DOI 10.37539/2949-1991.2025.28.5.011

Ogunniyi Oluwaseyi Olatunde,
Master's student, Faculty of International Economic Relations,
Financial University under the Government of the Russian Federation,
Russia Moscow
ORCID: 0009-0008-2405-9986

ОБЪЯСНИМОСТЬ И ДОВЕРИЕ В ПРОЦЕССЕ ПРИНЯТИЯ ФИНАНСОВЫХ РЕШЕНИЙ, ОСНОВАННЫХ НА ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ
EXPLAINABILITY AND TRUST IN AI-DRIVEN FINANCIAL DECISIONS

Аннотация: Искусственный интеллект (ИИ) играет важную роль в принятии решений на международном финансовом рынке. Однако его непрозрачность, называемая проблемой «чёрного ящика», вызывает обеспокоенность. Объяснимый ИИ (XAI) помогает понять логику, процессы и факторы, лежащие в основе решений. В статье рассматривается роль XAI в повышении прозрачности, справедливости и доверия в финансах. Анализируются технические аспекты, применение, этические нормы и перспективы развития. Цель способствовать ответственным и подотчётным инновациям на основе ИИ в финансовых учреждениях.

Abstract: Artificial Intelligence (AI) plays a key role in decision-making in the international financial market. However, its lack of transparency often called the “black box” problem raises concerns. Explainable AI (XAI) addresses this by making AI's logic, processes, and decision factors understandable. This article explores XAI’s role in enhancing transparency, fairness, and trust in finance. It examines its technical aspects, applications, ethical considerations, and future outlook. The goal is to promote responsible and accountable AI-driven innovation in financial institutions.

Ключевые слова: Объяснимый искусственный интеллект (XAI), прозрачность, доверие, интерпретируемость, черный ящик

Keywords: Explainable artificial intelligence (XAI), transparency, trust, interpretability, black-box

1. Introduction

Artificial intelligence (AI) is changing how financial decisions are made in banks, investment companies, and generally financial institutions. With the use of powerful tools, such as machine learning and deep learning, AI can analyze huge amounts of data quickly and make accurate predictions. With these, financial institutions can offer better services, spot fraud early, and make smarter investment choices. However, these powerful AI systems are complex and opaque, working as “black boxes,” which is a phenomenon, meaning it's not always clear AI arrive at their decisions.

As AI becomes a bigger part of our financial lives, knowing exactly why it makes certain decisions and how it does becomes extremely important. People need to trust that these AI systems are fair, ethical, and reliable, most especially when the decisions made directly tend to impact their lives, for example, getting approved for loans, insurance pricing, or investment advice. Regulators in the EU via the GDPR and AI Act, have started to emphasize the importance of transparency in AI systems, demanding clear explanations for automated financial decisions. Due to this, XAI has emerged as an important area of research. XAI helps users clearly see how AI models make decisions, giving them understandable explanations. This transparency helps people trust and accept AI-driven services more easily.

This article explores why explainability is critical to building trust in AI financial systems. It discusses the challenges faced when trying to create transparent AI, the current methods used in the finance sector to improve explainability, and how these efforts are shaping the future of financial decision-making.

2. The Need for Explainability in Financial AI

As Artificial Intelligence continues to significantly influence decision-making processes across various industries, the financial sector in particular stands out due to how critical the nature of its operations can be. Financial decisions that are powered by AI for example credit scoring, loan approvals, investment strategies, and fraud detection tend to have serious implications for individuals and businesses alike. While these advanced AI systems offer high level of efficiency and predictive accuracy, the way they work remains unknown, making it challenging for stakeholders to understand and put their trust in these decisions.

Explainable AI (XAI) has emerged as a response to these challenges, with the objective of making AI systems' decision-making processes transparent and interpretable to humans. According to research, "Interpretable models provide a path toward trustworthy AI by enabling humans to understand, validate, and trust AI-generated decisions" [9]. Financial institutions adopting XAI techniques can effectively understand AI decision-making, leading to greater level of trust among customers, regulators, and society at large.

Moreover, transparency and interpretability are features that are becoming increasingly mandated by regulatory frameworks globally. For instance, the General Data Protection Regulation (GDPR) enacted by the European Union stipulates that individuals have a "right to explanation," [7] particularly when AI automated decisions can significantly impact their lives. This regulation calls for an urgent need for financial institutions to incorporate explainability into their AI deployment strategies.

Furthermore, the adoption of XAI allows institutions to identify and correct certain biases that may exist within AI models, ensuring that there is fairness and ethical integrity is maintained. By adding explainability into their AI practices, financial organizations can mitigate risks associated with biased or unfair treatment, enhancing their reputation and customer loyalty.

Explainable AI is not just beneficial but rather an important requirement for the usage of AI within the financial sector. Transparency in AI decision-making is essential for compliance, ethical responsibility, and maintaining trust, making explainability a component that cannot be left out in modern financial practices.

3. Challenges of AI Opacity

Black box models, such as deep learning often produce highly accurate predictions but lack the transparency. These methods make compliance difficult, for example compliance with regulations such as the European Union’s General Data Protection Regulation (GDPR), which mandates the "right to explanation". Further, opaque algorithms can encourage biases and unfair treatment without accountability.

4. Principles of Explainable AI (XAI)

To develop and implement effective explainable AI systems, several foundational principles guide both research and application. These principles ensure that AI models are not only technically proficient but also ethically aligned and socially acceptable.

Transparency Transparency refers to the clarity with which an AI system communicates its decision-making process. This includes the structure of the model, the data it uses, and the logic it applies. A transparent system allows users to inspect how the AI algorithms functions and enables them understand the relationships that exists between inputs and outputs. A model is considered transparent by itself if it is understandable [1].

Interpretability Interpretability refers to how well a user can get the reason of a decision made by an AI model. By this, it is meant the ability to explain or be able to present the decision-making process in forms whereby its logic can be followed. This principle allows experts trace and find out how inputs turn into output which in turn is beneficial for the sake auditing, regulatory compliance especially in the finance sector.

Trustworthiness Trustworthiness is all about the confidence users have in an AI system's decisions. Trust is built when systems are transparent, are able to perform consistently, and provide explanations that tally with the underlying logic. Explanations must align with the model's actual behavior, not just plausible rationalizations.

Fairness Fairness ensures equity to all. It makes sure that the decisions made by AI are not bias, that is, they are not favourable to a certain group or set of users over that of another group, especially people or populations that are marginalized or underrepresented. XAI tools are instrumental in identifying biased patterns in data or model behavior, allowing corrective measures to be taken [4].

5. Applications in Financial Decisions

Credit Scoring

Credit scoring is an analysis that is usually performed by most financial institutions and lenders to determine the creditworthiness of an individual or a small, owner operated business. It is used to help decide whether a credit should be denied or extended. Overall, a credit score can impact other areas in terms of qualification for financial products like mortgages, credit cards and private loans etc. The credit score uses historical and current financial data and is influenced by factors such as the payment history, debts, length of credit history among others. Credit rating which is a similar concept is totally differenat. Most traditional AI models do operate as black boxes, leaving applicants and regulators not knowing how the decisions are made.

To address this situation, XAI techniques such as the Shapley Additive Explanations (SHap) and the Local Interpretable Model-Agnostic Explanations (LIME) can be used. These techniques provide intuitive, feature level attributions which helps applicants to understand the reasons why they were accepted or rejected, showing transparency, fairness on the path of the financial institutions [6]. XAI enables customers to challenge data inaccuracies such as amount of income, employment history that may have occurred and promotes ethical lending practices. It also enhances the internal governance of the model allowing the credit officers and compliance team to better monitor and validate AI decisions.

Fraud Detection

The issue of fraud is a complex one. It is more complex year after year and has become a major concern in the financial industry. Hence, the application of AI in detection of fraud is critical. With the use AI, little anomalies to transactions are detected which might have been difficult for humans to detect quickly. Despite this, without being able to explain or understand how it does, detection can be met with a lot of skepticism, especially when legitimate transactions are flagged incorrectly.

However, with the use of explainable models and visualization techniques such as decision trees, SHAP, LIME, financial institutions can justify fraud alerts to both risk teams and external regulators. Trust and transparency are enhanced with the use XAI in fraud detection and it has far more reaching implications for financial security. XAI supports human-AI collaboration, ensuring that fraud analysts can validate andrefine machine-generated predictions rather than blindly accepting AI decisions. This reduces therisk of both false positives and false negatives, making fraud detection systems more reliable and effective [5].

Regulatory Compliance

In regulated contexts, explainability serves many purposes such as building trust, it supports auditing, and makes sures that whatever AI decisions are made, it is made within the jurisdiction of the legal and ethical framework. These features come in handy when decisions probably impact the rights of individuals and overall well-being.

Financial institutions must comply with important regulations like GDPR, SEC and the AI Act, these frameworks and regulatory bodies call for transparency, accountability and fairness in

automated decision making thereby making its being able to explain not just a feature, but also a compliance imperative [8]. Explainable AI provides a robust mechanism for compliance by enabling clear documentation of decision-making processes. Regulators increasingly demand that AI models used in financial decisions are both transparent and auditable.

Algorithmic Auditing

Algorithmic auditing has become a vital practice to ensure AI systems are operating fairly, safely, and responsibly. It involves systematic evaluations by internal compliance teams or independent third parties to assess the behavior, risks, and ethical implications of AI models. Explainable AI techniques are applicable during audits as they are instrumental. Some auditing terminologies include:

Sensitivity analysis; can reveal how slight changes in inputs affect outcomes, highlighting vulnerabilities or biases.

Fairness metrics; can quantify the differences across demographic groups, providing an empirical basis for model adjustments.

Thorough documentation reviews; ensuring that models, training data, and decision rationales are properly recorded which further contributes to transparency.

In the financial sector, regular algorithmic audits supported by XAI tools are now becoming regulatory expectations. They help detect systemic biases eon time, they ensure that anti-discrimination laws are complied with, and they maintain public confidence in AI-driven financial systems.

6. Emerging Solutions

Several advanced approaches and tools are emerging to address explainability challenges in financial AI:

- Post-hoc explanation methods (e.g., SHAP, LIME)
- Intrinsically interpretable models (e.g., linear regression, decision trees)
- Counterfactual explanations: Illustrating alternative scenarios under which different outcomes would occur.
- Interactive visual analytics: Allowing stakeholders to explore model behaviors intuitively.

To bridge the gap between high-performing but black-boxed models and the need for interpretability in critical applications like finance, several sophisticated methods have been developed. These techniques allow data scientists, regulators, and end-users to understand the logic behind AI predictions and decisions. The most prominent approaches include post-hoc explanation methods, intrinsically interpretable models, counterfactual explanations, and interactive visual analytics.

Post-Hoc Explanation Methods

These are techniques which are applied after an AI model has been trained, providing explanations for decisions made by complex, black-box systems. Among the most widely used tools in this category are:

SHAP (SHapley Additive Explanations): Based on cooperative game theory, SHAP assigns each feature a contribution score that reflects its importance in a particular prediction. In financial applications such as credit scoring or fraud detection, SHAP helps explain why a model approved one applicant while rejecting another. It is highly regarded for offering both global and local explanations, giving insights into overall model behavior as well as individual cases.

LIME (Local Interpretable Model-Agnostic Explanations): LIME explains the predictions of any black-box classifier by approximating it locally with an interpretable model. For instance, in detecting suspicious transactions, LIME can produce a simple linear explanation of why a specific transaction was flagged, even when the underlying model is a deep neural network.

These tools are model-agnostic which they can be applied to any black model and can be integrated with any machine learning pipeline, making them indispensable in financial services where model transparency is essential for compliance and trust.

Intrinsically Interpretable Models

Unlike post-hoc methods, intrinsically interpretable models are designed to be transparent from the onset. That is, the models are built with explainability as a major feature, so that the decisions made clear to interpret. While they may not always match the performance of more complex algorithms, they provide clear insight into how decisions are made. Examples include:

Linear Regression and Logistic Regression: These models are easy to interpret, as they explicitly show the relationship between input features and the output variable. In finance, they are commonly used for assessing credit risk and estimating default probabilities.

Decision Trees**:** They are hierarchical models following a structure of feature-based splits, decision trees allow users to trace the logic behind each prediction. For example, a decision tree might reveal that applicants with a credit score above 700 and income over $50,000 are more likely to be approved for a loan.

Trees are graphically simple, thereby making it easy for anyone to understand. These models are particularly valuable in regulated industries, where accountability and explainability outweigh the marginal benefits of higher predictive accuracy. They however do have some limitations, for example, little changes in data can cause serious changes in the tree structure.

Counterfactual Explanations

Counterfactual explanations answer the question: "What would need to change in the input to receive a different outcome?" They show how a little change in input could lead to a different prediction. These explanations are intuitive and user-centric, making them particularly useful in consumer-facing financial applications. For instance, if an AI system denies a mortgage application, a counterfactual explanation might suggest that increasing the applicant's monthly income by $500 or reducing their credit utilization ratio could have resulted in approval. This form of feedback is actionable and supports fairness by providing users with a path to a different decision.

From a regulatory perspective, counterfactuals also support transparency, as they offer insight into decision boundaries and reveal the factors most influential in changing outcomes.

Interactive Visual Analytics

Interactive visual tools play a crucial role in making complex AI models more understandable, particularly for non-technical stakeholders. These platforms often integrate with SHAP, LIME, or proprietary interpretation engines to provide dynamic dashboards and explorable model visualizations.

Use in Credit Scoring**:** A bank's compliance officer can use visual dashboards to analyze how different variables affect the likelihood of loan approval across various customer segments.

Fraud Detection**:** Risk analysts can interact with real-time anomaly detection outputs, exploring patterns and refining alert thresholds based on visual trends. Interactive visual analytics promote transparency and collaboration across departments by translatingAbstract model mechanics into intuitive, actionable insights.

7. Challenges of Explainable AI

Despite its immense potential to enhance transparency and trust in AI systems, XAI is not without its several practical, technical, and philosophical challenges. These challenges complicate the development, adoption, and implementation of explainability frameworks, especially in high-stakes fields such as finance. Below are some of the most prominent challenges facing XAI today:

1. Balancing Performance and Interpretability

A fundamental challenge in XAI is the tension between model performance and interpretability. High-performing AI models, such as deep neural networks and ensemble methods, are typically opaque and difficult to interpret. Conversely, models that are inherently interpretable, such as decision trees or linear regressions, often underperform on complex tasks compared to black-

box models. Striking a balance between accuracy and explainability remains a key obstacle, particularly when financial institutions must choose between maximizing returns and ensuring transparency.

2. Lack of Standardized Evaluation Metrics

Unlike model accuracy or precision, explainability lacks standardized quantitative metrics. This makes it difficult to compare models or explanation techniques across applications. What is interpretable to a data scientist may not be understandable to a loan applicant or compliance officer. The subjective nature of interpretability complicates efforts to define universal standards, which is essential for regulatory acceptance and cross-industry adoption.

3. Model Complexity and Scalability

As financial data becomes increasingly high-dimensional and models more complex, the scalability of XAI techniques is a pressing concern. Many explanation methods are computationally expensive and not feasible for real-time or large-scale applications. For instance, SHAP and LIME, while effective, require multiple model evaluations to generate a single explanation, posing challenges for institutions with high-volume, low-latency environments like algorithmic trading.

4. Information Overload

Another paradox in XAI is the risk of overwhelming users with too much information. Detailed explanations that include feature importance scores, model graphs, or statistical justifications may not be useful or even comprehensible to non-expert stakeholders. Delivering explanations that are not only accurate but also comprehensible and context-appropriate is a significant communication challenge.

5. Security and Adversarial Risks

Revealing how an AI system functions may expose it to adversarial attacks. When its inner workings are understood, malicious actors could use explanation outputs to reverse-engineer models, identify weaknesses, and manipulate outcomes. In the financial industry, this could lead to fraud, regulatory breaches, or market manipulation. Therefore, XAI solutions must be carefully designed to balance transparency with model security.

6. Ethical and Legal Ambiguities

The legal implications of XAI remain underexplored. For example, does providing an explanation absolve institutions from liability, or does it increase it by formalizing the rationale for a potentially discriminatory outcome? Additionally, ethical concerns arise when explanations are manipulated to justify flawed decisions. These ambiguities necessitate clearer guidelines on the ethical deployment of XAI systems.

7. Human Factors and Cognitive Biases

Human users of XAI systems may interpret explanations through the lens of cognitive biases [2], leading to over-trust or under-trust in AI decisions. For instance, a visually appealing dashboard may lend undue credibility to a flawed model. Ensuring that explanations align with human reasoning without misleading or manipulating users requires careful attention to human-centered design principles.

Future Directions

The future of explainability in AI financial decisions points toward increased standardization of explainability methods, deeper integration of human-AI collaboration, and broader regulatory frameworks to enforce transparency. Developments such as neurosymbolic AI [3], combining symbolic reasoning with neural networks, promise further advancements in interpretability.

Conclusion

Explainable AI is becoming indispensable when it comes to building trust, transparency, and adhering to regulatory compliance within the financial industry. By addressing the black-boxed models, financial institutions can mitigate risks, improve level of trust with stakeholders, and fulfill

regulatory obligations. The proactive implementation of XAI methods puts institutions strategically in a good position for a future where the use of responsible AI is both a societal expectation and a regulatory requirement.

References:

1. Arrieta A. Rodriguez N. Ser J. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. URL: https://arxiv.org/pdf/1910.10045 (accessed: 26.04.2025).
2. Bertrand A. Belloum R. Eagan J. Maxwell W. How Cognitive biases affect XAI-assisted decision-making: A systematic review. URL: 10.1145/3514094.3534164 (accessed: 03.05.2025).
3. Colelough B. Regli W. Neuro-Symbolic AI in 2024: A Systematic Review. URL: 10.48550/arXiv.2501.05435 (accessed: 03.05.2025).
4. Deck L. Schomacker A. Speith T. Mapping the Potential of Explainable AI for Fairness Along the AI Lifecycle. URL: https://arxiv.org/pdf/2404.18736 (accessed: 26.04.2025).
5. Faruk N. Tariq A. Oladele S. Gok M. Explainable AI (XAI) for Fraud Detection: Building Trust and Transparency in AI-Driven Financial Security Systems. URL: https://www.researchgate.net/publication/390235753_Explainable_AI_XAI_for_Fraud_Detection_Building_Trust_and_Transparency_in_AI-Driven_Financial_Security_Systems (accessed 27.04.2025).
6. James C. Explainable AI in Credit Scoring: Improving Transparency and Accountability. URL: https://www.researchgate.net/publication/387456851_Explainable_AI_in_Credit_Scoring_Improving_Transparency_and_Accountability/citations (accessed: 27.04.2025).
7. Kaminski M. The Right to Explanation, Explained. URL: https://scholar.law.colorado.edu/faculty-articles/1227. (accessed: 22.04.25)
8. Rajuroy A. Regulatory-Complaint Explainable AI: Case studies and frameworks for financial, healthcare and government sectors. URL: https://www.researchgate.net/publication/390735985_Regulatory-Compliant_Explainable_AI_Case_Studies_and_Frameworks_for_Financial_Healthcare_and_Government_Sectors (accessed: 27.04.2025).
9. Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. URL: http://dx.doi.org/10.1038/s42256-019-0048-x (accessed: 22.04.25)