Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2023, 10(3):82-92

Research Article ISSN: 2394 - 658X

Explainable AI and Interpretable Machine Learning in Financial Industry Banking

Praneeth Reddy Amudala Puchakayala[1], Saurabh Kumar[2], Shafeeq Ur Rahaman[3]

[1]Data scientist, Regions Bank, AL, USA
[2]Kraft Heinz Foods, Senior Manager, Data Science
saurabh.hoa@gmail.com
[3]Monks, CA, USA

ABSTRACT

Because deep learning models are particularly good at processing large amounts of data and picking up intricate patterns, they have become widely used in a variety of industries, which is the result of their success. Their usage in crucial industries like finance and healthcare, where transparent decision-making is crucial, is fraught with serious risks, nevertheless, because of their explainability issues. AI gives machines the ability to learn from human experience, adapt to new inputs, and carry out jobs that resemble those of humans. Process automation, cognitive task augmentation, and intelligent process/data analytics are just a few of the ways artificial intelligence is quickly changing how businesses run. However, the key problem for human users would be to comprehend and properly trust the outcomes of AI algorithms and procedures.

In this research, we present a comparative overview of approaches designed to enhance the explainability of deep learning models in the financial domain. We categorise the collection of explainable AI methods based on their respective qualities, and we discuss the issues and obstacles of implementing explainable AI methods, as well as future paths that we believe acceptable and relevant.

Keywords: Deep Learning; Artificial Intelligence; Decision Making; Financial; Cognitive; Intelligent.

INTRODUCTION

A field that is always changing and has a long history dating back to the birth of human civilisation is finance. Efficient resource allocation is a primary responsibility of finance, exemplified by the management of money flows across diverse entities with varying requirements. These entities can be classified as individuals, corporations, or countries, resulting in the common categories of personal, corporate, and government finance. The sector can be traced back to 5000 years ago, when agrarian cultures had been created and developed for several thousand years. In fact, one of the earliest instances of banking, a key organisation in the financial sector, dates back to the Babylonian empire. Since then, developments in society and technology have forced the industry to adapt in a number of ways. These developments have been especially noticeable in the last 20 years because of how quickly technology is developing, particularly in the context of artificial intelligence. From digital transactions to investment management, risk management, algorithmic trading, and beyond, the latter has begun to disseminate across various financial sectors [1].

FinTech (Financial Technology) is the term for the new field that has grown significantly over the last 20 years [89] through the automation and enhancement of financial operations using innovative AI and non-AI technology. In this analysis, we look at AI-based technologies and machine learning for financial applications. Financial researchers and practitioners have used supervised, unsupervised, and semi-supervised machine learning methods, as well as reinforcement learning, to solve a wide range of problems. Credit evaluation, fraud detection, algorithmic trading, and wealth management are just a few examples. In supervised-based machine learning approaches, neural networks are commonly used to detect complicated correlations hidden in labelled data. Usually, domain experts give the labels. A domain expert could identify periods of favourable and negative returns when developing a stock-picking system, for example. The computer is then tasked with constructing the relationship between positive and negative returns of a given stock (or many stocks) and (potentially) high-dimensional data, and generalising to unseen data to, for example, anticipate

the behaviour of the stock in the future. In unsupervised machine learning approaches, the goal is to discover data with similar properties that may be clustered together without the need for domain-expert labelling [2].

For instance, one could consider employing similarity criteria like value, profitability, and risk to group all companies with comparable traits into clusters. Semi-supervised learning is a compromise between supervised and unsupervised learning in which just a portion of the data is labelled. Lastly, reinforcement learning seeks to maximise the cumulative reward that practitioners specify by executing a series of behaviours. In finance, reinforcement learning is applied to tasks like portfolio construction. Reinforcement learning is inextricably linked to Markov decision processes and differs significantly from supervised and unsupervised learning [3]. When it comes to complexity, controlled, unsupervised, and reinforcement learning methods are very different from one another. While certain methods—also known as "black-box" methods—are thought to be uninterpretable, others are thought to be easier to understand and, therefore, easier for practitioners to interpret. To this purpose, neural networks and deep learning algorithms, which underpin the bulk (if not the entirety) of modern machine learning approaches for financial applications, are regarded as black-box techniques (i.e., the reason for a particular prediction is not readily available).

This is a serious problem, particularly in industries that are high-risk and heavily regulated, like finance and healthcare, where a poor choice could result in the catastrophic loss of wealth or human life. Understanding the rationale (i.e., the facts and patterns) the computer utilised to reach a certain choice was therefore seen to be crucial. This includes the wide area of transparency in AI. The latter consists of three pillars:

(i) AI awareness,

(ii) AI model explainability, and

(iii) AI outcome explainability.

The first is to determine if a certain product uses artificial intelligence. The second must give a thorough description of the AI model, outlining all of its inputs and outputs. An in-depth explanation of how the inputs affected the AI model's results must be given by the third. We find a wide range of post-hoc interpretability techniques in this final category. We will assume that you are aware of AI, which means you know that it is used in a certain business process. This review will focus on AI explainability, which is also known as eXplainable AI or just XAI. Another distinction that is frequently emphasised is the interpretability and explainability of an AI model. Despite their widespread interchangeability, these two terms have a few significant differences. Interpretability describes a model's operation and rationale. Explainability is the capacity to interpret the findings in a way that makes sense to humans.

Many other techniques in finance are regarded as white-box techniques, whereas deep learning techniques are regarded as black-box techniques. One of the most hotly contested topics in the world of financial artificial intelligence is the trade-off between interpretability and complexity. On the one hand, white-box approaches are easily interpretable, yet they lack the ability to grasp complex linkages, commonly failing to satisfy performance expectations [4]. Conversely, black-box techniques don't allow for interpretation while typically (but not always) achieving the required results. As a result, it should come as no surprise that major attempts have been made in recent years to make black-box technologies more interpretable, with deep learning serving as a prime example.

In this study, we present a comprehensive overview of XAI techniques for the financial domain, which we refer to as FinXAI. Despite it being true that there have been several surveys on XAI techniques [6], these studies are not finance-specific and instead focus on generic XAI. Therefore, we explore explainability strategies that are specific to financial use cases.

To prepare this review, we looked at 69 publications, focussing primarily, but not completely, on the third pillar, namely the explainability of the inputs' contributions to the AI model's results. So, we thought about ways to improve the interpretability of black-box deep learning models using post-hoc methods, as well as models that are transparent from the start and don't need any further post-hoc interpretability. Although there aren't many collected papers in the subject of XAI, it's vital to remember that our primary goal is to concentrate on XAI methods that are relevant to the financial sector. This targeted strategy will provide significant insights for academics in related sectors, ultimately promoting innovation and advancement in the financial industry. With the growing demand for transparency and accountability of deep learning, the XAI community has published more works, although we focus here on financial use cases [7].

Because FinXAI is a very tiny subset of XAI, we aim to stay current with current methodologies by integrating existing works in a comprehensive manner. The publications were retrieved from Google Scholar and Scopus using a set of keywords relevant to works that have employed explainable AI techniques in financial use cases, which included "XAI, explainable AI, finance, financial sector, financial field, explainable ML".

We try to gather a wide range of papers that adequately address each topic, which are then compiled into tables 1, 2, and 3. We also discovered that the bulk of explanation types were limited to fact-based explanations, therefore we specifically searched for strategies explaining in the form of counterfactuals. Since the recipient usually prefers to understand why a particular prediction was made rather than the opposite, counterfactual explanations are thought to be a desirable kind of explanation. The primary outcomes of our study as a whole are:

• For academics who are interested in giving transparency in their solutions a higher priority, we offer a comprehensive study on consolidating XAI methodologies in the sector of finance, which we refer to as FinXAI.

• The FinXAI process is set up as a series of decision-making steps.

It is important that the XAI techniques are right for the audience. This framework aims to generate explanations that are focused on the needs of the audience and goal oriented. Examine existing FinXAI methodologies, assess how technically

they support ethical objectives, and enumerate significant implementation issues along with critical areas that need to be improved going forward.

AI EXPLAINABLE WITH BLACK BOX PROBLEM (XAI)

To comprehensively comprehend the relevance and advocacy of XAI, this study initially efforts to integrate the current understanding of AI applications in banking and financial services, eventually resulting in the "Black Box" challenge associated with AI. The proposed solution is XAI (refer to Figure 3). This section will also include a compilation of notable research articles about the concept and future prospective avenues for each level in Figure 1.

A. Banking and Finance Applications based on AI:

• Applying Forecasting and prediction model
• Managing the investment
• Managing the risk and regulators
• Banking

B. Black Box Problem

• Explainable AI design
• Inspecting the model
• Discuss the ML model

C. XAI

• Bank Bias Prevention
• Credit card risk payment fraud and prediction
• Monitoring and accounting

D. Trusted stakeholder

• Responsible
• Trustworthy
• Interpretability

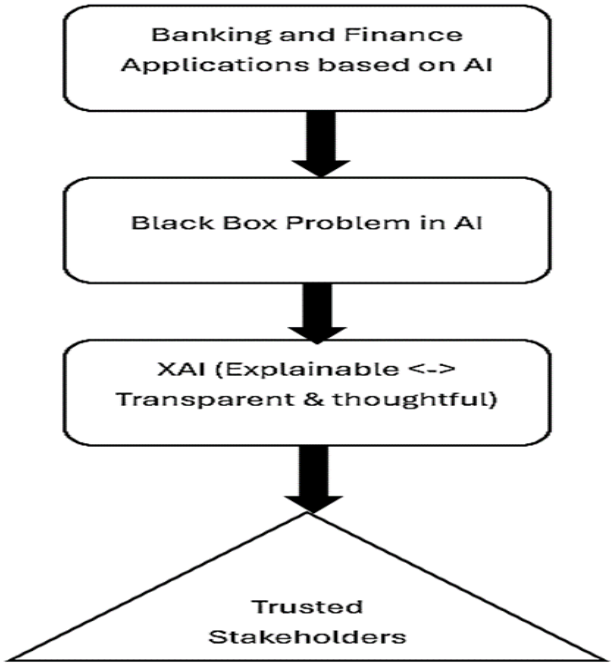

Figure 1. AI key Flow with XAI

AI models are trained and used to real-time data analysis in insurance, credit card, bank loan, and other retail banking products [8]. This process is done to detect fraud. [9] examined the machine learning model for highly accurate accounting fraud detection predictions on A-listed Chinese companies using financial ratios and unprocessed financial data. Although statistical and AI classifiers have been the subject of numerous studies, their effectiveness in credit risk forecasting needs to be evaluated in comparison to alternative techniques. In their study, integrated financial ratios (such as "solvency," "profitability," "cash flow," "capital structure," and "turnover") with corporate governance (CG) indicators (such as board structure, ownership structure, cash flow rights, and retention of key individuals) to forecast bankruptcy as in Figure 2. AI analyses and forecasts fraud in multiple dimensions of the financial system using previously existing data and customer behaviour, whereas XAI supplements by ensuring they are non-discriminatory and resilient, making them more useful to the industry [10].

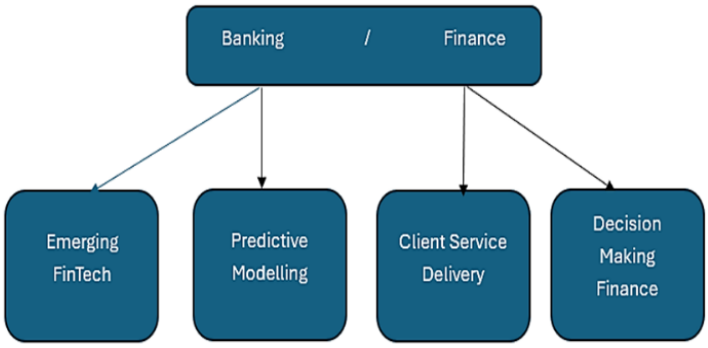

Figure 2. *AI Application in Finance and banking*

The current emphasis in corporate finance and risk management is on the forecast of bankruptcy and the monitoring of financial risks [11] Contemporary machine learning methodology is used in several crucial aspects of financial institution operations by creating and improving prediction models. Review articles have examined several machine learning approaches used in trading, risk and data management, forecasting, trend analysis and financial distress mitigation for medium and small firms. The "Ant Colony Optimisation (ACO)"-based "financial crisis prediction (FCP)" model was developed to anticipate financial crises and ML was used to predict firm financial solvency (Abdullah,2021). Furthermore, to demonstrate its superiority, it was compared to known methods for feature selection and data categorisation. Credit risk management, as measured by data mining, is a major challenge for the financial industry, resulting in considerable financial losses in both the banking and non-banking sectors. In order to expedite financial option pricing, particularly for high-dimensional models, [11] proposed the use of an artificial neural network (ANN).

For the purpose of forging monetary policy decisions, central banks depend on economic forecasting. By analysing vast datasets and identifying patterns in data that conventional data analysis methods would find challenging or impossible to identify, AI and ML can enhance the precision of economic forecasts [12]. Central banks manage financial stability by regulating financial institutions, monitoring systemic risks, and ensuring financial system stability. Artificial intelligence (AI) and machine learning (ML) can help central banks in real-time monitoring and identification of potential risks to financial stability, as well as the detection and investigation of fraudulent financial activities. Machine learning algorithms have been developed to identify anomalies and apparent risks in financial data, including market prices, trade volumes, and news sentiments. Money laundering, terrorist financing, and insider trading can all be discovered with machine learning algorithms that analyse transaction data. According to [13], there is a growing trend of technology breakthroughs in all industries, including finance.

Fintech start-ups are actively investigating and capitalising on the potential to contribute to Digital Financial Institutions (DFI), often referred to as "Industry 4.0" in the field of finance (Mhlanga, 2020). AI apps are currently used by many urban services. Urban artificial intelligences operate restaurants and enterprises characteristic of urban environments, maintain urban infrastructure, and manage traffic, air quality monitoring, garbage collection, and energy management [13]. The Peer-2-Peer (P2P) platform was explored using a comparison of ML algorithms (LightGBM vs XGboost) to determine prediction accuracy for loan default risk on these platforms. [14] proposed four kinds of variables that influence loan repayment on P2P platforms: "loan details", "financial status", "credit status", and "personal information". P2P lending platforms have also been researched to propose a cost-sensitive boosted tree loan appraisal model that employs cost-sensitive learning and extreme gradient boosting (XGBoost) to better identify potential defaulters. The implementation of AI is not without its obstacles, including customer resistance, high costs, unskilled personnel in AI, non-availability or poor quality of data, alignment with regulations, and security breaches [15].

Artificial intelligence (AI) appraises real-time service incidents in customer service by making use of data from digital and/or physical sources to provide targeted recommendations, alternatives, and responses to even the most complex client enquiries. It also provides practical tips for banks that are utilising AI customer care to stay connected with visitors. [16] investigated client attitudes, mass media, and interpersonal subjective norms based on technology familiarity and demographic variables to assess the financial sector's adoption of AI technology (robo-advisory). For analysing consumer comfort with AI in mobile banking for self-service delivery, perception of banking service; AI service, safety, and security for digital naives; and AI feature, namely perceived intelligence and anthropomorphism, were assessed. AI in mobile banking services was classed as utilitarian (transaction-oriented) or hedonic (relationship-oriented) in terms of value co-creation, consumer and company performance.

The use of chatbots in the financial industry is accompanied by a growing body of study on their interaction with people, focussing on principles of equality and data safety. The early stages of machine learning necessitate human intervention, which gives rise to prejudice against specific racial and gender groups and lowers public confidence in financial services [17]. This is known as the AI "black box" problem, which is the possibility of an AI delivering unintended consequences that go unnoticed or unanticipated because people are unable to understand its internal workings or its autonomous operation without human supervision or involvement.

AI - BLACK BOX PROBLEM

When individuals demonstrate knowledge of input and output but lack comprehension of the underlying process, it is commonly referred to as a black box in the context of artificial intelligence and its applications [18]. According to a survey on AI adoption in credit risk, thirty-eight percent of financial executives are concerned about explainability, with model governance being the most common worry. AI applications in the financial and banking industries offer a range of advantages, including enhanced efficiency and improved client experience. However, they also come with notable challenges and potential concerns. When AI systems are trained on data that does not reflect the entire population, they can become biassed. This can lead to biassed lending or insurance denials [19]. Understanding the decision-making process of AI models can be quite challenging due to their complex nature. When making important financial decisions, the absence of transparency can be concerning.

Cyberattacks targeting AI systems have the potential to compromise sensitive financial data, leading to financial losses and reputational harm. Legislation and compliance guidelines control the use of AI in finance and banking. AI systems may require auditing and certification from regulatory organisations, which can be a lengthy and costly process. As per the Organisation for Economic Co-operation and Development (OECD), incumbent banks encounter conflicting objectives that pose a challenge. Banks must, on the one hand, attain the quickness, dexterity, and adaptability inherent to fintechs. However, they still have to maintain the size, security requirements, and legal obligations of a conventional financial services company. One of the main challenges faced by banks is the absence of a well-defined AI strategy.

The list of prominent areas based on AI – Black Box Problem

- Credit Scoring
- Bankruptcy Prediction
- Fraud Prediction
- Risk Assessment based on investment
- Financial risk analysis
- Cyber Security

EXPLAINABLE AI MODELS

One of the USPs of AI models is model accuracy, which is traded off with transparency by a number of banking and financial institutions, prioritising XAI. The block-box nature of explainability (for customers or users), interpretability (for designers/programmers/users), and provenance (for regulatory purposes) is addressed by XAI in current AI models [20]. According to IBM, 68 percent of customers would demand greater explanations for AI judgements by 2024. XAI is a collection of tools, approaches, concepts, and strategies that help human stakeholders understand, evaluate, interpret, and enhance ML or AI model predictions and judgements. Defence Advanced Research Projects Agency (DARPA) claims that the present AI models are generating important Why, When, and How questions. Regarding the output of these models, XAI provides the solution and provides a summary of the distinction between the current artificial intelligence (AI) process and explainable artificial intelligence (XAI).

AI systems that are designed, built, and implemented with an emphasis on transparency and interpretation are commonly referred to as transparent AI. In other words, transparent AI systems are characterised by their simplicity and ease of understanding and explanation. AI needs transparency because it helps people trust AI systems and makes sure they are used in an accurate and accountable way. Users may better understand how the system generates decisions, as well as detect and rectify any biases or flaws in the system's logic, thanks to transparent AI. XAI approaches, which allow humans to understand the thinking behind an AI system's judgements, are one approach to achieving AI transparency. The level of transparency that different stakeholders experience can be categorised into algorithmic, interaction, and social [21]. The former two categories pertain to developers and regulators, while the latter two include designers and users. The creation of AI models and systems that are transparent, interpretable, and capable of offering acceptable explanations for their outputs and decision-making processes is known as XAI. XAI also fosters a positive environment for technologists by enhancing their ability to monitor, maintain as in figure 3,

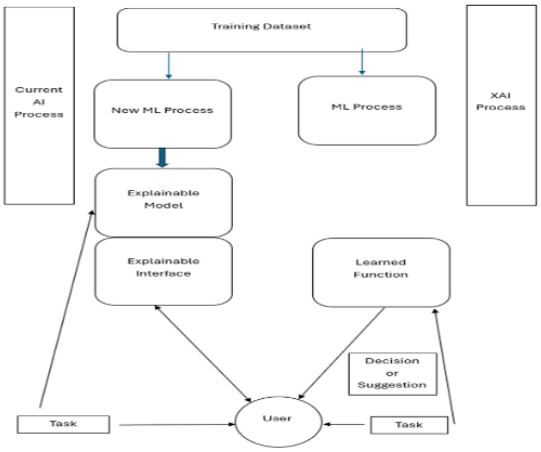

Figure. *3 XAI Framework*

and enhance AI models; business professionals by enabling trust in the output, providing suggestions and interventions to ensure that AI models are in alignment with the organization's objectives.

FORWARDING EXPLAINABLE AI TOWARDS FINANCIAL BANKING

In the financial sector, lending is one of the broad components that directly influence the economy. According to the Australian Bureau of Statistics' August 202146 figures, many Australians have billions of dollars in credit [22]. Any innovation that can help a banking sector better its profitability through the advancements it has or is seeking would be highly valued. Because of this, financial organizations—such as banks and private associations—are looking for methods to enhance, develop, and consider integrating artificial intelligence (AI) into decision-making procedures. This approach of applying AI is only the first step; further work needs to be done to accomplish the objectives. Lending money, a substantial source of income in financial institutions, has become an information issue as user data has grown, making it an ideal business application for machine learning.

The financial stability of the individual or firm determines a portion of the loan's value. The more information you have about the borrower (and how other people have previously returned obligations), the more you can assess their financial stability. Thus, assessments of the insurance's value (car, house, business, workmanship, etc.), the likelihood of further growth, and projections based on significant financial development are linked to the value of an advance. AI promises that, in theory, it is able to decipher these sources of information. It can assist with the framework needed to make a wise decision. Using AI algorithms, the business may calculate the borrower's risk in the risk matrix25 and automate scoring computations. One obstacle preventing AI from being used in the finance industry is its enigmatic character. Due to investor worries and AI's improvements and success, we're conducting scientific studies on XAI's current use in finance to assist AI's growth and convince investors to invest in intelligent decision systems.

A. Problem Consideration

AI gives machines the ability to learn from human experience, adapt to new inputs, and carry out jobs that resemble those of humans. From process automation to cognitive task augmentation and intelligent process/data analytics, artificial intelligence is advancing quickly and changing how businesses run 3, 10, 60. However, the key problem for human users would be to comprehend and properly trust the outcomes of AI algorithms and procedures. Using a compelling example from the banking industry, these methods have not yet been adopted by commercial systems because the ML system's incapacity to explain the outcome it generates. As black boxes by nature, ML systems are unable to provide the user with an explanation of what, how, or why [22]. A cost analysis to enhance the system and raise the risk is required, and this incapacity undermines the confidence of investors and stakeholders.

To improve financial systems with expanding data, we need intelligent tools that allow decision-making and compliance officers to interpret banking data. However, with the rising data domain, specialists are unable to cover all elements of data. Because of the rise of data, these specialists have constraints. On the other hand, to meeting the demands of financial corporations, AI systems must be able to justify their conclusions. The need of the hour is to have a system in place that can bridge this gap by giving a solution to the limitations of AI-based systems and assisting domain experts in understanding the workings of AI systems so that they can trust the decisions and proceed with them. The issue we are trying to solve in our study is this lack of confidence.

B. Explainable AI for Finance Sector

Explainable AI focusses on the last two components, namely model and stem explainability. Model explainability refers to the ability of an AI solution's internal operations to be understood, allowing humans to interpret the outcomes. Models with decreased complexity, such as linear and logistic regression, as well as decision trees, generally exhibit this characteristic. The inner workings of an AI solution are not entirely interpretable; hence people may not grasp the findings unless interpretability tools are used. This is true for complicated models—also known as black-box models—like deep neural networks. In these instances, it is usual to use model-agnostic post-hoc (and other) interpretability tools to grasp the AI's output in human terms. Similarly, XAI models can be classified into two main groups: intrinsically explainable, which include highly interpretable models such as linear and logistic regression, and extrinsically explainable, which need an additional tool to make them interpretable. Accordingly, these two classifications of models result in distinct levels of model transparency: simulatability, decomposability, and algorithmic transparency [23]. In each of these three classes, the characteristics of the previous class are inherited. Specifically, if a model is decomposable, it is also Simula table. Similarly, if a model is algorithmically transparent, it is both decomposable and Simula table.

If a model can let a person pretend to think about how it works, that's called simulatability. Decomposability means that every part of the model can be interpreted, such as the inputs, outputs, inner workings, and factors. User understanding of how the model reacts to different inputs and the capacity to reason about model faults are key to algorithmic transparency. A transparent model is a model that can be interpreted such that it can provide explanations that are understandable to humans. Transparency concerns extend beyond the model's boundaries to include the data and final product design process [24]. According to the European Union High-Level Expert Group on Artificial Intelligence [25], information that the model interacted with ought to be discoverable by human users at any given time.

Furthermore, the design process of the system should be unambiguous and coherent, therefore ensuring that it can be understood by relevant parties. Additional forms of information that can be explained include the people engaged in creating and implementing the AI system, as well as any principles or criteria that were put in place throughout its development [26]. An important justification for explainability is to establish the confidence of implicated stakeholders. Illustrations of such stakeholders encompass regulatory bodies, members of the board, auditors, end-users, and developers

[27]. To this purpose, the structure and level of elucidation differ among different audiences. Typically, the main message is communicated through reports tailored to the appropriateness of the audience consuming them. It is common knowledge that financial service providers are audited on often by supervisory agencies to ensure regulatory compliance and avoid potential fraud. The scrutiny that the authorities are likely to conduct is far more than what the service providers anticipate. This study, undertaken by [27], entails an initial inquiry to determine the specific categories of information that are considered essential from the standpoint of banks and regulatory bodies. Consequently, regulatory bodies deem all sorts of information as pertinent, whereas banks only take into account a selection of them. Accordingly, there is a discrepancy in the perception of requirement amongst corporations, which frequently causes the approval process for the implementation of financial services to be delayed. Indeed, the composition of a well-crafted explanation is mostly a matter of personal opinion. The level of necessary information typically escalates in a structured manner, beginning with the audience and extending to the regulatory authorities, as illustrated in Figure 4.

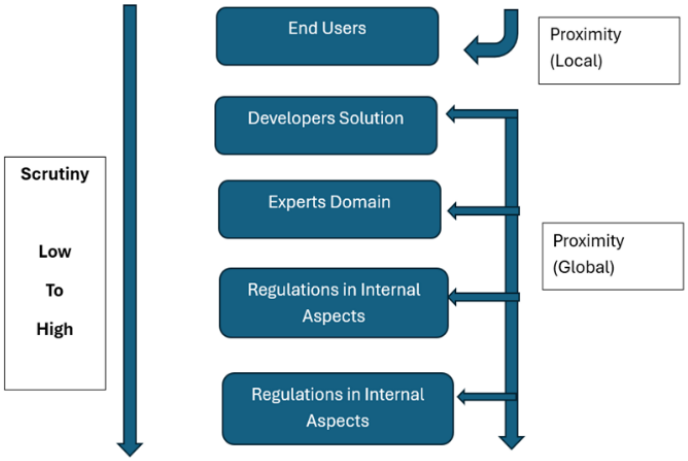

Figure 4. *User Requirement for different audience*

In this context, scrutiny denotes the quantity of information considered crucial. While end-users generally have less need for explanations regarding the cause of an outcome or data security, their primary focus is on addressing their practical issues. External regulators, on the other hand, want explanations regarding the finished item from head to toe (overall design guidelines, accountable and involved individuals, deployment procedure, internal training framework), as well as end-user requirements. Proximity determines the extent of explanation offered by the Explainable Artificial Intelligence (XAI) approach and can be characterised as local (explanatory factors concerning a specific result) and global (perception of the fundamental thinking and mechanics of the AI model). End consumers typically care about the local proximity—the method by which the consequence that affects them is supplied. For instance, an individual whose credit card application was denied would seek to ascertain the fundamental factors contributing to its rejection. Conversely, the solution providers and regulators arrange the internal operations and design workflow of the product to improve performance, ensure fairness in the model's assessment, and detect innate prejudices in the forecast (global proximity).

1. Local proximity Focus on a particular result
2. Global proximity Fundamental logic and workings of an artificial intelligence model.

An essential rationale for using explainable models is to guarantee that financial solutions comply with ethical norms established in the financial industry. According to the Monetary Authority of Singapore (MAS) guidelines, AI solutions must be created with the Fairness, Ethics, Accountability, and Transparency (FEAT) principles in mind [33]. The need for explainable models is growing primarily due to the quick development of AI solutions and the growing intricacy of their environments. Significantly, many instances of AI models exhibiting biases in their predictions intensify the need for remedies that can be explained. An infamous instance is Google's picture recognition software, which inadvertently breaks down individuals with dark complexion as gorillas.

These prejudices have the potential to reduce profits and harm the company's reputation. It is imperative to maintain the ethical standards of the financial industry in conjunction with the desirable concepts of AI ethics. These codes of financial ethics frequently coincide with ideas of artificial intelligence. An experiment employing 8 financial experts to investigate the relationship between the aforementioned sets was conducted in [34]. The findings indicate that there are many similarities between AI ethics (growth and sustainable development, human-centered values and fairness, transparency and explainability, safety and accountability) and financial ethics (integrity, objectivity, competence, fairness, confidentiality, professionalism, diligence). The strength of the ties between each factor was evaluated, with honesty and fairness having the most direct association with AI ethics. Undoubtedly, this is comprehensible considering that AI solutions should inherently possess these characteristics, irrespective of the targeted business.

C. Ethical Goals:

As previously said, the ethical objectives established by XAI solutions vary across different audiences, to provide different forms of explanations. Each audience is likely to be more impacted by one than the other. Figure 5 lists the

ethical goals endorsed for each set of viewers and demonstrates that some ethical goals overlap between audiences. [35], we present a concise elucidation of each ethical objective documented in Figure 2, adopting a financial viewpoint.

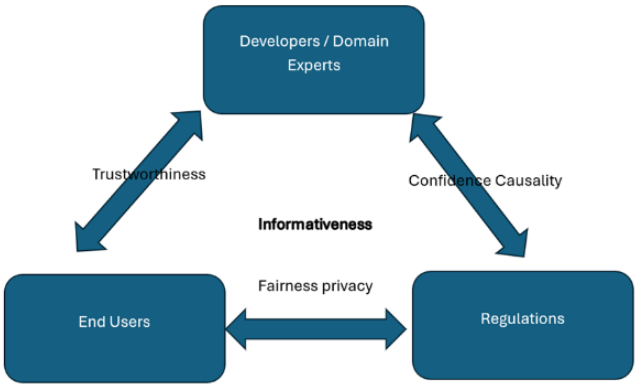

Figure 5. *Ethical Goals*

A. Accessibility:

Traditionally, the primary individuals who engage with algorithms are limited to AI developers or domain specialists. However, implementing accessibility services could enable non-experts to participate. This can be regarded as a crucial milestone in establishing AI as widespread and well embraced by the larger society. Similarly, financial institutions are discouraged from implementing these solutions due to the complexity of the algorithms, which necessitates intensive training and raises concerns about possible consequences in the event of miscalculation. A model can reduce user anxiety and encourage more organisations to implement similar practices if it can explain its workings in simple language [36].

B. Privacy Awareness:

Failure to understand the full extent of data accessible can lead to a breach of privacy. Similarly, such a matter raises concerns within the wider design process. Accountable persons involved in the design process should guarantee that third parties are granted limited access to the end-users data and take measures to prevent any misuse that may compromise the integrity of the data. Because of the volume and sensitive nature of the data being collected, privacy awareness is particularly crucial in the financial industry [37]. Confidence: The artificial intelligence model should not only deliver an outcome, but also the confidence it has in the decision-making process. This will enable domain experts to discover ambiguity in both the outcomes of the model and the zone of data that was recorded.

Stability in prediction can be used to assess a model's confidence, however explanations supplied by the model can only be considered if they produce consistent findings across multiple data inputs. Typically, developers or experts find it advantageous to comprehend the causal relationship between data aspects. Nevertheless, substantiating it is a challenging endeavour that necessitates thorough experimentation [38]. While correlation might play a role in evaluating causality, it is often not indicative of causal effect. Because AI models only find correlations in the data they learn from, domain specialists are typically required to do a more in-depth investigation of causal linkages.

C. Transferability:

The field of study focused on distilling knowledge acquired from AI models is vast. One significant advantage is that it enables the reuse of various models and prevents the need for substantial re-training. Nevertheless, the intricacy of the algorithms restricts professionals from use of trained models in various fields. A model trained to estimate future stock prices, for example, can most likely be used to predict other financial variables such as bond price, market volatility, or creditworthiness if the model's behaviour in these situations is understood. Providing a clear understanding of the internal mechanisms can alleviate the responsibility of specialists in adjusting acquired knowledge, therefore decreasing the need for meticulous perfection. Possibly one of the fundamental characteristics for enhancing future AI models is transferability [39].

CHALLENGES AND PERFORMANCE LISTED BASED ON FINANCIAL SECTOR

One of the USPs of AI models is model accuracy, which is traded off with transparency by a number of banking and financial institutions, prioritising XAI [40]. The block-box nature of explainability (for customers or users), interpretability (for designers/programmers/users), and provenance (for regulatory purposes) is addressed by XAI in current AI model. According to IBM, 68 percent of customers would demand greater explanations for AI judgements by 2024. XAI is a collection of tools, approaches, concepts, and strategies that help human stakeholders understand, evaluate, interpret, and enhance ML or AI model predictions and judgements [41]. Defence Advanced Research Projects Agency (DARPA) claims that the present AI models are generating important Why, When, and How questions. Regarding the output of these models, XAI is the solution.

AI systems that are designed, built, and implemented with an emphasis on transparency and interpretation are commonly referred to as transparent AI. In other words, transparent AI systems are characterised by their simplicity and ease of comprehension. AI needs transparency because it helps people trust AI systems and makes sure they are used in an honest

and responsible way. Users may better understand how the system generates decisions, as well as detect and rectify any biases or flaws in the system's logic, thanks to transparent AI. XAI approaches, which allow humans to understand the thinking behind an AI system's judgements, are one approach to achieving AI transparency. The level of transparency that different stakeholders experience can be categorised into algorithmic, interaction, and social [42]. The former two categories pertain to developers and regulators, while the latter two include designers and users. The creation of AI models and systems that are transparent, interpretable, and capable of offering acceptable explanations for their outputs and decision-making processes is known as XAI [43]. XAI also creates a conducive environment for technologists by increasing their efficiency to monitor, maintain, and improve AI models; business professionals by enabling trust in the output, providing suggestions and interventions for AI model alignment with organisational objectives; and legal and risk professionals by encouraging them to complain about regulatory and customer preferences [44] and [45].

CONCLUSION

Overall, explainability will remain a major area of effort in FinTech as companies strive to establish trust and confidence with both consumers and authorities. A complete assessment of XAI tools in the financial domain (FinXAI) has been published by us as a conclusion to our study. This analysis highlights the tremendous progress that has been made in recent years towards the development of explainable AI models for financial applications. This comprises both intrinsically transparent models and post-hoc explainability strategies, with the former requiring further improvement. We proposed a framework that sets the selection of relevant FinXAI tools as a sequential decision-making process, with a significant emphasis placed on the audience and an iterative evaluation of the explanation that was created. The reviewed works are organised according to their unique features for easy access by interested readers. The contributions of contemporary FinXAI to a number of ethical goals, including as trustworthiness, fairness, informativeness, accessibility, privacy, confidence, causation, and transparency, are also one of the topics that we investigate. This review demonstrates that there are several limitations and obstacles linked with FinXAI, despite the fact that there have been many amazing efforts done so far. This includes proper measures for assessing the faithfulness and plausibility of explanations, as well as concerns about over-reliance on potentially deceptive explanations.

Future research should address these problems while also investigating new FinXAI avenues, such as incorporating NLP into explanation-generating systems and emphasising intrinsically transparent models. However, there is significant potential for XAI techniques to improve transparency, trust, and accountability in the financial sector. This highlights how important it is for there to be ongoing research and development in this section of the industry.

REFERENCES

[1]. Hasib, K.M.; Tanzim, A.; Shin, J.; Faruk, K.O.; Al Mahmud, J.; Mridha, M. BMNet-5: A novel approach of neural network to classify the genre of Bengali music based on audio features. IEEE Access 2022, 10, 108545–108563.

[2]. Hasib, K.M.; Iqbal, M.; Shah, F.M.; Mahmud, J.A.; Popel, M.H.; Showrov, M.; Hossain, I.; Ahmed, S.; Rahman, O. A survey of methods for managing the classification and solution of data imbalance problem. arXiv 2020, arXiv:2012.11870.

[3]. Maitra, S.; Hossain, T.; Hasib, K.M.; Shishir, F.S. Graph theory for dimensionality reduction: A case study to prognosticate parkinson's. In Proceedings of the 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 4–7 November 2020; IEEE: New York, NY, USA, 2020; pp. 134–140.

[4]. Jahan, S.; Islam, M.R.; Hasib, K.M.; Naseem, U.; Islam, M.S. Active Learning with an Adaptive Classifier for Inaccessible Big Data Analysis. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–7.

[5]. Varmedja, D.; Karanovic, M.; Sladojevic, S.; Arsenovic, M.; Anderla, A. Credit card fraud detection-machine learning methods. In Proceedings of the 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), Novi Sad, Serbia, 20–22 March 2019; IEEE: New York, NY, USA, 2019; pp. 1–5.

[6]. Pech, R. Fraud Detection in Mobile Money Transfer as Binary Classification Problem; Eagle Technilogies Inc Publ: Arlington, VA, USA, 2019; pp. 1–15.

[7]. Kurshan, E.; Shen, H.; Yu, H. Financial crime & fraud detection using graph computing: Application considerations & outlook. In Proceedings of the 2020 Second International Conference on Transdisciplinary AI (TransAI), Irvine, CA, USA, 21–23 September 2020; IEEE: New York, NY, USA, 2020; pp. 125–130.

[8]. Pambudi, B.N.; Hidayah, I.; Fauziati, S. Improving money laundering detection using optimized support vector machine. In Proceedings of the 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 5–6 December 2019; IEEE: New York, NY, USA, 2019; pp. 273–278.

[9]. Zhang, Y.; Trubey, P. Machine learning and sampling scheme: An empirical study of money laundering detection. Comput. Econ. 2019, 54, 1043–1063.

[10]. Raiter, O. Applying supervised machine learning algorithms for fraud detection in anti-money laundering. J. Mod. Issues Bus. Res. 2021, 1, 14–26.
[11]. Lopez-Rojas, E.A.; Barneaud, C. Advantages of the PaySim simulator for improving financial fraud controls. In Intelligent Computing: Proceedings of the 2019 Computing Conference, Volume 2; Springer: Berlin/Heidelberg, Germany, 2019; pp. 727–736.
[12]. Kuppa, A.; Le-Khac, N.A. Adversarial XAI methods in cybersecurity. IEEE Trans. Inf. Forensics Secur. 2021, 16, 4924–4938. [CrossRef]
[13]. Ngai, E.W.; Hu, Y.; Wong, Y.H.; Chen, Y.; Sun, X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decis. Support Syst. 2011, 50, 559–569. [CrossRef]
[14]. Saia, R.; Carta, S. Evaluating Credit Card Transactions in the Frequency Domain for a Proactive Fraud Detection Approach; SECRYPT: Berlin, Germany, 2017; pp. 335–342.
[15]. Carcillo, F.; Le Borgne, Y.A.; Caelen, O.; Kessaci, Y.; Oblé, F.; Bontempi, G. Combining unsupervised and supervised learning in credit card fraud detection. Inf. Sci. 2021, 557, 317–331.
[16]. Zhao, Z.; Bai, T. Financial Fraud Detection and Prediction in Listed Companies Using SMOTE and Machine Learning Algorithms. Entropy 2022, 24, 1157.
[17]. Khatri, S.; Arora, A.; Agrawal, A.P. Supervised machine learning algorithms for credit card fraud detection: A comparison. In Proceedings of the 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 29–31 January 2020; IEEE: New York, NY, USA, 2020; pp. 680–683.
[18]. Hema, A. Machine Learning methods for Discovering Credit Card Fraud. IRJCS Int. Res. J. Comput. Sci. 2020, III, 1–6. 27.
[19]. Kumar, M.S.; Soundarya, V.; Kavitha, S.; Keerthika, E.; Aswini, E. Credit card fraud detection using random forest algorithm. In Proceedings of the 2019 3rd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 21–22 February 2019; IEEE: New York, NY, USA, 2019; pp. 149–153.
[20]. Seera, M.; Lim, C.P.; Kumar, A.; Dhamotharan, L.; Tan, K.H. An intelligent payment card fraud detection system. Ann. Oper. Res. 2021, 334, 445–467.
[21]. Puh, M.; Brki´c, L. Detecting credit card fraud using selected machine learning algorithms. In Proceedings of the 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 20–24 May 2019; IEEE: New York, NY, USA, 2019; pp. 1250–1255.
[22]. Lopez-Rojas, E.; Elmir, A.; Axelsson, S. PaySim: A financial mobile money simulator for fraud detection. In Proceedings of the 28th European Modeling and Simulation Symposium, EMSS, Larnaca, Cyprus, 26–28 September 2016; Dime University of Genoa: Genoa, Italy; 2016; pp. 249–255.
[23]. Hasnat, F.; Hasan, M.M.; Nasib, A.U.; Adnan, A.; Khanom, N.; Islam, S.M.; Mehedi, M.H.K.; Iqbal, S.; Rasel, A.A. Understanding Sarcasm from Reddit texts using Supervised Algorithms. In Proceedings of the 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC), Hyderabad, India, 6–18 September 2022; IEEE: New York, NY, USA, 2022; pp. 1–6.
[24]. Niklas Bussmann, Paolo Giudici, Dimitri Marinelli and Jochen Papenbrock, Explainable Machine Learning in Credit Risk Management, Computational Economics (2021) 57:203–216.
[25]. Ambreen Hanif, Towards Explainable Artificial Intelligence in Banking and Financial Services, arXiv:2112.08441v1 [cs.LG] 14 Dec 2021.
[26]. Černevičienė, Jurgita ; Kabašinskas, Audrius, On multi-criteria decision-making methods in finance using explainable artificial intelligence, AMSS 2022: 13th conference on data analysis methods for software systems, Druskininkai, Lithuania, December 1–3, 2022.
[27]. Islam, M.T.; Hasib, K.M.; Rahman, M.M.; Tusher, A.N.; Alam, M.S.; Islam, M.R. Convolutional Auto-Encoder and Independent Component Analysis Based Automatic Place Recognition for Moving Robot in Invariant Season Condition. Hum. Centric Intell. Syst. 2022, 3, 13–24. [CrossRef]
[28]. Wang, Y.; Pan, Z.; Zheng, J.; Qian, L.; Li, M. A hybrid ensemble method for pulsar candidate classification. Astrophys. Space Sci. 2019, 364, 139.
[29]. Cody, C.; Ford, V.; Siraj, A. Decision tree learning for fraud detection in consumer energy consumption. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 9–11 December 2015; IEEE: New York, NY, USA, 2015; pp. 1175–1179.
[30]. Javed Mehedi Shamrat, F.; Ranjan, R.; Hasib, K.M.; Yadav, A.; Siddique, A.H. Performance evaluation among id3, c4. 5, and cart decision tree algorithm. In Pervasive Computing and Social Networking: Proceedings of ICPCSN 2021; Springer: Berlin/Heidelberg, Germany, 2022; pp. 127–142.

[31]. Ruiz-Gonzalez, R.; Gomez-Gil, J.; Gomez-Gil, F.J.; Martínez-Martínez, V. An SVM-based classifier for estimating the state of various rotating components in agro-industrial machinery with a vibration signal acquired from a single point on the machine chassis. Sensors 2014, 14, 20713–20735.
[32]. 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.
[33]. Saurabh Kumar, "Difference-in-Differences in Action: Measuring Brand Marketing Campaign Impact Through Survey Responses", International Journal of Science and Research (IJSR), Volume 7 Issue 9, September 2018, pp. 1669-1673, https://www.ijsr.net/getabstract.php?paperid=SR18920585948