Journal of Frontiers in Multidisciplinary Research

Explainable Artificial Intelligence for Financial Crime Prevention: Translating Machine Learning Outputs into Regulatory and Complia Decision-Making

Okolie Awele [1*], Daniel Oghenekome Erebi [2], Bright Kofi Ladzro [3], Oluwatosin Lawal [4], Didunoluwa Olukoya [5], Samson Onaopemipo Amoran [6]

[1] School of Computing and Data Science, Wentworth Institute of Technology, Boston, USA
[2] Department of Electrical and Electronic Engineering, Federal University of Petroleum Resources, Effurun, Nigeria
[3] Department of Mathematics and Statistics, College of Arts and Sciences, American University, Washington, DC, United States
[4] Department of Mathematics Statistical Analytics, Computing and Modeling, Texas A&MUniversity, Kingsville, USA
[5] Independent Researcher, USA
[6] Department of Computer Science, Western Illinois University, USA.

* Corresponding Author: **Okolie Awele**

Article Info

E-ISSN: 3050-9726
P-ISSN: 3050-9718
Impact Factor (RSIF): 8.10
Volume: 05
Issue: 02
July - December 2024
Received: 27-09-2024
Accepted: 29-10-2024
Published: 01-12-2024
Page No: 148-153

Abstract

Financial institutions increasingly rely on machine learning systems to detect fraudulent activities and prevent financial crime across complex transactional environments. Models like these are exemplified by their remarkable prediction accuracy, but at the same time, their lack of interpretability leads to considerable problems in terms of compliance with regulations, auditing, and operational trust. In the case of financial transactions that carry a high level of risk, regulators and compliance professionals want not just precise risk forecasts but also the reasoning behind the decisions to be open and understandable. The presented research introduces a regulatory-compliant explainable artificial intelligence (XAI) framework that connects machine learning results and financial crime decision-making processes. The framework does not innovate but rather focuses on converting risk scores and explainability outputs of predictive models into interpretable decision artifacts that can be easily reviewed for compliance, supervisory oversight, and human-in-the-loop validation. The proposed methodology combines explainability mechanisms with governance-oriented design principles, allowing for consistent justification of flagged transactions, enhanced audit trails, and increased accountability in automated systems for financial crimes detection and prevention. Case study illustrations demonstrate how explainable AI can support escalation, investigation, and reporting decisions in fraud and anti-money laundering contexts. The findings highlight the role of explainable AI as a critical enabler for aligning machine learning innovation with regulatory expectations, contributing to more transparent, trustworthy, and responsible financial crime prevention systems.

DOI: https://doi.org/10.54660/.IJFMR.2024.5.2.148-153

Keywords: Explainable Artificial Intelligence, Financial Crime Prevention, Regulatory Compliance, Decision Support Systems, Fraud Detection, Governance.

1. Introduction

Financial crime such as fraud and money laundering, is still a major problem and a source of global financial instability and economic insecurity. The digitization of financial services which envelops mobile payments, online banking, and cryptocurrency transactions, has not only increased the number of illicit financial activities but also made them more complicated.

Financial institutions have reacted by progressively applying machine learning-based systems to detect suspicious transactions and handle financial crime risk more efficiently (Dal Pozzolo et al., 2018; Bahnsen et al., 2015) [4,2,3]. These systems, despite being challenging to understand and interpret, have occupied predictive capability much improved over rule-based approaches. The heavy dependence on complex and opaque models has however, resulted in new challenges concerning the issues of transparency, accountability, and the trust of regulators. The regulatory authorities and supervisory bodies insist on the explainability and justification in the financial crime decision-making process. Institutions must state clearly the reasons for the transaction flagging, the customer risk assessments, and the reporting decisions, especially in the context of audits and supervisory reviews (Basel Committee on Banking Supervision, 2015; Financial Action Task Force, 2021) [1]. Nonetheless, numerous top-performing machine learning models like ensemble methods and deep learning architectures are working as black boxes, generating risk scores without providing understandable reasons. This disparity between predictive accuracy and regulatory usability limits the practical application of advanced analytics within compliance-critical environments and exposes institutions to governance and model risk concerns.

Explainable artificial intelligence (XAI) has emerged as a promising approach to address these challenges by enhancing the interpretability of machine learning models without significantly compromising performance (Doshi-Velez & Kim, 2017; Lundberg & Lee, 2017) [5,8]. Existing XAI research in financial crime detection has primarily focused on post-hoc explanation techniques, such as feature importance analysis and local explanation methods, to improve transparency for analysts and model developers (Guidotti et al., 2019) [7]. While these approaches provide valuable insights into model behavior, they are often treated as supplementary visualizations rather than integral components of regulatory decision workflows. As a result, explainability outputs are rarely structured in a manner that directly supports compliance validation, audit documentation, or supervisory oversight.

Consequently, the research on explainable machine learning has created a critical gap between the operational realities of financial regulation. In high-stakes decision environments, the need for explainability goes beyond merely offering technical interpretability and involves providing the support for the automated decisions to be accountable, consistent, and defensible (Rudin, 2019) [10]. Regulators and compliance teams are in need of such decision artifacts that would not only show the model outputs but also demonstrate the risk drivers, the contextual reasoning and the human judgment that link them, and the transparent escalation, investigation and reporting processes enabled. Without this kind of alignment, the use of machine learning systems for the prevention of financial crime is still to a great extent limited by lack of trust, bias, and governance concerns. To tackle this issue, this research suggests a framework of regulatory-aligned explainable artificial intelligence designed to link the machine learning outputs to the decision-making processes in financial crimes. The work does not aim at the development of new predictive models or at the improvement of classification performance but rather aims at making the interpretability of the model understandable in a regulatory and compliance context. The proposed framework brings together explainability and governance-oriented design principles, which in turn help in converting model risk scores into comprehensible decision rationales which can be used during audit review, supervisory evaluation, and human-in-the-loop oversight. By repositioning explainable AI as a decision-support and governance tool, this research offers a structured method for the alignment of machine learning innovation with regulatory expectations in financial crime detection systems.

2. Background and Related Work

2.1. Machine Learning in Financial Crime Prevention

Machine learning has turned out to be a main stay in the detection and prevention of financial crime, particularly in the areas of fraud, money laundering, and unscrupulous transaction activities. The traditional rule-based systems have proved to be inadequate most of the time because of the huge volume and the intricate nature of the current financial transactions (Bahnsen et al., 2015; Dal Pozzolo et al., 2018) [4,2,3]. Supervised learning techniques, one of which is deep learning, including the use of ensembles, yield excellent results in predicting anomalous transactions across the various financial platforms. Nevertheless, the majority of these systems function as “black boxes” that offer little or no insight into the processes that led to the predictions. Such non-transparency bears down on the institutions' capability of providing justifications for the flagged transactions and meeting regulatory and audit requirements (Rudin, 2019) [10].

2.2. Explainable Artificial Intelligence (XAI)

The Explainable AI (XAI) initiative faces the challenge of reconciling high predictive performance with interpretability. Among these are SHAP (Lundberg & Lee, 2017) [8] and LIME (Ribeiro et al., 2016) [9], which are called post-hoc model interpretations; by translating the outputs of complex models into human-interpretable feature-level contributions, they do the job of explaining the predictions inferred by the models. In the areas of financial crime, these techniques can be of great assistance to the analysts as they will be able to comprehend the reasoning behind a flagged transaction, thus, the efficiency of the investigation would possibly be improved and the number of false positives decreased (Doshi-Velez & Kim, 2017; Guidotti et al., 2019) [5,7]. However, despite such progress, the focus of the current studies has been on understanding by analysts, thus, regulatory scrutiny and compliance decision-making are still unmet.

2.3. Regulatory and Compliance Challenges

Financial regulatory frameworks have notably stressed the importance of accountability, auditability, and transparency in automated decision systems such as those issued by the Financial Action Task Force (FATF, 2021) and Basel Committee on Banking Supervision (2015) [1]. This is, in part, due to the fact that the regulation requires the institutions to not only identify questionable activities but also to explain the standing of their decisions in a clear and auditable way. Black-box models, which are very good at making predictions, still are not able to meet these regulatory standards and thus must not be adopted in compliance-critical environments. Furthermore, explainability is not enough by itself unless it is woven into decision-making workflows where the outputs can be converted into actionable insights for investigators, auditors, and supervisory authorities.

2.4. Gap in Current Research

Current research on XAI in financial crime primarily focuses on technical interpretability or analyst support, with limited exploration of regulatory-aligned frameworks. Few studies systematically address how explainability can be operationalized to meet governance requirements, support audit trails, or facilitate human-in-the-loop oversight (Rudin, 2019; Guidotti et al., 2019) [10,7]. This gap motivates the development of a regulatory-aligned XAI framework, capable of translating machine learning outputs into interpretable, actionable, and audit-ready decision artifacts.

3. Regulatory and Compliance Challenges

The increasing complexity of financial transactions and the widespread use of machine learning systems in fraud detection and anti-money laundering (AML) have introduced significant regulatory and compliance challenges. Regulatory authorities, including national supervisory agencies and international bodies such as the Financial Action Task Force (FATF, 2021) and the Basel Committee on Banking Supervision (2015), mandate that institutions maintain transparency, accountability, and auditability in all decision-making processes related to financial crime.

3.1. Transparency and Accountability

One of the main requirements that the regulatory frameworks imposed on the financial sector was transparency. The use of automated decision-making systems needed to be accompanied by clear justifications for actions like transaction flagging, risk scoring, and customer classification. Models of the black-box type, particularly those based on ensemble or deep learning, are often unable to be used for regulatory examination purposes, thus increasing the fear of compliance breaches and the burden of proving accountability on the part of the institution (Rudin, 2019) [10]. The lack of interpretable outputs makes it difficult for the organizations to exhibit the rationale for the flagged activities, which consequently puts them at the legal and operational risk.

3.2. Auditability and Documentation

Audit teams and regulators expect every step of the financial crime detection process to be documented in detail. This not only entails but also requires providing the logic behind risk scoring and the reasons for escalation or dismissal, thereby including clearly indicating the transactions that were flagged and giving the rationale for doing so. Conventional machine learning models deliver very high accuracy but at the same time, they are very limited in terms of their interpretability, which in turn causes the difficulty of producing audit-ready reports that meet the requirements of the regulators (Basel Committee on Banking Supervision, 2015; Guidotti et al., 2019) [7]. In the case of the AML compliance, especially in the area of audits, financial institutions must ensure that the automated decision can always be traced, explained, and justified.

3.3. Human-in-the-Loop Oversight

Regulations are more and more stressing the necessity of human oversight in decision-making processes. Automated systems will not take over the role of humans in evaluating suspicious activities, they will only be their partners (FATF, 2021). The use of human-in-the-loop decision-making along with explainable AI will make it possible for compliance teams to check the outputs of AI, change the limits of acceptance, and assign risk scores based on the broad company policies. One of the benefits of having humans in the loop is that regulatory requirements are met, and at the same time, the risk of bias and error in automated decisions is reduced (Doshi-Velez & Kim, 2017) [5].

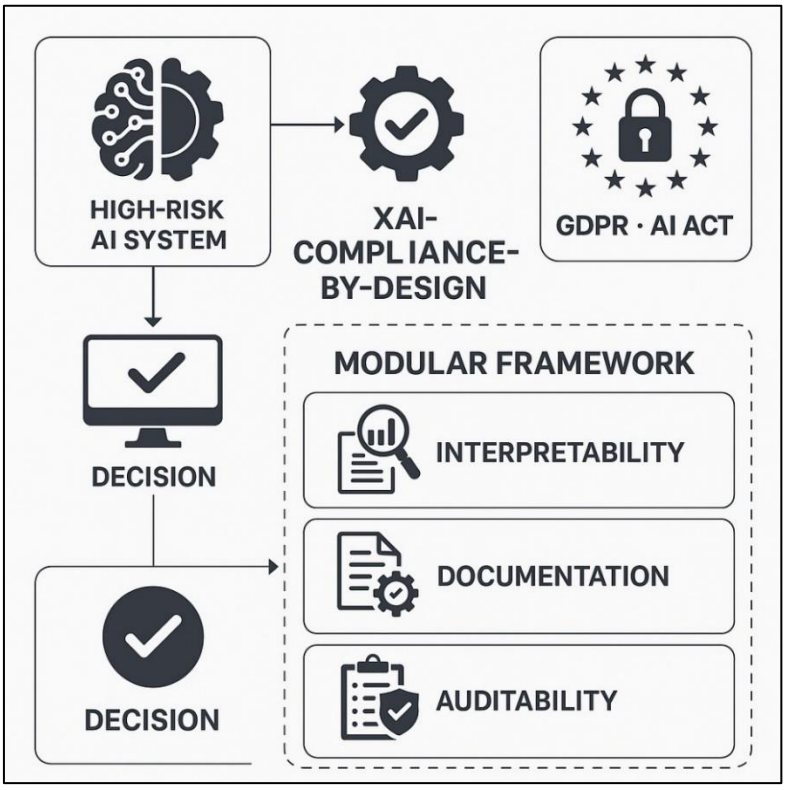

Fig 1: Xai Compliance by Design

3.4. Challenges in Operational Deployment

Using XAI in the context of the financial market wherein there are heavy regulations is not easy. One of the difficulties that organizations come across is coming up with one standard way of interpreting across all the different models; another is how to incorporate XAI outputs into the already existing compliance workflows and finally, they have to see to it that it is done according to what the supervisors expect. On top of that, they have to deal with the fact that the need for the explainability of the models to non-technical stakeholders is a feature of the regulators (Lundberg & Lee, 2017; Guidotti et al., 2019) [8,7]. These operational challenges highlight the need for a structured framework that will be able to convert model predictions into decision artifacts that are actionable and compliance-aligned.

4. Framework Architecture

The framework consists of three layers, each translating predictive outputs into regulatory-compliant insights:

1. **Predictive Layer:** The input for this layer is transactional data, and the output is risk scores calculated by the use of pre-trained machine learning models. The system is independent of the models used, and it can easily handle traditional classifiers (e.g., Random Forest) as well as more sophisticated methods (e.g., XGBoost, neural networks). The focus here is not on creating new models but rather on integrating existing ones into a regulatory-compliant workflow.
2. **Explainability Layer:** The explainability layer converts risk scores into feature-level contributions which provide both global and local interpretability. Global explanations: They help to identify patterns or features that are always important across the dataset (e.g., a high transaction frequency that is positively correlated with an increase in fraud risk). Local explanations: They offer reasoning specific to the transaction, portraying exactly how a particular transaction was marked (e.g., an unusually large transaction from a region that is considered high-risk). The compliance teams, by utilizing both global and local insights, reap the benefit of gaining a macro-level understanding and at the same time receiving a micro-level actionable guidance.
3. **Regulatory Interpretation Layer:** This layer transforms the results of the explainability process into the artifacts of the decision that are appropriate to be used for audits, reporting, and compliance:
4. **Justification reports for flagging:** Highlight the main risk factors and suggest measures.
5. **Workflows for escalation:** Classify cases with the highest risk for human review or regulatory submission.
6. **Trails for audit:** Document choices, justifications, and reviewer contributions for the purpose of verifiability.
7. **Integration of human intervention:** Create platforms where officers can approve, reject, or provide feedback on model decisions, thus ensuring supervision and responsibility are upheld.

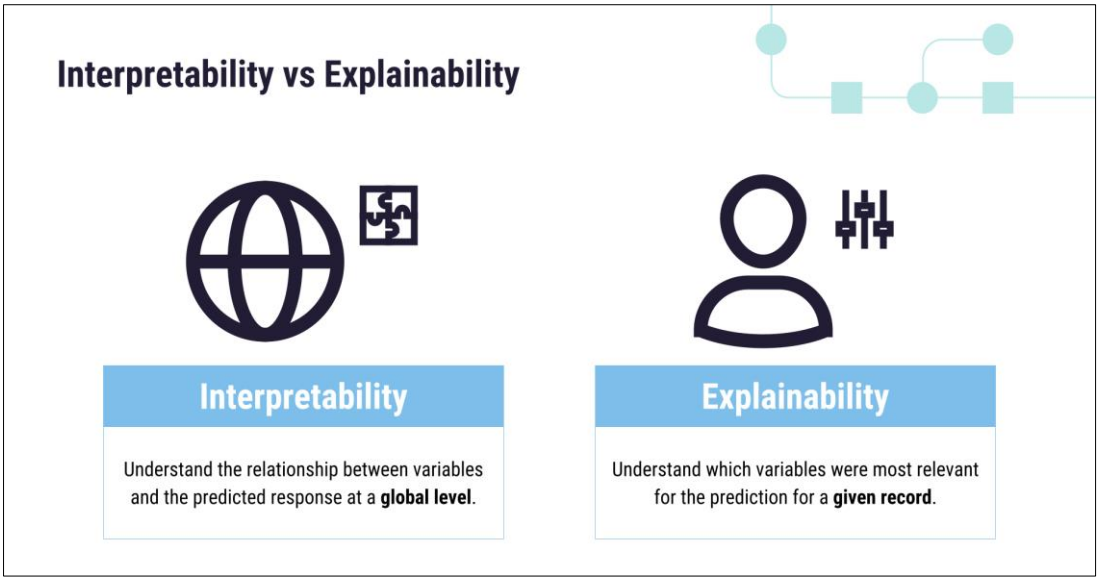

Fig 2: Interpretability vs Explainability

4.1. Operational Workflow

A typical workflow under this framework is as follows:

1. **Transaction Input:** Real-time or batch transactions are fed into the predictive layer.
2. **Risk Scoring:** ML models generate risk scores for each transaction.
3. **Explanation Generation:** The explainability layer computes feature contributions and interprets the output in human-readable terms.
4. **Decision Translation:** Outputs are structured into compliance-ready reports and risk escalation recommendations.
5. **Human Oversight:** Compliance officers review flagged transactions, validate explanations, adjust thresholds if necessary, and document the decision.
6. **Audit and Reporting:** All scores, explanations, and human decisions are stored in an auditable repository to satisfy regulatory requirements.

This workflow ensures that machine learning predictions are fully operationalized for regulatory compliance without sacrificing model transparency or predictive power.

4.2. Advantages and Contributions

The proposed framework offers a range of advantages that are significant:

1. **Regulatory Alignment:** Outputs of the ML model are interpreted in the form of insights that are compatible with the guidelines of FATF and Basel.
2. **Improved Accountability:** The use of standardized explanations and human-in-the-loop review guarantees that every decision can be traced back.
3. **Audit-Readiness:** The whole process involves the creation of structured records for all transactions and decisions which subsequently helps in compliance risk reduction.
4. **Decision Support:** The process of prioritizing, escalating, and reporting of suspicious activities becomes more efficient than before.
5. **Operational Integration:** The design which is independent of the model allows financial crime detection systems that are already in place to adopt it with ease.

The framework which is centered on explainability and regulatory usability, not only covers the critical gap in the current literature but also allows high-performing machine learning systems to be trusted, auditable, and actionable in compliance-critical environments.

6. Explainable AI as a Decision-Support Tool

The regulatory-aligned XAI framework transforms machine learning predictions into actionable, auditable, and interpretable insights for compliance officers, auditors, and supervisory authorities. By providing both global and local explanations of model outputs, the framework enables human reviewers to make informed decisions, ensuring accountability, transparency, and regulatory compliance.

7. Case Study: High-Risk Transaction Detection

Consider a high-risk mobile money transaction flagged by a predictive model. Using the explainability layer:

1. **Feature-Level Explanation:** SHAP values indicate that the unusually large transaction amount, a previously unrecognized recipient, and a high frequency of prior transactions were the primary contributors to the elevated risk score (Lundberg & Lee, 2017) [8].
2. **Decision Artifact:** The framework generates a flagging justification report, summarizing the key risk drivers and recommending escalation to a compliance officer.
3. **Human Oversight:** The compliance officer reviews the explanation, validates the recipient account, and determines whether to approve, investigate further, or block the transaction.

This approach ensures decisions are traceable, auditable, and aligned with regulatory expectations, reducing false positives and operational risk.

8. Case Study: Money Laundering Pattern Recognition

In a scenario involving cryptocurrency transactions, the framework identifies patterns indicative of potential money laundering:

1. **Global Explanation:** Aggregated feature contributions highlight that transactions routed through multiple wallets or originating from high-risk jurisdictions are consistently associated with higher risk.
2. **Local Explanation:** Individual transaction chains are flagged, with detailed reasoning for each feature's contribution to the risk score.
3. **Regulatory Reporting:** The regulatory interpretation layer produces a structured report suitable for supervisory authorities, clearly documenting the rationale behind flagged activities and recommended actions.

By linking model outputs to regulatory categories, the framework enhances compliance readiness, facilitates efficient human review, and supports consistent decision-making across cases.

9. Conclusion and Future Work

This study presents a regulatory-aligned explainable AI framework that bridges machine learning outputs and actionable decision-making in financial crime prevention. By providing interpretable and auditable explanations, the framework enables compliance officers and auditors to make informed decisions, improving transparency, accountability, and operational efficiency. Case studies demonstrate how high-risk transactions and potential money laundering patterns can be identified, explained, and validated effectively, supporting both regulatory compliance and human oversight.

Future work can explore real-time implementation of the framework, integration of multimodal data sources, assessment of fairness and bias, and adaptation across different regulatory jurisdictions. These extensions will further enhance the practical applicability and scalability of explainable AI in high-stakes financial environments.

10. References

1. Basel Committee on Banking Supervision. Sound management of risks related to money laundering and financing of terrorism. Basel: Bank for International Settlements; 2015. Available from: https://www.bis.org/bcbs/publ/d405.htm.
2. Bahnsen AC, Aouada D, Ottersten B. Cost-sensitive decision trees for fraud detection. Expert Syst Appl. 2015;42(13):5558-67. doi:10.1016/j.eswa.2015.02.023.
3. Bahnsen AC, Aouada D, Ottersten B. Ensemble of example-dependent cost-sensitive decision trees. arXiv. 2015. Available from: https://arxiv.org/abs/1505.04637.
4. Dal Pozzolo A, Bontempi G, Snoeck M, Waterschoot S. Adversarial drift detection in fraud detection. IEEE Comput Intell Mag. 2018;13(3):38-48. doi:10.1109/MCI.2018.2840734.
5. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv. 2017. Available from: https://arxiv.org/abs/1702.08608.
6. Financial Action Task Force. Risk-based approach for anti-money laundering and counter-terrorist financing. Paris: FATF; 2021. Available from: https://www.fatf-gafi.org/recommendations.html.
7. Guidotti R, Monreale A, Ruggieri S, Turini F, Pedreschi D, Giannotti F. A survey of methods for explaining black box models. ACM Comput Surv. 2019;51(5):93. doi:10.1145/3236009.
8. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765-74. Available from: https://arxiv.org/abs/1705.07874.
9. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International

Conference on Knowledge Discovery and Data Mining; 2016. p. 1135-44. doi:10.1145/2939672.2939778.

10. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206-15. doi:10.1038/s42256-019-0048-x