ABNUS JOURNAL
Your Research Partner

Corelytics: Insights into Business & Accounting Systems (Corelytics)
E-ISSN: 3164-3337 | Vol. 1, No. 4 (2026) | pp. 213-224

# The Role and Challenges of Artificial Intelligence in Mitigating Fraud Risk in Digital Wallet Use

**Khalishah Nuriella Ramadhani[1*], Satryo[2]**

*Universitas Padjadjaran, Jatinangor, Indonesia[1,2]*

***Corresponding author:** khalishahnr04@gmail.com |

Received: 30 Agustus 2026 | Revised: 14 September 2026 | Revised: 10 October 2026 | Published: 5 November 2026

**Abstract**

**Purpose:** This study examines the effectiveness of Artificial Intelligence (AI) in detecting fraud in digital financial transactions compared with manual methods from an internal control perspective, identifies its limitations, particularly false positives and machine learning blind spots, and explains the role of system calibration and human intervention.

**Research Methodology:** The study used a literature review of scientific articles, professional reports, and regulations on AI based fraud detection in digital financial transactions, including digital wallets. The literature was screened for relevance to the research questions and synthesized thematically.

**Results:** AI outperforms manual and rule based methods in speed, scale, and detection accuracy because it can analyze entire transaction populations in real time. However, AI has structural limitations in the form of false positives and blind spots toward new fraud patterns caused by concept drift, data bias, adversarial behavior, and limited labeled data. Regular calibration and human professional judgment remain necessary to verify alerts and make final decisions, in line with regulations that require human involvement in automated decisions.

**Conclusions:** Collaboration between AI and humans, with AI as the initial screening tool and humans as the final decision makers, is the most effective model for fraud detection.

**Limitations:** The study is a literature review and has not been tested empirically with real transaction data.

**Contributions:** The study integrates evidence on AI effectiveness and limitations into a human in the loop framework and offers guidance for financial institutions, regulators, and researchers.

***Keywords:*** *Artificial Intelligence, Digital Wallet, False Positive, Fraud Detection, Human Oversight*

**How to Cite:** Ramadhani, K. N., & Satryo. (2026). The Role and Challenges of Artificial Intelligence in Mitigating Fraud Risk in Digital Wallet Use. *Corelytics: Insights into Business & Accounting Systems (Corelytics)*, 1(4), 213-224.

## 1. Introduction

Technology plays an important role in many areas of life and has brought major changes to society, including the economy. The development of digital technology has transformed the landscape of financial transactions, as economic activities shift from cash to digital systems. Digital wallets, payments based on quick response codes, and online interbank transfers have become an inseparable part of daily life, reflecting a shift in consumer behavior toward a fully digital financial ecosystem. The fintech revolution has changed how financial services are designed, delivered, and regulated (Gomber, Kauffman, Parker, & Weber, 2018), and digital finance can expand financial inclusion when users trust and can safely access digital channels (Ozili, 2018). Artificial intelligence also contributes to digital financial inclusion by enabling faster and more personalized services (Mhlanga, 2020).

© 2026 Abnus | Open Access — CC BY 4.0 | abnus.co | DOI:10.52909/corelytics.v1i4.467 

Digital transactions, however, carry consequences that cannot be ignored. The increasing volume and complexity of digital financial transactions directly open wider gaps for fraud. Schemes such as credit card fraud, account hacking, identity theft, and money laundering have become more common amid massive and fast transaction flows. In 2024, the Global Anti Scam Alliance estimated that global losses to scams exceeded US$1 trillion in a single year, which shows that financial stability and public trust are threatened by the prevalence of such cases (Global & Feedzai, 2024). Research on mobile payments shows that trust and perceived risk are among the strongest determinants of whether consumers continue to use digital payment services (Patil, Tamilmani, Rana, & Raghavan, 2020), so fraud directly threatens the sustainability of the digital payment ecosystem.

Facing these challenges, financial institutions are required to have internal control systems that are more adaptive and responsive. Manual fraud detection methods are considered limited in terms of speed and capacity to handle the volume and complexity of transaction data. These limitations have encouraged financial institutions and digital payment providers to adopt Artificial Intelligence (AI) as the front line in mitigating fraud risk efficiently. Reviews of financial fraud detection confirm that data driven and intelligent methods have become essential because static rules cannot keep pace with evolving fraud schemes (West, & Bhattacharya, 2016; (Hilal, Gadsden, & Yawney, 2022)).

AI offers analytical capabilities far beyond human limits, and systems based on machine learning and deep learning can achieve very high detection rates (Ali, Othman, Eisa, Al-Dhaqm, Nasser, Elhassan, Elshafie, & Saif, 2022). Nevertheless, AI is not infallible and still has weaknesses. Challenges such as a high number of false positives and the existence of machine learning blind spots require further attention (Dal, Boracchi, Caelen, Alippi, & Bontempi, 2018). These conditions show that although AI contributes greatly to the efficiency and accuracy of fraud detection, it is not an error free solution. Calibration of algorithms and human analytical intervention are still needed. Research on management and AI also shows that the best outcomes often arise when AI augments human decision making rather than fully replacing it (Jarrahi, 2018).

It is therefore important to examine in more depth the role of AI in detecting fraud and the challenges that accompany it, in order to provide a more comprehensive picture for financial institutions and regulators in designing more accountable risk mitigation systems. This study addresses three research questions. The first concerns the effectiveness and accuracy of AI in detecting fraud compared with manual methods from an internal control perspective. The second concerns the limitations and gaps that arise from the implementation of AI in fraud detection, particularly false positives and machine learning blind spots. The third concerns the role of calibration and human intervention in overcoming the challenges of AI. The study contributes theoretically to the fields of accounting, auditing, and internal control by integrating evidence on the effectiveness and limitations of AI, and practically by offering guidance for financial institutions, regulators, academics, and the public.

## 2. Literature Review and Hypothesis/es Development

### *2.1 Fraud*

Fraud is an intentional act or omission committed to achieve a particular objective, such as manipulating or presenting false financial information to other parties, by individuals or groups inside or outside an organization, with the aim of obtaining direct or indirect gains that ultimately harm others. According to the Association of Certified Fraud Examiners, fraud includes asset misappropriation, corruption, and financial statement fraud, and these three types are interrelated and are more likely to occur in organizations with weak internal control systems (Association of Certified Fraud Examiners, 2024). Theories of fraud have evolved from the fraud triangle of pressure, opportunity, and rationalization to models that include capability, ego, and collusion, because the frequency and sophistication of fraud have increased (Vousinas, 2019). In the context of digital financial transactions, fraud appears in various schemes, including identity theft, card not present fraud, social engineering, the creation of fictitious identities, and money laundering through complex digital transactions.

### *2.2 Digital Financial Transactions*

Digital financial transactions refer to any transfer of economic value carried out through electronic means without directly involving physical money. Digital wallets, interbank transfers through mobile banking, payments using the Quick Response Code Indonesian Standard (QRIS), online credit and debit card transactions, peer to peer lending, and cryptocurrency transactions are common forms of such transactions. Cardless payment systems through digital wallets, contactless payments, and real time payment systems are growing globally and replacing conventional payment methods (Gomber et al., 2018). Digital wallets are particularly attractive to users because they combine convenience with speed, but research on mobile payment adoption shows that anxiety about security and trust in providers remain central to their continued use (Patil et al., 2020).

Alongside these conveniences, the rapid expansion of digital financial transactions has also widened the opportunities for fraud. Digital finance can broaden access to financial services for people who were previously excluded from the formal banking system, and it supports financial inclusion and economic stability when it is accompanied by adequate safeguards (Ozili, 2018). The use of artificial intelligence and other Industry 4.0 technologies has further accelerated this inclusion by making financial services cheaper, faster, and easier to reach for individuals and small businesses (Mhlanga, 2020). However, the same characteristics that make digital transactions attractive, namely their speed, high volume, and independence from physical presence, also make them vulnerable to misuse. Fraudsters exploit these features through schemes such as identity theft, account takeover, phishing, social engineering, and unauthorized card use, which can be completed within seconds and are often difficult to reverse once the funds have been transferred (Abdallah et al., 2016). Because millions of transactions are processed every day, fraudulent activity represents only a very small fraction of the total and is hidden among legitimate transactions, which makes it hard to identify with conventional rule based or manual methods (West & Bhattacharya, 2016). Fraud patterns also change continuously as perpetrators adapt to new security measures, so detection systems must be able to recognize anomalies that have never been seen before rather than relying only on known patterns (Hilal et al., 2022). These conditions show that the growth of digital financial transactions must be matched by more adaptive and intelligent fraud detection mechanisms to protect users and maintain trust in the digital financial system.

### *2.3 Artificial Intelligence in Fraud Detection*

Artificial intelligence is a branch of computer science that aims to create systems that can imitate human intelligence, including the ability to learn, reason, understand language, and make decisions. This technology enables the rapid and accurate analysis of large amounts of data (Jarrahi, 2018). In fraud detection, AI is applied through supervised learning, which learns from transactions labeled as legitimate or fraudulent, and unsupervised learning, which identifies transactions that deviate from normal patterns (Abdallah et al., 2016, Hilal et al., 2022). Deep learning methods have expanded the ability to detect complex anomalies in high dimensional data (Pang et al., 2022). Combining unsupervised and supervised learning improves detection performance (Carcillo et al., 2021), ensemble methods such as AdaBoost and majority voting increase accuracy (Randhawa et al., 2018), and features that capture customers' spending behavior over time substantially improve detection (Correa Bahnsen et al., 2016). A champion challenger comparison at a card issuer also showed that a deep learning model outperformed a hybrid ensemble model in both offline and live tests (Kim et al., 2019).

### *2.4 AI in Auditing and Internal Control*

AI is also changing the practice of auditing and internal control. Kokina and Davenport (2017) explain that automation and AI allow auditors to analyze entire populations of transactions rather than samples, to detect anomalies continuously, and to focus human effort on judgment intensive tasks. Issa et al. (2016) describe how AI can formalize parts of the audit process and supplement the audit workforce, while emphasizing that professional judgment remains essential for complex decisions.

From an internal control perspective, AI based fraud detection therefore functions as a detective control that operates in real time, while human reviewers act as a compensating control for the errors and limitations of automated systems.

The value of AI in internal control, however, depends on how well its outputs can be understood, trusted, and acted upon by the people responsible for oversight. Jarrahi (2018) argues that AI and humans have complementary strengths, with AI excelling at processing large volumes of data quickly and consistently, while humans contribute intuition, contextual understanding, and the ability to handle ambiguous situations. Raisch and Krakowski (2021) similarly caution that organizations which rely on AI purely for automation risk losing the human expertise needed to question and correct automated outputs, and they recommend combining automation with augmentation so that machines and people learn from each other over time. This balance is especially important in fraud detection because many machine learning models operate as black boxes whose reasoning cannot easily be explained to auditors, managers, or customers whose transactions have been blocked (Adadi & Berrada, 2018). Rudin (2019) contends that for high stakes decisions, organizations should prefer interpretable models or at least ensure that the basis of each decision can be traced, since unexplained outputs make it difficult to establish accountability. Trust also plays a central role, because users and reviewers who do not understand how an AI system reaches its conclusions may either over rely on it or ignore its alerts altogether (Glikson & Woolley, 2020). In addition, fraud patterns shift continuously, so models trained on historical data gradually lose accuracy unless they are monitored and retrained, a problem known as concept drift (Lu et al., 2019). For these reasons, an effective internal control design does not treat AI as a replacement for human judgment but positions it within a structured process of review, documentation, and periodic evaluation, in which human oversight ensures that automated decisions remain accurate, explainable, and aligned with organizational policies.

### *2.5 Machine Learning Blind Spots*

In machine learning, a blind spot refers to a situation in which an AI model fails to detect fraud patterns it has never seen before, such as zero day patterns, or in which the training data are not representative of real conditions. Fraud perpetrators exploit technological gaps, social engineering, and new attack vectors to take advantage of weaknesses in detection systems. Blind spots have several causes, including concept drift, data bias, adversarial attacks, and limited data. Concept drift occurs when the statistical properties of the data change over time, so that a model trained on past data loses accuracy on new data (Lu et al., 2019). In fraud detection, concept drift is particularly important because fraudsters continuously change their strategies, and the extreme imbalance between legitimate and fraudulent transactions makes learning more difficult (Dal Pozzolo et al., 2018).

### *2.6 Human Intervention, Explainability, and System Calibration*

The adoption of AI in fraud detection requires attention to data governance and human oversight. AI systems cannot operate optimally without regular calibration and tuning. Thresholds must be adjusted to changes in transaction behavior, models must be updated periodically to avoid concept drift, and the features used by models must be evaluated and refreshed, all of which require human intervention. Human oversight continues to play an important role in maintaining compliance (KPMG, 2025). Collaboration between AI and humans allows institutions to use the speed and scalability of AI while relying on the critical and contextual judgment that only humans can provide (Raisch & Krakowski, 2021).

Explainability is an important condition for effective human oversight. Explainable AI refers to methods that make AI decisions understandable to humans and is regarded as essential for responsible AI (Barredo Arrieta et al., 2020, Adadi & Berrada, 2018). For high stakes decisions, interpretable models may be preferable to black box models (Rudin, 2019), and explainable machine learning has been applied successfully in financial risk management (Bussmann et al., 2021). Explainability also increases users' trust in and acceptance of AI systems (Shin, 2021), while trust in AI depends on its

transparency and reliability (Glikson & Woolley, 2020, Choung et al., 2023). The European General Data Protection Regulation contains provisions on automated decision making and human involvement, although scholars debate whether it creates a full right to explanation (Wachter et al., 2017).

## 3. Research Methodology

### *3.1 Research Design*

This study used a literature review approach. A literature review was chosen because the research questions concern the synthesis of existing knowledge about the effectiveness, limitations, and governance of AI in fraud detection rather than the testing of a specific model on a particular dataset. Literature reviews are suitable for integrating findings from different disciplines and for developing conceptual frameworks (Snyder, 2019), and this study followed general guidelines for literature reviews in business research (Paul & Criado, 2020) and the reporting principles of the PRISMA 2020 statement where applicable (Page et al., 2021).

### *3.2 Data Sources and Analysis*

The literature consisted of scientific journal articles, professional reports, and regulatory documents related to AI based fraud detection in digital financial transactions, auditing, and internal control. The literature was identified using keywords such as artificial intelligence, machine learning, fraud detection, false positives, concept drift, human oversight, and digital wallets. Sources were screened for relevance to the three research questions, and the selected literature was synthesized thematically into three themes, namely the effectiveness of AI compared with manual methods, the limitations of AI, and the role of calibration and human intervention.

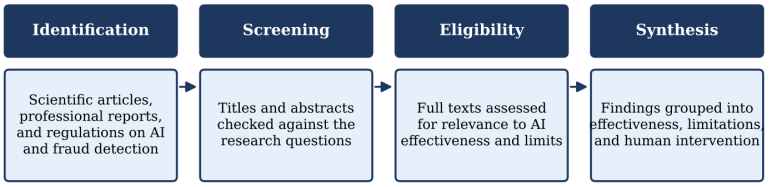



Figure 1. Stages of the literature review process

Figure 1 shows the four stages of the literature review used in this study. In the identification stage, scientific articles, professional reports, and regulations on AI and fraud detection were collected. In the screening stage, titles and abstracts were checked against the research questions. In the eligibility stage, the full texts were assessed for their relevance to the effectiveness and limitations of AI. In the synthesis stage, the findings were grouped into three themes, namely effectiveness, limitations, and human intervention.

## 4. Results and Discussions

### *4.1 Effectiveness and Accuracy of AI Compared with Manual Methods*

AI has proven superior to rule based fraud detection in several respects. Machine learning and deep learning have become the main foundations of fraud detection across the financial sector (Ali, Othman, Eisa, Al-Dhaqm, Nasser, Elhassan, Elshafie, & Saif, 2022; Hilal, Gadsden, & Yawney, 2022). AI systems can analyze thousands of transactions per second, identify anomalies within milliseconds, and issue warnings before a transaction is completed, which is impossible for human auditors to achieve. Reviews of machine learning based fraud detection show that ensemble and deep learning models consistently achieve high accuracy and recall on benchmark and industry datasets

© 2026 Abnus | Open Access — CC BY 4.0 | abnus.co | DOI:10.52909/corelytics.v1i4.467 

(Randhawa, Loo, Seera, Lim, & Nandi, 2018), and deployments at card issuers show that advanced models can outperform existing systems in live operation (Kim, Lee, Shin, Yang, Cho, Nam, Song, Yoon, & Kim, 2019).

Several financial institutions have reported successful implementation of AI in fraud detection. In Indonesia, several banks and digital wallet providers have adopted such solutions to improve transaction security. These systems enable the early detection of suspicious transactions, which ultimately protects customers from financial losses. Large banks have also begun integrating AI into their transaction monitoring systems to detect money laundering and other illegal transactions (KPMG, 2025).

The difference in speed between AI and manual audit procedures is highly significant in the context of digital transactions, which occur very quickly. While auditors may need hours or days to review a sample of transactions, AI can process the entire population of transactions simultaneously (Kokina & Davenport, 2017). Beyond time efficiency, automation reduces the operational cost of large scale manual testing and allows auditors to focus on investigations that require professional judgment (Issa, Sun, & Vasarhelyi, 2016). From an internal control perspective, AI therefore strengthens detective controls by expanding coverage from samples to full populations and by shortening the time between a fraudulent event and its detection.

Table 1. Comparison between manual or rule based fraud detection and AI based fraud detection

| Aspect | Manual or Rule Based Detection | AI Based Detection |
|---|---|---|
| Coverage | Samples of transactions | Entire population of transactions |
| Speed | Hours or days for review | Real time analysis within milliseconds |
| Adaptability | Fixed rules updated manually | Models that learn from new data |
| Detection of complex patterns | Limited to known rules | Detection of nonlinear and hidden anomalies |
| Main weakness | Low scalability and slow response | False positives and blind spots |

Table 1 compares manual or rule based fraud detection with AI based fraud detection across five aspects. Manual and rule based methods cover only samples of transactions, require hours or days for review, rely on fixed rules that must be updated manually, and are limited to known rules, so their main weakness is low scalability and slow response. AI based detection covers the entire population of transactions, analyzes them in real time within milliseconds, learns from new data, and can detect nonlinear and hidden anomalies, but its main weaknesses are false positives and blind spots. The table shows that AI improves the effectiveness of internal control while introducing new risks that must be managed. The table was compiled by the authors from the synthesis of the reviewed literature.

### *4.2 Limitations of AI, False Positives, and Machine Learning Blind Spots*

Although AI shows impressive performance, the system has limitations that cannot be ignored. Based on the literature review, two main challenges emerge, namely high false positives and machine learning blind spots. These are considered structural limitations of AI that cannot be fully eliminated merely through better data or algorithms. False positives occur when an AI system classifies normal transactions as suspicious or fraudulent. Industry reports note that false positive rates remain substantial even in optimized systems, which leads to customer friction, unnecessary blocking of legitimate transactions, and alert fatigue among investigators (The Payments Association, 2025). Research using real transaction data confirms that investigators can review only a limited number of alerts, so the precision of alerts is a critical measure of system usefulness (Dal Pozzolo et al., 2018).

Machine learning blind spots refer to the fundamental weakness of AI systems in recognizing fraud patterns that have never been encountered before, so that in some cases fraudulent transactions escape detection. Concept drift causes models trained on historical data to lose accuracy as fraud strategies change (Lu et al., 2019). Data bias and limited labeled data cause models to perform poorly on

underrepresented groups or rare types of fraud, and adversarial perpetrators deliberately adapt their behavior to avoid detection (Hilal et al., 2022). Deep anomaly detection methods can help detect unknown patterns, but they also face challenges of interpretability and high false alarm rates (Pang et al., 2022).

Table 2. Main limitations of AI based fraud detection and their causes

| Limitation | Description | Main Causes | Consequence |
|---|---|---|---|
| False positives | Legitimate transactions classified as suspicious or fraudulent | Strict thresholds, class imbalance, incomplete behavioral profiles | Customer friction and alert fatigue |
| Blind spots | Failure to detect fraud patterns never seen before | Concept drift, data bias, adversarial attacks, limited labeled data | Undetected fraud and financial losses |
| Black box decisions | Decisions that cannot be explained to users or auditors | Complex models without explanation methods | Low trust and difficulty in accountability |

Table 2 summarizes the main limitations of AI based fraud detection. False positives occur when legitimate transactions are classified as suspicious or fraudulent because of strict thresholds, class imbalance, and incomplete behavioral profiles, which leads to customer friction and alert fatigue. Blind spots occur when systems fail to detect fraud patterns never seen before because of concept drift, data bias, adversarial attacks, and limited labeled data, which results in undetected fraud and financial losses. Black box decisions occur when complex models cannot explain their decisions to users or auditors, which reduces trust and makes accountability difficult. The table was compiled by the authors from the synthesis of the reviewed literature.

### *4.3 The Role of Calibration and Human Intervention*

Although AI offers significant advantages, the technology is not free from errors, and AI systems cannot operate optimally without regular calibration and tuning. Human intervention is essential for reducing the negative effects of false positives, because professionals with audit competence exercise professional judgment that cannot be replaced by AI systems (Kokina & Davenport, 2017). In many jurisdictions, regulations have begun to require human involvement in automated decision making that has significant effects on individuals. Examples of relevant regulations include the European General Data Protection Regulation, the European Digital Operational Resilience Act, and Indonesia's Personal Data Protection Law, which indicate that AI cannot fully replace humans in decision making with significant legal and financial consequences (Wachter, Mittelstadt, & Floridi, 2017).

Based on the literature, the most effective model of fraud detection is collaboration between humans and AI rather than the dominance of either party. In this model, AI acts as a tool that performs initial screening, identifies anomalies, and issues alerts, while humans act as the final decision makers who conduct further investigation, verify the findings of AI, and determine appropriate and rational actions based on the facts found in each case. This collaboration allows financial institutions to use the speed and scalability of AI while relying on the critical and contextual judgment that only humans can provide (Jarrahi, 2018). Explainable AI supports this collaboration by giving investigators understandable reasons for each alert, which speeds up verification and improves accountability (Barredo, Díaz-Rodríguez, Bennetot, Tabik, Barbado, Garcia, Gil-Lopez, Molina, Benjamins, Chatila, & Herrera, 2020; Bussmann, Giudici, Marinelli, & Papenbrock, 2021).

© 2026 Abnus | Open Access — CC BY 4.0 | abnus.co | DOI:10.52909/corelytics.v1i4.467 

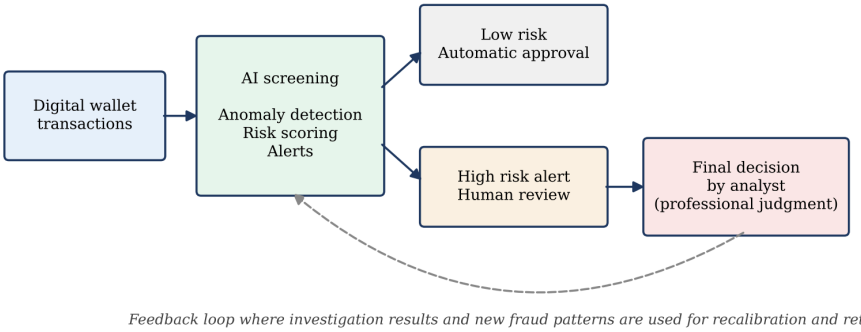



Figure 2. Human in the loop framework for AI based fraud detection in digital wallets

Figure 2 shows the human in the loop framework derived from the review. Digital wallet transactions on the left are first screened by AI through anomaly detection, risk scoring, and alerts. Transactions assessed as low risk are approved automatically, while transactions that trigger high risk alerts are sent for human review. Analysts then make the final decision based on professional judgment. The dashed arrow represents a feedback loop in which the results of human investigations and newly discovered fraud patterns are used to recalibrate and retrain the AI model, which reduces false positives and blind spots over time.

Table 3. Division of roles between AI and humans in fraud detection

| Stage | Role of AI | Role of Humans |
|---|---|---|
| Screening | Analyze all transactions in real time and score risk | Set thresholds and risk appetite |
| Alerting | Flag anomalies and prioritize alerts | Review high risk alerts and verify evidence |
| Decision | Provide explanations for flagged transactions | Make final decisions with professional judgment |
| Learning | Retrain models with new data | Label investigated cases and recalibrate models |

Table 3 shows the division of roles between AI and humans across four stages of fraud detection. In the screening stage, AI analyzes all transactions in real time and scores their risk, while humans set thresholds and risk appetite. In the alerting stage, AI flags anomalies and prioritizes alerts, while humans review high risk alerts and verify the evidence. In the decision stage, AI provides explanations for flagged transactions, while humans make the final decisions using professional judgment. In the learning stage, AI retrains its models with new data, while humans label investigated cases and recalibrate the models. The table shows that AI and humans perform complementary rather than competing functions. The table was compiled by the authors from the synthesis of the reviewed literature.

### *4.4 Discussion*

The findings show that AI is more effective and accurate than manual methods for detecting fraud, but that its effectiveness depends on how it is governed. This conclusion is consistent with research showing that AI in management creates value through augmentation, in which humans and machines combine their strengths, rather than through full automation (Raisch & Krakowski, 2021). It is also consistent with evidence from auditing that AI changes rather than eliminates the role of auditors, shifting their work toward the interpretation of results and the exercise of judgment (Kokina & Davenport, 2017).

For digital wallet providers in Indonesia, the findings imply that investments in AI should be accompanied by investments in human expertise, data governance, and explainability. Trust in digital

payment systems depends on both the security of transactions and the fairness of decisions that affect users, such as blocking accounts or transactions (Glikson, & Woolley, 2020; Patil, Tamilmani, Rana, & Raghavan, 2020). Systems that frequently block legitimate transactions or cannot explain their decisions may undermine user trust even when they detect fraud effectively. Regulators can support this balance by setting standards for model validation, human review of high impact decisions, and the explainability of AI systems used in financial services.

### *4.5 Implications for Internal Control and Governance*

From the perspective of internal control, the findings suggest that AI should be treated as one component of a broader control system rather than as a standalone solution. Preventive controls, such as strong authentication and limits on transaction amounts, reduce the opportunity for fraud before it occurs. Detective controls based on AI identify suspicious transactions as they occur, and corrective controls, such as investigation, account recovery, and the reimbursement of victims, address the consequences of fraud that was not prevented. Human reviewers connect these layers by interpreting alerts, deciding on responses, and feeding lessons back into the design of preventive and detective controls (Kokina, & Davenport, 2017; Issa, Sun, & Vasarhelyi, 2016).

Governance of AI models is equally important. Financial institutions should document the data, assumptions, and performance of their fraud detection models, monitor them continuously for concept drift, and validate them periodically using independent reviewers (Lu, Liu, Dong, Gu, Gama, & Zhang, 2019). Performance should not be evaluated only by overall accuracy, because fraud data are highly imbalanced and a model can appear accurate while missing many fraud cases or generating many false alarms (Dal, Boracchi, Caelen, Alippi, & Bontempi, 2018). Measures such as precision, recall, and the cost of errors should therefore be reported to management and auditors. Explainable models and explanation tools help internal auditors assess whether AI decisions are consistent with policies and regulations (Rudin, 2019). Finally, institutions should consider the experience of customers, because unnecessary blocking of legitimate transactions can reduce trust in digital wallets, while transparent communication about why a transaction was flagged can strengthen it (Glikson & Woolley, 2020).

## 5. Conclusions

Based on the discussion, AI has proven to be more effective and accurate than manual methods in detecting fraud because it can process all transactions in real time with much higher speed and detection rates while reducing the operational costs of internal control. However, AI still has structural limitations in the form of high false positives and machine learning blind spots toward new fraud patterns that have never been recognized before, so it cannot be relied upon entirely on its own. Regular system calibration and human intervention through professional judgment remain necessary to verify detection results and make final decisions, in line with regulatory requirements such as the General Data Protection Regulation, the Digital Operational Resilience Act, and Indonesia's Personal Data Protection Law. Collaboration between AI and humans is therefore the most optimal approach to fraud detection in the financial sector, with AI acting as the initial screening tool and humans as the final decision makers.

Based on these conclusions, financial institutions are advised to ensure that every AI alert, especially high risk alerts, is verified by humans before final decisions are made, to calibrate systems regularly with the latest transaction data so that they can recognize new fraud patterns and reduce the risk of blind spots, and to build feedback mechanisms that use the results of human investigations of undetected fraud as learning data to improve the accuracy of AI systems continuously.

This study is a literature review and has not been tested empirically. Future research is advised to conduct empirical studies using real transaction data, to compare the effectiveness of different AI algorithms, and to examine potential algorithmic bias in more depth, so that the results can make a more applicable contribution to the financial industry in Indonesia.

© 2026 Abnus | Open Access — CC BY 4.0 | abnus.co | DOI:10.52909/corelytics.v1i4.467 

## Acknowledgement

The authors thank Universitas Padjadjaran and the organizers of the National Accounting Paper Competition 2026 for the opportunity and support provided during the preparation of this work.

## Author Contributions

KNR contributed to the conceptualization of the study, the literature search, the synthesis of findings, and the writing and revision of the manuscript. S contributed to the conceptualization of the study, the literature search, the synthesis of findings, and the writing and revision of the manuscript. Both authors read and approved the final manuscript.

## Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

## References

Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications, 68, 90–113*. https://doi.org/10.1016/j.jnca.2016.04.007

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6, 52138–52160*. https://doi.org/10.1109/ACCESS.2018.2870052

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. *Applied Sciences*, *12*(19). https://doi.org/10.3390/app12199637

Association of Certified Fraud Examiners. (2024). *Occupational fraud 2024: A report to the nations*.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58, 82–115*. https://doi.org/10.1016/j.inffus.2019.12.012

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, *57*(1), 203-216. https://doi.org/10.1007/s10614-020-10042-0

Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences, 557, 317–331*. https://doi.org/10.1016/j.ins.2019.05.042

Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human–Computer Interaction*, *39*(9), 1727-1739. https://doi.org/10.1080/10447318.2022.2050543

Correa Bahnsen, A., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications, 51, 134–142*. https://doi.org/10.1016/j.eswa.2015.12.030

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(8), 3784-3797. https://doi.org/10.1109/TNNLS.2017.2736643

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627-660. https://doi.org/10.5465/annals.2018.0057

Global Anti Scam Alliance & Feedzai. (2024). *Global state of scams report 2024*.

Gomber, P., Kauffman, R. J., Parker, C., & Weber, B. W. (2018). On the fintech revolution: Interpreting the forces of innovation, disruption, and transformation in financial services. *Journal of Management Information Systems*, *35*(1), 220-265. https://doi.org/10.1080/07421222.2018.1440766

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. *Expert Systems with Applications, 193, Article 116429*. https://doi.org/10.1016/j.eswa.2021.116429

Issa, H., Sun, T., & Vasarhelyi, M. A. (2016). Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation. *Journal of Emerging Technologies in Accounting*, *13*(2), 1-20. https://doi.org/10.2308/jeta-10511

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577-586. https://doi.org/10.1016/j.bushor.2018.03.007

Kim, E., Lee, J., Shin, H., Yang, H., Cho, S., Nam, S., Song, Y., Yoon, J., & Kim, J. (2019). Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning. *Expert Systems with Applications, 128, 214–224*. https://doi.org/10.1016/j.eswa.2019.03.042

Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. *Journal of Emerging Technologies in Accounting*, *14*(1), 115-122. https://doi.org/10.2308/jeta-51730

KPMG. (2025). *The role of AI in financial crime compliance*.

Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2019). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, *31*(12), 2346-2363. https://doi.org/10.1109/TKDE.2018.2876857

Mhlanga, D. (2020). Industry 4.0 in finance: The impact of artificial intelligence (AI) on digital financial inclusion. *International Journal of Financial Studies*, *8*(3). https://doi.org/10.3390/ijfs8030045

Ozili, P. K. (2018). Impact of digital finance on financial inclusion and stability. *Borsa Istanbul Review*, *18*(4), 329-340. https://doi.org/10.1016/j.bir.2017.12.003

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372, Article n71*. https://doi.org/10.1136/bmj.n71

Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2022). Deep learning for anomaly detection: A review. *ACM Computing Surveys*, *54*(2). https://doi.org/10.1145/3439950

Patil, P., Tamilmani, K., Rana, N. P., & Raghavan, V. (2020). Understanding consumer adoption of mobile payment in India: Extending Meta-UTAUT model with personal innovativeness, anxiety, trust, and grievance redressal. *International Journal of Information Management, 54, Article 102144*. https://doi.org/10.1016/j.ijinfomgt.2020.102144

Paul, J., & Criado, A. R. (2020). The art of writing literature review: What do we know and what do we need to know? International Business Review, 29(4), Article 101717. https://doi.org/10.1016/j.ibusrev.2020.101717

The Payments Association. (2025). *AI and fraud prevention: The hidden risks of false positives and black-box models*.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192-210. https://doi.org/10.5465/amr.2018.0072

Randhawa, K., Loo, C. K., Seera, M., Lim, C. P., & Nandi, A. K. (2018). Credit card fraud detection using AdaBoost and majority voting. *IEEE Access, 6, 14277–14284*. https://doi.org/10.1109/ACCESS.2018.2806420

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies, 146, Article 102551*. https://doi.org/10.1016/j.ijhcs.2020.102551

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104, 333–339*. https://doi.org/10.1016/j.jbusres.2019.07.039

Vousinas, G. L. (2019). Advancing theory of fraud: The S.C.O.R.E. *model. Journal of Financial Crime*, *26*(1), 372-381. https://doi.org/10.1108/JFC-12-2017-0128

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, *7*(2), 76-99. https://doi.org/10.1093/idpl/ipx005

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security, 57, 47–66*. https://doi.org/10.1016/j.cose.2015.09.005

© 2026 Abnus | Open Access — CC BY 4.0 | abnus.co | DOI:10.52909/corelytics.v1i4.467 