

RESEARCH

The Human-Centric Algorithm: Synergising AI-Driven Fraud Detection with Ethical Human Oversight to Optimise Sustainable Banking Integrity

Aminu Usman Masanawa*DBA Research Scholar, Rushford Business School*Email: dba1030@rushford.eu

ORCID: 0009-0005-4158-3075

Ahmad Shaheen*College of Management and Technology,**Arab Academy for Science, Technology and Maritime Transport, Egypt*Email: ahmad.yahya.shaheen@aast.edu

ORCID: 0009-0002-5677-1990

ABSTRACT

PURPOSE: The research aims to analyse whether Artificial Intelligence (AI) and Machine Learning (ML) alone are efficient and self-sustainable in detecting financial fraud, or if they need human intervention and supervision for better fraud detection activity in digital finance.

DESIGN/METHODOLOGY: The Kaggle credit card fraud dataset of 284,807 transactions was used to determine the grey zone of suspicious flagging. To analyse the occurrence of false positives, a linear regression model was applied to the first 10,00 anonymised banking transactions.

FINDINGS: Results demonstrated a high degree of false-positive exposure in automated anomaly-detection systems. The regression analysis also yielded a statistically significant model ($F = 4.255$; $p < 0.001$), which underscores the complexity and unpredictability of fraud behaviour in online banking contexts.

CITATION: Masanawa, A. U. and Shaheen, A. (2026): The Human-Centric Algorithm: Synergising AI-Driven Fraud Detection with Ethical Human Oversight to Optimise Sustainable Banking Integrity. *World Journal of Entrepreneurship, Management and Sustainable Development (WJEMSD)*, Vol. 22, No. 6, pp. 509-526.

RECEIVED: 19 December 2025 / **REVISED:** 14 June 2026 / **ACCEPTED:** 14 July 2026 / **PUBLISHED:** 13 September 2026

COPYRIGHT: © 2026 by all the authors of the article above. The article is published as an open access article by WASD under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

ORIGINALITY/VALUE: The current research advances the investigation into the likelihood of false positives detected by AI and whether further human supervision is required to intervene in the detection or validation of suspicious flagging.

RESEARCH LIMITATIONS: The study is limited to the first 1,000 observations from the dataset to ensure manageable statistical interpretation and analytical consistency.

PRACTICAL IMPLICATIONS: The research offers a transformational roadmap for applying human contextual judgement to banking ecosystems, balancing the predictive accuracy of AI with the moral responsibility of human intelligence to create a resilient, inclusive, and sustainable financial future.

KEYWORDS: *Digital Fraud Detection; Suspicious Transactions; Online Banking; Fraud Monitoring; Credit Card Fraud; Digital Finance.*

INTRODUCTION

In 2025, financial security in the global digital economy has been a major challenge, as digital payments by value rose to \$2.0 quadrillion, with 3.6 trillion transactions, among which \$1 trillion were fraudulent (McKinsey and Company, 2025). Such a high volume of transactions led to a transformation in the banking industry guided by Artificial Intelligence (AI), Machine Learning (ML), cloud computing and advanced digital payment systems (Shafi *et al.*, 2023). AI-powered technologies are becoming a necessity in financial institutions, driving automation, enhancing customer experience, and improving fraud detection (Mujtaba and Yuille, 2024). As AI is adopted more widely in e-commerce and online banking, the risks of financial cybercrime and payment fraud are growing. Banks are no exception to the ongoing digital transformation, in which fraud detection has become critical. In the evolving AI environment, it is difficult to maintain the security of the banking institutions and customers' trust, thereby hampering sustainability. However, in today's banking landscape, automated fraud detection has become a crucial component, playing a key role in minimising financial losses and improving cybersecurity (Nwadiokwu, 2025).

Though AI-based fraud detection solutions can detect anomalies through predictive analytics and real-time monitoring (Faisal *et al.*, 2024), they are associated with ethical and operational issues. One of the important problems is false positives, that is, transactions made by a customer that are mistakenly categorised as fraudulent (Bello *et al.*, 2023). All of these issues could result in transaction declines, reputational harm, heightened disappointment, and monetary chaos (Faraji *et al.*, 2024). While AI can help make financial services affordable and efficient, it also poses risks of financial exclusion due to the incompleteness or rigidity of algorithms, particularly for customers

engaged in irregular financial transactions (Qatawneh *et al.*, 2026). In addition, the opacity of algorithms can lead to bias and a lack of transparency in AI-driven decision-making processes, such as loan approvals, thereby deepening concerns about transparency, accountability, and fairness in digital banking governance (Wijethilake, 2025). Not only is this inefficient, but it is also an issue of sustainability, financial equity, and access to banking services.

The research investigates the effectiveness of fraud detectors, the ethical boundaries of automated banking, and recommends a Human-Centric Fraud Governance Model that integrates AI/ML detectors and Ethical Human Oversight (EHO). The study has two research questions (RQs).

RQ1: What is the effectiveness of AI-based fraud detection tools for detecting financial fraud?

RQ2: How does EHO contribute to minimising false positives and rejecting fraudulent transactions?

Theoretical Foundation

The EHO theory is one of the most vital to this research on AI-based fraud detection, as it ensures that automated technology does not operate in ways that erode the moral understanding of human oversight (Kenny, 2026). The integration of the Human-in-the-Loop mechanism will improve efficiency by balancing algorithmic speed and human nuance, thereby preventing ML bias. Using this approach, the likelihood of false positives will be verified and nullified with human validation. Ultimately, EHO provides critical accountability, ensuring that humans retain the final authority over the AI to identify patterns amid the high stakes of financial innovation.

Key Variables and Relationships

This study examines the connection between AI fraud-detection practices and sustainable banking integrity. The dependent variable is fraud classification behaviour, measured by the results of legitimate and fraudulent transactions. V4 (Fraud Behaviour), V10 (Strong Fraud Indicator), V12 (Fraud Anomaly), V14 (Fraud Anomaly), and V17 (Fraud Pattern) reflect behavioural variables after Principal Component Analysis (PCA), and are independent variables correlated with anomaly-detection patterns. EHO Theory is a form of moderating governance that aims to reduce the risk of false positives and improve contextual transaction evaluation.

Contribution of the Study

This study extends fraud detection beyond a purely technology-oriented construct into a wider domain of sustainable financial governance. It deviates from existing studies, which have focused only on the computational efficiency of AI fraud detection, by proposing a human-centric governance perspective to examine the ethical implications of decision-making in digital banking. Four major contributions of this study are as follows:

- Conceptualising fraud detection as a sustainability challenge
- Introduction of grey-zone transactions (referring to ambiguous transactions where legitimate customer behaviour overlaps with this fraud-like anomaly pattern)
- Developing a human-centric algorithmic framework
- Bridging ML fraud detection, ethical AI governance and sustainable banking into a unified analytical framework

LITERATURE REVIEW

Artificial Intelligence in Banking

AI, with the potential to automate processes, has been used to enhance financial security systems in the banking sector. As per Uzoamaka and Ade (2026), with the increasing digitisation of financial services, online transactions, and payment systems, AI is proving to be a valuable support for customer analytics, risk management, and fraud detection. In addition, Shah and Kumar (2025) also found that AI-based systems have substantially improved operational efficiency, minimised human involvement, and enhanced the precision of anomaly detection in intricate financial landscapes. However, Fonkem (2025) suggests that the rise of AI in banking is most often linked to the rise of the sophistication of cybercrime and digital payment fraud. This indicates that the existing fraud monitoring architecture is not very effective at addressing new fraud trends, as it relies on a set of pre-programmed behavioural conditions. Predictive analytics, therefore, has become one of the major weapons in the arsenal of fraud prevention, with constantly evolving AI and ML systems, which are more adaptable and flexible than conventional anti-fraud systems.

ML Fraud Detection Models

ML models are very effective in fraud detection, as they can process highly imbalanced financial data and uncover hidden behavioural anomalies, which is why they are becoming increasingly used. Leevy *et al.* (2023) recommend that logistic regression is widely used in financial fraud analysis for its ease of interpretation and effectiveness in binary classification. Logistic Regression (LR) has been shown to be effective for structured transactions, but is less predictive when transactions become highly nonlinear and complex in their behaviour. However, Ayorinde (2025) notes that

the linear regression model provides more interpretable estimates of variable influence; it is more suitable for examining fraud-related behavioural patterns and false-positive risk within a human-centric framework.

Ethical Challenges in Algorithmic Banking

Ridzuan *et al.* (2024) have highlighted the ethical issues associated with computerised banking services. False positives are among the largest issues, as they occur when a legitimate customer is incorrectly labelled as a scam. When the algorithm is oversensitive, it leads to customers having transactions declined, disrupting financial services, harming their reputation, and eventually weakening trust in financial institutions. False positives have been found to have economic and psychological effects that are increasingly attributable not only to operational inefficiencies but also to socio-economic exclusion.

Furthermore, Chaudhary (2024) argues that AI systems are black-box decision-making systems, in which users (and even institutions) do not know how they make decisions. This obscurity compromises the financial governance system's transparency and accountability. In addition, algorithmic bias is a concern, as ML systems are based on past transactions and can inadvertently learn discriminatory patterns. It also implies a risk of excessive screening and/or limitations on transactions for vulnerable customer groups and those with irregular financial behaviour. According to Wijethilake (2025), these issues emphasise the importance of AI regulation as both a technological and an ethical and sustainability issue in the context of digital banking.

Human-in-the-Loop Systems

As a result of ethical issues, Nguyen (2025) has suggested that, in recent years, there has been a growing number of studies promoting Human-in-the-Loop approaches that utilise human input to obtain context-driven judgements, thereby enhancing the efficiency of AI in banking operations. According to Kandikatla and Radeljic (2025), Human-in-the-Loop systems provide a step of EHO before any banking action for ambiguous and/or high-risk transactions. Institutions can continue to operate at the same efficiency and scale as AI does, without unfairly rejecting transactions or becoming inflexible in their algorithms. In situations where transactions behave in an unusual yet legitimate manner, automated systems may struggle to distinguish them from fraud, and contextual interpretation is particularly important. Kumar (2024) highlights that AI-human interaction has been shown to improve organisational accountability, increase customer trust, and reduce the risks of financial exclusion. That is why more and more financial decision-making systems are accepting Human-in-the-Loop governance as a viable alternative to full autonomy.

Sustainable Banking Integrity

Sustainable banking integrity extends beyond profitability and security infrastructure, also focusing on accountability, financial inclusion, and AI governance (Vasudevan *et al.*, 2026; Kothandapani, 2025). There is a growing expectation for financial institutions to further advance technology and enhance transparency to support inclusivity in financial services. Responsible AI governance focuses on the efficiency of prediction, fairness and socio-economic inclusion. Further, Shehadeh (2025) says that sustainable banking frameworks recognise that algorithmic accountability is part of the broader challenge of financial accessibility and is one of the pillars of the resilience of future sustainable financial institutions. The existing research on AI fraud detection has largely focused on predictive accuracy, model optimisation, and computational efficiency.

Few studies have examined the connections between fraud detection systems and ethical governance, sustainable banking and/or banking integrity. Moreover, the social and economic impacts of false-positive financial exclusion have not been studied in depth in prior literature.

RESEARCH METHODS

This study adopts a quantitative experimental design and a human-centric evaluation approach for analysing the relationship between AI in Fraud Detection and Ethical Banking Governance. The dataset used is the Kaggle Credit Card Fraud Dataset, which comprises 284,807 anonymised transactions, highly imbalanced fraud classification, and PCA-transformed variables (V4, V10, V12, V14, and V17) (Kaggle, 2025).

Kaggle dataset link: <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>

The methodological framework involves a linear regression model to evaluate the efficiency of predictive fraud detection and false-positive behaviour (Yaseen and Al-Amarneh, 2025). Further, the data were scrutinised and classified using convenience sampling into the first 1,000 transactions, based on PCA-transformed variables (X1-X5), for regression analysis.

The study particularly highlights the use of PCA to identify grey-zone transaction clusters, explaining that once the algorithms are hardcoded, they will treat some transactions as outliers and then reject them.

A Human-in-the-Loop design is combined with automated anomaly detection and ethical human decision-making to cut down on financial exclusion and algorithmic redlining. The main areas of research focus are: the accuracy of fraud classification, the cost of false-positive classifications, sensitivity to anomalies and the socio-economic consequences of banking decisions made using automated information.

Within the dataset, data limitations were addressed to meet the research purpose. Thereafter, the data was analysed to identify a clear pattern before distinguishing between suspicious and normal

transactions. If the fraud behaviour time is more than 2000, a strong indicator is less than -5, and a fraud anomaly is less than -5, the parameters indicate a suspicious behaviour pattern; otherwise, the transaction is classified as normal.

The complete Excel solution can be viewed at:

<https://drive.google.com/drive/folders/15rFJpTf4Yw1vsKZ6SeXVN2zXICgJdjD8?usp=sharing>

ETHICAL STATEMENT

The study utilised publicly available secondary data from the Kaggle repository (<https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>); therefore, there is no concern regarding personal identifiers or privacy issues. The Kaggle Credit Card Fraud Dataset is a well-known and widely accepted dataset in fraud analytics, and it can serve as a benchmark for empirical investigation. Its open licensing policy, anonymised data, and reproducibility attributes make it transparent, credible, and ethically suitable to research.

RESULTS AND DISCUSSION

Quantitative Component

The quantitative analysis used the Kaggle Credit Card Fraud Dataset to investigate the correlations between the variables related to AI fraud detection and bank transaction behaviour. This research employed distribution analysis, transaction amount analysis, grey-zone transaction analysis, false-positive evaluation, and regression-based predictive analysis to investigate the effectiveness of automated fraud detection systems and the ethical constraints of these systems in sustainable digital banking settings.

Distribution Analysis: This data comprised anonymised banking transactions from 284,807 people. Analysis of the transactions revealed 142,423 fraudulent and 142,384 valid transactions. The results demonstrated a highly sensitive fraud detection system, with suspicious behavioural indicators widely present across both genuine and fraudulent observations.

The classification of fraud was represented by:

Class = {0, Legitimate Transaction and 1, Fraudulent Transaction}

There was also considerable overlap between the behaviour of the fraudulent transactions and that of legitimate banking transactions for the PCA-transformed variables V10, V12, and V14. It indicates that fraud-detection AI can be used in different ways, with no clear sign that the behaviour is being perpetrated solely for financial gain.

The study thus supports the trend towards significant institutional dependence on AI-powered systems capable of identifying behavioural anomalies in the vast volume of digital banking data. However, the combination of legitimate and fraudulent transaction activity poses the risk that algorithms may incorrectly classify transactions, leading to false positives.

Transaction Amount Analysis: A transaction amount analysis was used to determine if the amount of a financial transaction had a significant effect on fraud classification behaviour. The results showed that the total amount of transactions committed by fraudsters was 12,653,519.54, and the total amount of transactions committed by authentic users was 12,509,070.47, both of which were similar in size.

The total transaction amount can be represented as:

$$TA = \sum_{i=1}^n i$$

Where:

- TA = Total transaction amount
- Amount i = Individual transaction value
- n = Total number of transactions

The difference in transaction type is relatively small, suggesting that transaction value is not a good discriminator between genuine and fraudulent financial activity (Pratama and Wahid, 2025). It is a significant finding, since the AI-powered fraud detection systems often focus on unusual transaction volumes when detecting anomalies.

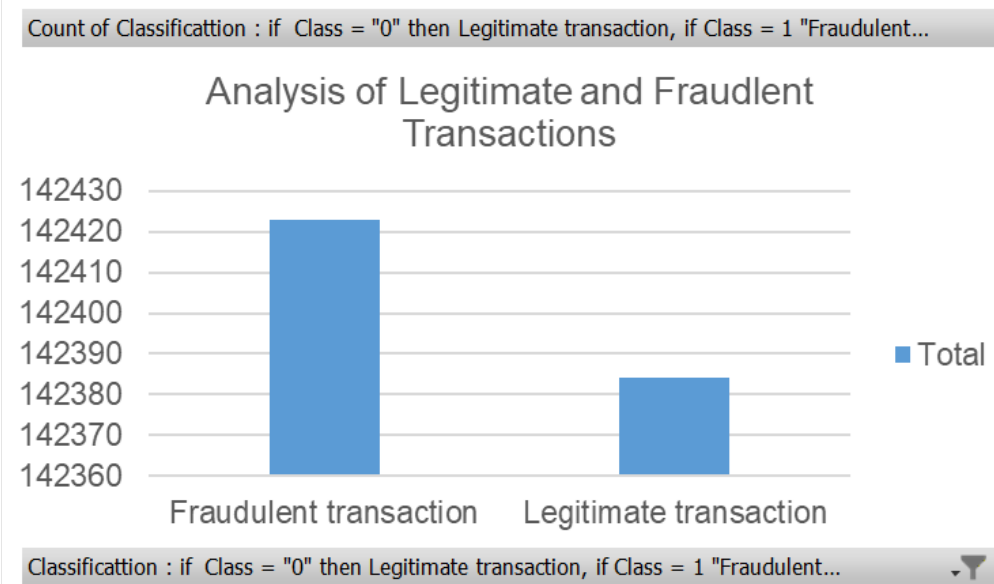


Figure 1: Determination of Legitimate and Fraudulent Transactions

Source: Developed by authors using a pivot table (Editable format of image can be viewed at: <https://drive.google.com/drive/folders/15rFJpTf4Yw1vsKZ6SeXVN2zXICgJdjD8?usp=sharing>)

The results shown in Figure 1 accordingly indicate that financial transaction values cannot be the only basis for fraud detection; there should be mechanisms for behaviour interpretation and contextual evaluation.

Grey-Zone Transaction Analysis: Grey-zone transactions are “behaviourally ambiguous” in some way or contain elements that could be legitimate and elements that could be AI-generated fraud.

AIFlag= Suspicious if Amount $\geq$ Threshold, otherwise Normal

The analysis in Table 1 shows that transactions classified as ‘Suspicious’ in the AI system had a total transaction value of 24,994,248.97. Amongst them, 12,557,093.89 transactions were fraudulent and 12,437,155.08 transactions were true. The narrow distinction between these sorts of transactions strongly shows that important and legitimate banking transactions might be mistaken for fraudulent transactions by AI systems.

Table 1: Grey-Zone Transaction Analysis

Row Labels	Fraudulent transaction	Legitimate transaction	Grand Total
Normal	96425.65	71915.39	168341.04
Suspicious	12557093.89	12437155.08	24994248.97
Grand Total	12653519.54	12509070.47	25162590.01

Source: Processed by authors from Kaggle Dataset using pivot table and charts (Kaggle, 2025)

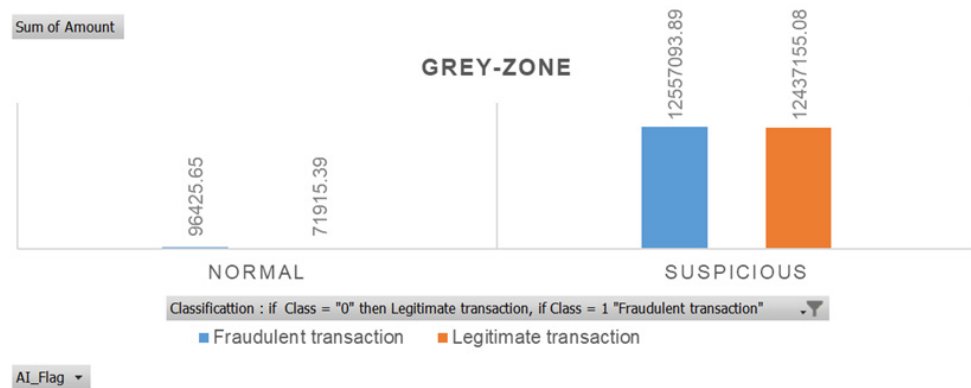


Figure 2: Comprehension of Fraudulent Transaction and Legitimate Transaction Flagged as AI Suspicious

Source: Developed by authors from Kaggle Dataset using pivot table and charts (Kaggle, 2025). (Editable format of image can be viewed at:

<https://drive.google.com/drive/folders/15rFJpTf4Yw1vsKZ6SeXVN2zXlCgJdjD8?usp=sharing>)

The outcome supports the existence of suspicious/legitimate transaction clusters with behavioural overlap. Figure 2 shows that of the total normal transactions (168,341.04), 96,425.65 were fraudulent, and 71,915.39 were legitimate. This highlights that AI systems that rely solely on anomaly-detection logic cannot fully distinguish between malicious and legitimate behaviour.

Given the high volume of legitimate suspicious transactions, Human-Centric Human-in-the-Loop governance systems are even more important in the banking industry. Uncontrolled AI systems can yield substantial false-positive rates, leading to declined transactions, delayed financial access, reputational issues, and customer dissatisfaction.

False Positive Risk Analysis: The false-positive risk analysis considers the ethical aspects of AI in fraud detection. In automated anomaly detection, false positives occur when a legitimate customer's transaction is incorrectly identified as fraudulent.

The False Positive Rate (FPR) can be defined as:

$$\text{FPR} = \text{False Positives} / (\text{False Positive} + \text{True Negative})$$

$$\text{FPR} = 12437155.08 / (12437155.08 + 71915.39)$$

$$= 99.43\%$$

The results showed that 12,437,155.08 suspicious transactions were incorrectly classified, indicating that many legitimate transactions were included in the AI classification system (99.43%). The findings suggest that algorithmic rigidity and customer disruption may be unintended consequences of AI systems, with a focus on behavioural abnormality. However, among the AI-marked suspicious transactions, 49.76% are false negatives [$12437155.08 / (12437155.08 + 12557093.89)$]. The results for false negatives among detected or marked significant transactions show a significant need for EHO to minimise algorithmic misclassifications and financial exclusion.

Technical prediction error is not the only impact of false positives in a sustainability context. Suspicious customers who are actually legitimate can suffer financial disruption, reduced access to accounts, slow transactions, and loss of confidence in banking institutions.

The results of this study confirm the need to incorporate the EHO into fraud detection systems. AI systems are not as capable as human analysts at applying contextual reasoning. Human oversight reduces the risk of mistakes in rejecting transactions to prevent algorithmic exclusion based on an understanding of behaviour and context.

Regression-Based Predictive Analysis: The relationship between selected transaction variables and fraud classification behaviour was investigated using regression analysis on the first 1,000 transactions in the dataset. Five behavioural variables related to anomaly-detection patterns were analysed.

The regression model is represented as:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + \epsilon$$

Where:

- Y = Fraud classification behaviour
- a = Intercept
- b = Regression Coefficients
- X1–X5 = Selected PCA behavioural variables
- X1: Fraud Behaviour
- X2: Strong Fraud Indicator

- X3: Fraud Anomaly
- X4: Fraud Anomaly
- X5: Fraud Pattern
- ϵ = Error term

Regression analysis was performed using the Excel Data Analysis Toolpack to avoid manual calculation errors. The regression results yielded a Multiple R of 0.1447 and an R Square of 0.0209, indicating that the selected variables account for 2.1% of fraud classification behaviour.

Although the data exhibited somewhat complex behaviour, the Analysis of Variance (ANOVA) yielded an F-statistic of 4.255 and a significance level of 0.000780, indicating that the regression model remained statistically significant.

The coefficient analysis also revealed that the transaction variables exhibit different behavioural relationships with the fraud classification results. One variable was statistically significant, with a positive coefficient (15.118) and a p-value of 0.0017, and was correlated with suspicious transaction behaviour. Another variable that resulted in a statistically significant negative coefficient was the p-value of -22.400, with a p-value of 0.00079, indicating a negative behavioural effect on fraud prediction. Table 2 shows the complete regression analysis.

Table 2: Regression Analysis for the First 1000 Transaction Data

SUMMARY OUTPUT										
Regression Statistics										
Multiple R	0.144766									
R Square	0.020957									
Adjusted R-Square	0.016033									
Standard Error	185.7574									
Observations	1000									
ANOVA										
	Df	SS	MS	F	Significance F					
Regression	5	734195.4	146839.1	4.255489	0.000781					
Residual	994	34298771	34505.81							
Total	999	35032967								
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%		
Intercept	66.59974	6.729027	9.89738	4.32E-22	53.39501	79.80447	53.39501	79.80447		
1.378155224	15.11808	4.829008	3.13068	0.001795	5.64186	24.5943	5.64186	24.5943		
0.090794172	-22.4004	6.6538	-3.36655	0.00079	-35.4575	-9.34325	-35.4575	-9.34325		
-0.617800856	-14.151	9.341682	-1.51482	0.130136	-32.4826	4.180728	-32.4826	4.180728		
-0.311169354	1.627408	8.072785	0.201592	0.840277	-14.2142	17.46907	-14.2142	17.46907		
0.207971242	-10.7401	8.71872	-1.23184	0.218301	-27.8493	6.369151	-27.8493	6.369151		

Source: Measured by authors

Abbreviations used in Table 2: Df = degrees of freedom; SS= Sum of Squares; MS= Mean Square; F= F-statistics

The statistical significance condition is expressed as $p < 0.05$. The complete Excel solution can be viewed at:

<https://drive.google.com/drive/folders/15rFJpTf4Yw1vsKZ6SeXVN2zXICgJdjD8?usp=sharing>

Prediction inconsistencies between observations were also identified using residual analysis, which revealed grey-zone behavioural clusters in the data. Based on the results, it can be concluded that AI systems play a crucial role in fraud classification, but they are not sufficient to guarantee fairness and accuracy without proper human context (RQ1 answered).

The analysis was conducted on the first 1000 transactions, primarily due to time and tool constraints. The authors also observed that using a larger dataset would not significantly alter the results, which was one of the primary reasons for choosing a smaller dataset. It might have affected generalisability to some extent, but the research provides a pathway for future studies to conduct a more comprehensive, in-depth analysis.

Qualitative Interpretation

The interpretation examined the ethical and sustainability aspects of digital banking governance and the relevance of AI-based fraud-detection systems. The results indicate that AI enhances the effectiveness of operational fraud detection through real-time anomaly monitoring and predictive behavioural analysis. However, over-reliance on automated decision-making may increase the risk of false-positive classification and financial exclusion. The inclusion of a Human-Centric Algorithm and EHO improves efficiency in the digital banking landscape by enhancing accountability, transparency, customer trust, and inclusivity, thereby reducing reliance on AI-based detection (RQ2 answered).

Due to computational limitations associated with Excel-based regression analysis, the study utilised only the first 1,000 observations from the dataset to ensure manageable statistical interpretation and analytical consistency. The selected sample remained sufficient to identify behavioural patterns, false-positive risks, and grey-zone transaction dynamics relevant to the study's Human-Centric AI framework.

Figure 3 indicates that improved Human-in-the-Loop integration has the potential to enhance ethical banking governance by ensuring equitable financial outcomes, consumer trust, and inclusive digital banking practices, as well as increasing the efficiency of predictive AI capabilities. This framework emerged from a deductive-inductive approach that integrates literature findings, EHO theory and empirical dataset analysis. The conceptual analysis showcases AI fraud detection, Human-in-the-loop, and ethical banking governance, whereas empirical analysis guides the framework from grey-zone transaction analysis, false-positive risk evaluation, and regression analysis to understand behavioural overlap between legitimate and fraudulent transactions. As a result, the Human-Centric AI Fraud Governance Framework improves contextual judgement, fairness, and sustainable banking integrity.

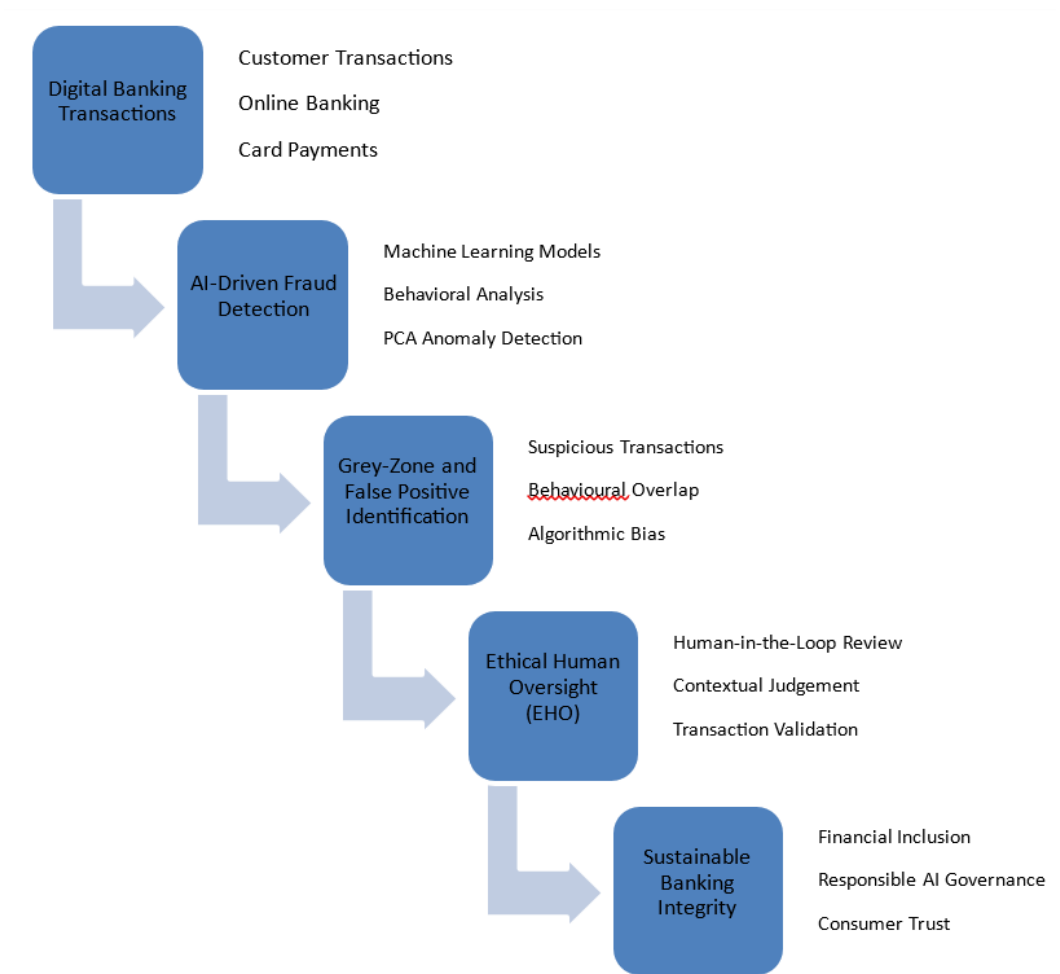


Figure 3: Human-Centric AI Fraud Governance Framework

Source: Constructed by the authors

CONCLUSIONS

The study focused on identifying the role of AI in detecting digital fraud in digital banking, and critically explored the ethical challenges of an all-digital financial governance. The results confirmed that ML-based fraud detection systems can identify anomalies within sub-seconds and monitor transactions, enhancing operational effectiveness. The regression-based analysis and interpretation of grey-zone transactions further verified that AI systems could successfully identify suspicious

transaction patterns in extremely imbalanced datasets. However, the results also emphasised the need to complement automated fraud detection systems with human intervention, including EHO and behavioural interpretation. The Human-Centric AI Fraud Governance Framework could be expanded by adding Explainable Artificial Intelligence (XAI) and real-time ethical auditing systems, thereby introducing transparency, interpretability, and institutional accountability into AI-driven banking governance.

REFERENCES

- Ayorinde, A.S. (2025): Explainable Deep Learning Models for Detecting Sophisticated Cyber-Enabled Financial Fraud Across Multi-Layered FinTech Infrastructure. *International Journal of Cybersecurity and Digital Forensics*, 5(3), pp.241-263. Available at <https://doi.org/10.30574/ijdsra.2025.16.1.2134>
- Bello, O.A., Ogundipe, A., Mohammed, D., Adebola, F. and Alonge, O.A. (2023): AI-driven approaches for real-time fraud detection in US financial transactions: Challenges and opportunities. *European Journal of Computer Science and Information Technology*, 11(6), pp.84-102. Available at: https://www.researchgate.net/publication/381548442_AI-Driven_Approaches_for_Real-Time_Fraud_Detection_in_US_Financial_Transactions_Challenges_and_Opportunities
- Chaudhary, G. (2024): Unveiling the black box: Bringing algorithmic transparency to AI. *Masaryk University Journal of Law and Technology*, 18(1), pp.93-122. Available at: <https://www.cceol.com/search/article-detail?id=1287954>
- Faisal, N.A., Nahar, J., Sultana, N. and Mintoo, A. A. (2024): Fraud detection in banking leveraging AI to identify and prevent fraudulent activities in real-time. *Journal of Machine Learning, Data Engineering and Data Science*, 1(01), pp.181-197. Available at: <https://doi.org/10.70008/jmldeds.v1i01.53>
- Faraji, M.R., Shikder, F., Hasan, M.H., Islam, M.M. and Akter, U.K. (2024): Examining the role of artificial intelligence in cyber security (CS): A systematic review for preventing prospective solutions in financial transactions. *International Journal*, 5(10), pp.4766-4782. Available at: <https://doi.org/10.61707/7rfyma13>
- Fonkem, B.N. (2025): AI-Powered Risk Scoring Models for Real-Time Fraud Detection in Digital Banking Ecosystems. *Journal of Computational Analysis and Applications*, 34(11), pp.349-371. Available at: <https://eudoxuspress.com/index.php/pub/article/view/4135>
- Kaggle (2025): *Credit Card Fraud Detection*. Available at: <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud> (Accessed: 26 May 2026)
- Kandikatla, L. and Radeljic, B. (2025): AI and Human Oversight: A Risk-Based Framework for Alignment. *arXiv preprint arXiv:2510.09090*. Available at: <https://doi.org/10.48550/arXiv.2510.09090>
- Kenny, K.R.S. (2026): The Monetary Flow of a Coming Paradigm. *Available at SSRN 6604018*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6604018

- Kothandapani, H.P. (2025): Ai-driven regulatory compliance: Transforming financial oversight through large language models and automation. *Emerging Science Research*, 12(1), pp.12-24. Available at: <https://doi.org/10.6084/m9.figshare.28251608>
- Kumar, P. (2024): AI-Powered Fraud Prevention in Digital Payment Ecosystems: Leveraging Machine Learning for Real-Time Anomaly Detection and Risk Mitigation. *Journal of Information Systems Engineering and Management*, 9(4). Available at: <https://doi.org/10.52783/jisem.v9i4.102>
- Leevy, J.L., Hancock, J., Khoshgoftaar, T.M. and Abdollah Zadeh, A. (2023): Investigating the effectiveness of one-class and binary classification for fraud detection. *Journal of Big Data*, 10(1), p.157. Available at: <https://doi.org/10.1186/s40537-023-00825-1>
- McKinsey & Company. (2025): *The 2025 McKinsey Global Payments Report: Competing systems, contested outcomes*. McKinsey & Company. Available at: <https://www.mckinsey.com/industries/financial-services/our-insights/global-payments-report> (Accessed: 26 May 2026)
- Mujtaba, N. and Yuille, A. (2024): AI-Powered Financial Services: Enhancing Fraud Detection and Risk Assessment with Predictive Analytics. *Unpublished. DOI, 10*. Available at: <https://doi.org/10.13140/RG.2.2.23580.09603>
- Nguyen, M. (2025): *Enhancing Decision-Making in Anti-Money Laundering with Explainable AI (XAI) Visualization: Aligning Human and Machine Learning Decision Making Processes* (Master's thesis). Available at: <https://studenttheses.uu.nl/handle/20.500.12932/49340>
- Nwadiokwu, O.T. (2025): Advanced AI-Driven Threat Intelligence Systems for Proactive Detection and Mitigation of Cyber Fraud in Financial Institutions. *International Journal of Computer Applications Technology and Research*, 14(2), p.214-230. Available at: <https://doi.org/10.7753/IJCATR1402.1015>
- Pratama, S.F. and Wahid, A.M.A. (2025): Fraudulent transaction detection in online systems using random forest and gradient boosting. *Journal of Cyber Law*, 1(1), pp.88-115. Available at: <https://doi.org/10.63913/jcl.v1i1.5>
- Qatawneh, A., Alfityani, A., Abuorabi, Y. K. and El-Khateeb, A. (2026): Role of Artificial Intelligence-Based Automation and Accuracy of Financial Reporting in AIS in Increasing Sustainability in Corporate Practices. *World Journal of Science, Technology and Sustainable Development (WJSTSD)*, 21(1/2/3), pp. 163-179. Available at: <https://doi.org/10.47556/J.WJSTSD.21.1-2-3.2026.10>
- Ridzuan, N.N., Masri, M., Anshari, M., Fitriyani, N.L. and Syafrudin, M. (2024): AI in the financial sector: The line between innovation, regulation and ethical responsibility. *Information*, 15(8), p.432. Available at: <https://doi.org/10.3390/info15080432>
- Shafi, S.B., Mahmud, I., Rahul, D.D., Aqif, H., Tariquzzaman, M.D., Osman, F.Y., Akter, S. and Sozib, H.M. (2023): Artificial intelligence in banking: Transforming bank management and customer experience. *Cuestiones de Fisioterapia*, 52(3), pp.337-348. Available at: <https://doi.org/10.48047/kydnzm47>

- Shah, K. and Kumar, L. (2025): AI-Driven Anomaly Detection in Financial Consolidation: Enhancing Accuracy and Compliance. *International Journal of Enhanced Research in Science Technology & Engineering*, 14(4), pp.2319–7463. Available at: https://www.researchgate.net/publication/397427201_AI-Driven_Anomaly_Detection_in_Financial_Consolidation_Enhancing_Accuracy_and_Compliance
- Shehadeh, M. (2025): Exploring cutting-edge algorithms in FinTech: Leveraging machine learning and artificial intelligence for achieving sustainability. Available at: <https://doi.org/10.1108/978-1-83549-608-420251011>
- Uzoamaka, CHIKELUBA and Ade, B.S. (2026): Artificial Intelligence, Automation, and Data-Driven Decision-Making in Enhancing Banking Efficiency and Business Process Optimisation in Nigeria. *International Journal of Scientific Research and Management (IJSRM)*, 14(01), pp.10312-10328. Available at: <https://doi.org/10.18535/ijssrm/v14i01.em13>
- Vasudevan, A., ŞİMŞEK, S. A and Jamil, N. A. (2026): Digital Financial Services and Financial Inclusion: The effect of e-levy on the Adoption of Mobile Money in a Sustainable Economy. *World Journal of Entrepreneurship, Management and Sustainable Development (WJEMSD)*, 22(4), pp. 315-331. Available at: <https://doi.org/10.47556/WJEMSD.22.4.2026.2>
- Wijethilake, C. (2025): Ethical challenges in utilizing Artificial Intelligence in banking operations: A systematic review. Available at: https://jyx.jyu.fi/jyx/Record/jyx_123456789_102739
- Yaseen, H. and Al-Amarneh, A.A. (2025): Adoption of artificial intelligence-driven fraud detection in banking: The role of trust, transparency, and fairness perception in financial institutions in the United Arab Emirates and Qatar. *Journal of Risk and Financial Management*, 18(4), p.217. Available at: <https://doi.org/10.3390/jrfm18040217>

BIOGRAPHY



Aminu Usman Masanawa is a banking professional with over 12 years of experience in development financing. Aminu earned a Bachelor's degree in Electrical and Electronics Engineering from the University of Bradford and an MBA from the University of South Wales. Over the years, Aminu has worked with various departments in the bank spanning across Operations, Finance and Micro, Small and Medium-sized Enterprises (MSMEs). Outside of professional life, he enjoys reading, travelling, and volunteering in community development programmes. He finds joy in helping businesses connect meaningfully with their desired customers.



Dr Ahmad Shaheen is an Assistant Professor of Marketing and Business Administration within the Marketing Department at the College of Management and Technology, Arab Academy for Science, Technology and Maritime Transport, Egypt. He holds a PhD in Marketing from an AACSB-accredited institution and possesses over a decade of extensive experience in both academia and professional consulting. His expertise encompasses AI-enabled marketing, influencer strategy, and consumer behaviour. Dr Shaheen is an active researcher whose work is published in Scopus-indexed journals. He is dedicated to teaching at the undergraduate, MBA, and doctoral levels while simultaneously providing strategic guidance to SMEs on data-driven digital transformation.

