos. Technol.* **2024**, *37*, 4. [[CrossRef](#)]
49. Batko, K.; Ślęzak, A. The use of Big Data Analytics in healthcare. *J. Big Data* **2022**, *9*, 3. [[CrossRef](#)]
50. Borisov, V.; Leemann, T.; Seßler, K.; Haug, J.; Pawelczyk, M.; Kasneci, G. Deep Neural Networks and Tabular Data: A Survey. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 7499–7519. [[CrossRef](#)]
51. Mbanaso, M.U.; Abrahams, L.; Okafor, K.C. Data Collection, Presentation and Analysis. In *Research Techniques for Computer Science, Information Systems and Cybersecurity*; Springer: Cham, Switzerland, 2023; pp. 115–138. [[CrossRef](#)]
52. Tjoa, E.; Guan, C. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Trans. Neural Networks Learn. Syst.* **2021**, *32*, 4793–4813. [[CrossRef](#)]
53. Gajcin, J.; Dusparic, I. Redefining Counterfactual Explanations for Reinforcement Learning: Overview, Challenges and Opportunities. *ACM Comput. Surv.* **2024**, *56*, 219. [[CrossRef](#)]
54. Hassija, V.; Chamola, V.; Mahapatra, A.; Singal, A.; Goel, D.; Huang, K.; Scardapane, S.; Spinelli, I.; Mahmud, M.; Hussain, A. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cogn. Comput.* **2023**, *16*, 45–74. [[CrossRef](#)]
55. Lötsch, J.; Kringel, D.; Ultsch, A. Explainable Artificial Intelligence (XAI) in Biomedicine: Making AI Decisions Trustworthy for Physicians and Patients. *BioMedInformatics* **2021**, *2*, 1–17. [[CrossRef](#)]
56. Hossain, I.; Zamzmi, G.; Mouton, P.R.; Salekin, S.; Sun, Y.; Goldgof, D. Explainable AI for Medical Data: Current Methods, Limitations, and Future Directions. *ACM Comput. Surv.* **2023**. [[CrossRef](#)]
57. Rudin, C.; Chen, C.; Chen, Z.; Huang, H.; Semenova, L.; Zhong, C. Interpretable machine learning: Fundamental principles and 10 grand challenges. *Stat. Surv.* **2022**, *16*, 1–85. [[CrossRef](#)]
58. Zhong, X.; Gallagher, B.; Liu, S.; Kailkhura, B.; Hiszpanski, A.; Han, T.Y.-J. Explainable machine learning in materials science. *NPJ Comput. Mater.* **2022**, *8*, 204. [[CrossRef](#)]
59. Ekanayake, I.; Meddage, D.; Rathnayake, U. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). *Case Stud. Constr. Mater.* **2022**, *16*, e01059. [[CrossRef](#)]
60. Černevičienė, J.; Kabašinskas, A. Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artif. Intell. Rev.* **2024**, *57*, 216. [[CrossRef](#)]
61. Weber, P.; Carl, K.V.; Hinz, O. Applications of Explainable Artificial Intelligence in Finance—A systematic review of Finance, Information Systems, and Computer Science literature. *Manag. Rev. Q.* **2024**, *74*, 867–907. [[CrossRef](#)]

62. Leijnen, S.; Kuiper, O.; van der Berg, M. Impact Your Future Xai in the Financial Sector a Conceptual Framework for Explainable Ai (Xai). Hogeschool Utrecht, Lectoraat Artificial Intelligence, Whitepaper, Version 1, 1–24. 2020. Available online: <https://www.hu.nl/onderzoek/projecten/uitlegbare-ai-in-de-financiele-sector> (accessed on 2 August 2024).
63. Dastile, X.; Celik, T. Counterfactual Explanations with Multiple Properties in Credit Scoring. *IEEE Access* **2024**, *12*, 110713–110728. [[CrossRef](#)]
64. Martins, T.; de Almeida, A.M.; Cardoso, E.; Nunes, L. Explainable Artificial Intelligence (XAI): A Systematic Literature Review on Taxonomies and Applications in Finance. *IEEE Access* **2023**, *12*, 618–629. [[CrossRef](#)]
65. Kalra, A.; Mittal, R. Explainable AI for Improved Financial Decision Support in Trading. In Proceedings of the 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 14–15 March 2024; pp. 1–6.
66. 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](#)]
67. Li, Y.; Song, X.; Wei, T.; Zhu, B. Counterfactual learning in customer churn prediction under class imbalance. In Proceedings of the 2023 6th International Conference on Big Data Technologies (ICBDT '23), Qingdao, China, 22–24 September 2023; pp. 96–102.
68. Zhang, L.; Zhu, Y.; Ni, Q.; Zheng, X.; Gao, Z.; Zhao, Q. Local/Global explainability empowered expert-involved frameworks for essential tremor action recognition. *Biomed. Signal Process. Control* **2024**, *95*, 106457. [[CrossRef](#)]
69. Sadeghi, Z.; Alizadehsani, R.; Cifci, M.A.; Kausar, S.; Rehman, R.; Mahanta, P.; Bora, P.K.; Almasri, A.; Alkhalwaldeh, R.S.; Hussain, S.; et al. A review of Explainable Artificial Intelligence in healthcare. *Comput. Electr. Eng.* **2024**, *118*, 109370. [[CrossRef](#)]
70. Alizadehsani, R.; Oyelere, S.S.; Hussain, S.; Jagatheesaperumal, S.K.; Calixto, R.R.; Rahouti, M.; Roshanzamir, M.; De Albuquerque, V.H.C. Explainable Artificial Intelligence for Drug Discovery and Development: A Comprehensive Survey. *IEEE Access* **2024**, *12*, 35796–35812. [[CrossRef](#)]
71. Murindanyi, S.; Mugalu, B.W.; Nakatumba-Nabende, J.; Marvin, G. Interpretable Machine Learning for Predicting Customer Churn in Retail Banking. In Proceedings of the 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 11–13 April 2023; pp. 967–974.
72. Mill, E.; Garn, W.; Ryman-Tubb, N.; Turner, C. Opportunities in Real Time Fraud Detection: An Explainable Artificial Intelligence (XAI) Research Agenda. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*, 1172–1186. [[CrossRef](#)]
73. Dutta, J.; Puthal, D.; Yeun, C.Y. Next Generation Healthcare with Explainable AI: IoMT-Edge-Cloud Based Advanced eHealth. In Proceedings of the IEEE Global Communications Conference, GLOBECOM, Kuala Lumpur, Malaysia, 4–8 December 2023; pp. 7327–7332.
74. Njoku, J.N.; Nwakanma, C.I.; Lee, J.-M.; Kim, D.-S. Evaluating regression techniques for service advisor performance analysis in automotive dealerships. *J. Retail. Consum. Serv.* **2024**, *80*, 103933. [[CrossRef](#)]
75. Agostinho, C.; Dikopoulou, Z.; Lavasa, E.; Perakis, K.; Pitsios, S.; Branco, R.; Reji, S.; Hetterich, J.; Biliri, E.; Lampathaki, F.; et al. Explainability as the key ingredient for AI adoption in Industry 5.0 settings. *Front. Artif. Intell.* **2023**, *6*, 1264372. [[CrossRef](#)]
76. Finzel, B.; Tafler, D.E.; Thaler, A.M.; Schmid, U. Multimodal Explanations for User-centric Medical Decision Support Systems. *CEUR Workshop Proc.* **2021**, *3068*, 1–6.
77. Brochado, F.; Rocha, E.M.; Addo, E.; Silva, S. Performance Evaluation and Explainability of Last-Mile Delivery. *Procedia Comput. Sci.* **2024**, *232*, 2478–2487. [[CrossRef](#)]
78. Kostopoulos, G.; Davrazos, G.; Kotsiantis, S. Explainable Artificial Intelligence-Based Decision Support Systems: A Recent Review. *Electronics* **2024**, *13*, 2842. [[CrossRef](#)]
79. Nyrupe, R.; Robinson, D. Explanatory pragmatism: A context-sensitive framework for explainable medical AI. *Ethics Inf. Technol.* **2022**, *24*, 13. [[CrossRef](#)]
80. Talaat, F.M.; Aljadani, A.; Alharthi, B.; Farsi, M.A.; Badawy, M.; Elhosseini, M. A Mathematical Model for Customer Segmentation Leveraging Deep Learning, Explainable AI, and RFM Analysis in Targeted Marketing. *Mathematics* **2023**, *11*, 3930. [[CrossRef](#)]
81. Kulkarni, S.; Rodd, S.F. Context Aware Recommendation Systems: A review of the state of the art techniques. *Comput. Sci. Rev.* **2020**, *37*, 100255. [[CrossRef](#)]
82. Sarker, A.A.; Shanmugam, B.; Azam, S.; Thennadil, S. Enhancing smart grid load forecasting: An attention-based deep learning model integrated with federated learning and XAI for security and interpretability. *Intell. Syst. Appl.* **2024**, *23*, 200422. [[CrossRef](#)]
83. Nnadi, L.C.; Watanobe, Y.; Rahman, M.; John-Otumu, A.M. Prediction of Students' Adaptability Using Explainable AI in Educational Machine Learning Models. *Appl. Sci.* **2024**, *14*, 5141. [[CrossRef](#)]
84. Vellido, A.; Martín-Guerrero, J.D.; Lisboa, P.J.G. Making machine learning models interpretable. In Proceedings of the 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 25–27 April 2012; pp. 163–172. Available online: <https://www.esann.org/sites/default/files/proceedings/legacy/es2012-7.pdf> (accessed on 16 August 2024).
85. Alkhatib, A.; Ennadir, S.; Boström, H.; Vazirgiannis, M. Interpretable Graph Neural Networks for Tabular Data. In Proceedings of the ICLR 2024 Data-Centric Machine Learning Research (DMLR) Workshop, Vienna, Austria, 26–27 July 2024; pp. 1–35. Available online: <https://openreview.net/pdf/60ce21fd5bcf7b6442b1c9138d40e45251d03791.pdf> (accessed on 23 August 2024).
86. Saeed, W.; Omlin, C. Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowl. Based Syst.* **2023**, *263*, 110273. [[CrossRef](#)]

87. de Oliveira, R.M.B.; Martens, D. A Framework and Benchmarking Study for Counterfactual Generating Methods on Tabular Data. *Appl. Sci.* **2021**, *11*, 7274. [[CrossRef](#)]
88. Bienefeld, N.; Boss, J.M.; Lüthy, R.; Brodbeck, D.; Azzati, J.; Blaser, M.; Willms, J.; Keller, E. Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. *NPJ Digit. Med.* **2023**, *6*, 94. [[CrossRef](#)]
89. Molnar, C.; Casalicchio, G.; Bischl, B. Interpretable Machine Learning—A Brief History, State-of-the-Art and Challenges. In *ECML PKDD 2020 Workshops, Proceedings of the ECML PKDD 2020, Ghent, Belgium, 14–18 September 2020*; Koprinska, I., Kamp, M., Appice, A., Loglisci, C., Antonie, L., Zimmermann, A., Guidotti, R., Özgöbek, Ö., Ribeiro, R.P., Gavaldà, R., et al., Eds.; Springer: Cham, Switzerland, 2021; Volume 1323, pp. 417–431. [[CrossRef](#)]
90. Pawlicki, M.; Pawlicka, A.; Kozik, R.; Choraś, M. Advanced insights through systematic analysis: Mapping future research directions and opportunities for xAI in deep learning and artificial intelligence used in cybersecurity. *Neurocomputing* **2024**, *590*, 127759. [[CrossRef](#)]
91. Hartog, P.B.R.; Krüger, F.; Genheden, S.; Tetko, I.V. Using test-time augmentation to investigate explainable AI: Inconsistencies between method, model and human intuition. *J. Cheminform.* **2024**, *16*, 39. [[CrossRef](#)] [[PubMed](#)]
92. Srinivasu, P.N.; Sandhya, N.; Jhaveri, R.H.; Raut, R. From Blackbox to Explainable AI in Healthcare: Existing Tools and Case Studies. *Mob. Inf. Syst.* **2022**, *2022*, 167821. [[CrossRef](#)]
93. Rong, Y.; Leemann, T.; Nguyen, T.-T.; Fiedler, L.; Qian, P.; Unhelkar, V.; Seidel, T.; Kasneci, G.; Kasneci, E. Towards Human-Centered Explainable AI: A Survey of User Studies for Model Explanations. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 2104–2122. [[CrossRef](#)]
94. Baniecki, H.; Biecek, P. Adversarial attacks and defenses in explainable artificial intelligence: A survey. *Inf. Fusion* **2024**, *107*, 102303. [[CrossRef](#)]
95. Panigutti, C.; Hamon, R.; Hupont, I.; Llorca, D.F.; Yela, D.F.; Junklewitz, H.; Scalzo, S.; Mazzini, G.; Sanchez, I.; Garrido, J.S.; et al. The role of explainable AI in the context of the AI Act. In *Proceedings of the FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, Chicago, IL, USA, 12–15 June 2023; pp. 1139–1150.
96. Madiaga, T.; Chahri, S. EU Legislation in Progress: Artificial Intelligence Act, 1–12. 2024. Available online: [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS\\_BRI\(2021\)698792\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf) (accessed on 16 August 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.