ABNUS JOURNAL
Your Research Partner

Financial Analytics and Economic Innovation Journal (FAEIJ)
E-ISSN: 3164-337X | Vol. 1, No. 4 (2026) | pp. 213-226

Artificial Intelligence for Digital Transaction Fraud Detection and the Credibility of Indonesian Cashless Payments

Erwin Xandra[1*], Ernesta Angelica[2], Ayrin Lolita Chandra[3]

Universitas Kristen Krida Wacana, Jakarta, Indonesia[1,2,3]

***Corresponding author:** erwin.322024053@civitas.ukrida.ac.id |

Received: 28 August 2026 | Revised: 7 September 2026 | Revised: 23 September 2026 | Published: 29 October 2026

Abstract

Purpose: This study analyzes the role of Artificial Intelligence (AI) in improving the effectiveness of fraud detection through real time anomaly detection, behavioral profiling, and machine learning, and evaluates its contribution to the credibility of cashless payment systems in Indonesia.

Research Methodology: A systematic literature review was conducted on national and international scientific articles published between 2020 and 2025 in Scopus, Google Scholar, Garuda, and SINTA, complemented by official reports from Bank Indonesia, the Financial Services Authority, and the Indonesian Fintech Association. The literature was selected using inclusion and exclusion criteria and analyzed thematically.

Results: The most effective detection approach combines real time anomaly detection for speed, behavioral profiling for misuse of valid credentials, and ensemble machine learning for accuracy, forming a layered defense against increasingly competent fraud perpetrators. Credibility depends not only on algorithmic accuracy but also on ability, integrity through explainability, and benevolence through victim protection. Effective stakeholder synergy is determined mainly by gaps in adoption capacity among institutions.

Conclusions: AI should be positioned as a strategic foundation for the credibility of digital payments, supported by explainable systems and shared data infrastructure.

Liimitations: The study is a literature review and does not empirically test the relationship between AI transparency and user trust.

Contributions: The study integrates fraud, trust, and diffusion theories to explain how AI strengthens payment credibility and offers recommendations for regulators, payment providers, and researchers.

Keywords: *Artificial Intelligence, Cashless Payment, Credibility, Digital Payment System, Fraud Detection*

How to Cite: Xandra, E., Angelica, E., & Chandra, A. L. (2026). Artificial Intelligence for Digital Transaction Fraud Detection and the Credibility of Indonesian Cashless Payments. *Financial Analytics and Economic Innovation Journal (FAEIJ)*, 1(4), 213-226.

1. Introduction

Indonesia is undergoing a digital transformation toward a cashless society, marked by exponential growth in the use of digital payment systems such as the Quick Response Code Indonesian Standard (QRIS), electronic wallets, and mobile banking. According to Bank Indonesia, QRIS had been used by about 57 million users and 39.3 million merchants by the first half of 2025, with a total of 6.05 billion transactions worth IDR 579 trillion (Bank, 2025). Digital finance can expand financial inclusion and support economic stability, particularly in emerging economies where many people previously had limited access to formal financial services (Ozili, 2018). The fintech revolution has transformed how

© 2026 Abnus | Open Access — CC BY 4.0 | abnus.co | DOI:10.52909/faeij.v1i4.464

financial services are designed, delivered, and regulated, creating new opportunities as well as new risks (Gomber, Kauffman, Parker, & Weber, 2018).

The growth in transaction volume, however, has been followed by an escalation of fraud threats, ranging from the misuse of fake QRIS codes and the theft of One Time Passwords (OTP) to phishing attacks and account takeover. These threats cause financial losses and erode public trust in the national digital payment ecosystem. Data compiled from the Financial Services Authority (*Otoritas Jasa Keuangan*/OJK) show that from November 2024 to November 2025, fictitious online shopping fraud was reported in 64.933 cases with losses of IDR 1.14 trillion, while impersonation fraud was reported in 39.978 cases with losses of IDR 1.54 trillion (CNBC, 2025). The Indonesia Anti Scam Centre (IASC) of OJK recorded more than 432.000 reports of digital fraud with total losses of IDR 9.1 trillion between November 2024 and January 2026 (Babel, 2026).

The credibility of digital payment systems is an essential foundation for the success of a cashless society in Indonesia. When the public loses confidence in the security of payment systems, the adoption of financial technology is hindered and the national target for financial inclusion may be threatened. Research on mobile payment adoption shows that trust and perceived risk are among the strongest determinants of whether consumers adopt and continue to use digital payment services (Patil, Tamilmani, Rana, & Raghavan, 2020). Effective and responsive strategies for mitigating fraud risk are therefore an urgent need.

Artificial intelligence has proven to be a highly effective solution for detecting fraud in the global financial industry. By using machine learning and deep learning algorithms, Artificial Intelligence (AI) can analyze massive volumes of transaction data in real time, identify anomalous patterns, and predict fraudulent activity with high accuracy (West, & Bhattacharya, 2016; Hilal, Gadsden, & Yawney, 2022). The technology applies innovative approaches such as real time anomaly detection, behavioral profiling, and ensemble methods that combine multiple algorithms to improve the robustness and accuracy of prediction (Carcillo, Caelen, Kessaci, Oblé, & Bontempi, 2021). AI can also support digital financial inclusion by enabling safer and more efficient financial services (Mhlanga, 2020).

Nevertheless, most research on the application of AI to digital transaction fraud detection still focuses on the technical evaluation of machine learning models. Systematic reviews show that the literature concentrates on algorithm performance metrics such as accuracy, precision, and recall (Ali, Othman, Eisa, Al-Dhaqm, Nasser, Elhassan, Elshafie, & Saif, 2022), while studies that specifically analyze how AI implementation strengthens the credibility of and public trust in cashless payment ecosystems in Indonesia remain relatively limited. This gap underlies the need for the present study.

Based on this gap, the study addresses three research questions. The first concerns how AI mechanisms, namely real time anomaly detection, behavioral profiling, and machine learning, improve the speed and accuracy of fraud detection in digital payment transactions in Indonesia. The second concerns the extent to which AI implementation in fraud detection contributes to the security, transparency, and public trust that underpin the credibility of cashless payment systems. The third concerns the factors that influence the effectiveness of synergy among digital payment providers, banks, regulators, and technology developers in implementing AI in an integrated manner. The study contributes theoretically by integrating fraud theory, trust theory, and diffusion of innovation theory, and practically by offering recommendations for regulators, payment providers, users, and researchers.

2. Literature Review

2.1 Fraud Pentagon Theory

The fraud pentagon theory extends the earlier fraud triangle and fraud diamond models. The fraud triangle originally contained three elements, namely pressure, opportunity, and rationalization, and the fraud pentagon adds two further elements, namely competence and arrogance (Astuti & Geraldina,

2023). Pressure is a financial or nonfinancial drive that motivates a person to commit fraud. Opportunity is a gap or weakness in a control system that allows fraud to occur without being detected. Rationalization is the justification that a perpetrator constructs so that the fraudulent act appears reasonable. Competence is the ability of a perpetrator to override internal controls, conceal actions, and control social situations for personal gain. Arrogance is an attitude of superiority or a belief that controls and legal consequences do not apply to oneself. Reviews of fraud theory show that models of fraud have been continually extended to capture new drivers such as capability, ego, and collusion, because traditional models no longer explain the increasing frequency and sophistication of fraud (Vousinas, 2019).

This theory is relevant for understanding the evolution of cybercrime in digital payment systems. Modern fraud perpetrators are no longer individuals with limited technical knowledge but people with high technical capability, or competence, who can design phishing schemes, counterfeit QRIS codes, and break into OTP systems, and who act with strong confidence that they can avoid detection, or arrogance. The theory is therefore used as a framework for understanding that fraud mitigation in cashless payment systems cannot rely only on closing opportunities but must also keep pace with the increasing competence of perpetrators.

2.2 Technology Acceptance Model and Trust Theory

The technology acceptance model explains an individual's acceptance of technology through two main perceptions, namely perceived usefulness and perceived ease of use (Davis, 1989). Pavlou (2003) extended the model by adding trust and perceived risk, because electronic transactions involve little face to face interaction and therefore create greater uncertainty than conventional transactions. The unified theory of acceptance and use of technology adds performance expectancy, effort expectancy, social influence, and facilitating conditions as determinants of technology use (Venkatesh et al., 2016), and studies of electronic services show that perceived usefulness and ease of use mediate the effects of other factors on adoption (Chen & Aklikokou, 2020). In Indonesia, research on QRIS adoption among micro, small, and medium enterprises shows that technology readiness and trust jointly shape adoption decisions (Nurqamarani et al., 2024).

To explain how trust is formed, this study refers to the integrative model of Mayer et al. (1995), which describes three dimensions of trust. Ability is the belief in the competence of the trusted party, integrity is the belief in its honesty and consistency, and benevolence is the belief that the trusted party prioritizes the interests of its users. These three dimensions are relevant for explaining how AI based fraud detection systems build users' sense of security. Ability is reflected in the accuracy of real time anomaly detection, integrity in the transparency of verification processes, and benevolence in a system orientation that favors user protection. The lower the fraud risk that users perceive, the stronger the credibility of the cashless payment ecosystem as a whole. Reviews of human trust in AI further show that trust depends on the tangibility, transparency, reliability, and immediacy of AI systems (Glikson & Woolley, 2020), and that trust plays a central role in the acceptance of AI technologies (Choung et al., 2023).

2.3 Behavioral Profiling

In cybersecurity research, behavioral profiling, also known as user and entity behavior analytics, is an approach that builds a baseline profile of the normal habits of each user or entity in a system, including the usual time, location, device, and amount of transactions. Current activity is then continuously compared with the baseline, so that deviations such as logging in from a new device, transaction amounts outside usual patterns, or sudden spikes in transaction frequency can be flagged immediately as risky. Research on credit card fraud shows that features that capture customers' spending behavior over time, such as the periodicity of transaction times, substantially improve detection performance (Correa Bahnsen et al., 2016).

The advantage of this approach over static rule based systems lies in its ability to form a distinctive behavioral identity for each user, so that it can still recognize fraud that uses the valid credentials of

victims, such as account takeover after OTP theft. Such cases are difficult to detect when a system only verifies login credentials. Behavioral profiles are also dynamic and are updated automatically through continuous learning, so that their accuracy improves over time. This approach is the main basis for answering the research question that highlights the role of behavioral profiling in recognizing user behavior that indicates potential digital crime.

2.4 Machine Learning and Anomaly Detection

Anomaly detection is a technique for recognizing data or activities that deviate significantly from the normal patterns of a system (Pang et al., 2022). While behavioral profiling focuses on the profile of each individual, anomaly detection works at the level of aggregate transaction patterns through two main approaches. Supervised learning methods such as random forest, support vector machine, and logistic regression are trained on data labeled as legitimate or fraudulent, while unsupervised methods such as k means clustering and isolation forest group unlabeled data to discover patterns that deviate from the majority (Hilal, Gadsden, & Yawney, 2022).

Combining several algorithms through ensemble learning has been shown in many studies to improve the accuracy and recall of detection compared with single algorithms. Randhawa et al. (2018) found that AdaBoost and majority voting improve the detection of credit card fraud, and Carcillo et al. (2021) showed that combining unsupervised and supervised learning improves detection accuracy. Research using real transaction data also emphasizes that fraud detection systems must handle concept drift, class imbalance, and the limited capacity of investigators to review alerts (Dal Pozzolo et al., 2018). A champion challenger comparison at a card issuer found that a deep learning model outperformed a hybrid ensemble model in both offline and live tests (Kim et al., 2019).

2.5 Explainable Artificial Intelligence

Many high performing AI models operate as black boxes whose internal reasoning is difficult for humans to understand. Explainable artificial intelligence refers to methods that make the decisions of AI systems understandable to humans and is regarded as essential for responsible AI (Barredo, Díaz-Rodríguez, Bennetot, Tabik, Barbado, Garcia, Gil-Lopez, Molina, Benjamins, Chatila, & Herrera, 2020; Adadi, & Berrada, 2018). Rudin (2019) argues that for high stakes decisions, organizations should prefer inherently interpretable models rather than trying to explain black box models after the fact. Experimental research shows that explainability and causability increase users' trust in and acceptance of AI systems (Shin, 2021). In finance, explainable machine learning has been applied to credit risk management to make AI predictions transparent for regulators and users (Bussmann, Giudici, Marinelli, & Papenbrock, 2021).

2.6 Diffusion of Innovation Theory

Diffusion of innovation theory explains how an innovation spreads in society through five attributes (Rogers, 1962). Relative advantage is the degree to which an innovation is better than existing methods. Compatibility is the consistency of an innovation with the habits and values of users. Complexity is the degree of difficulty in using the innovation. Trialability is the opportunity to try the innovation on a small scale, and observability is the extent to which its benefits are visible to others. The higher the relative advantage and compatibility and the lower the complexity of an innovation, the faster it is adopted. In Indonesia, diffusion theory has been used to explain the adoption of fintech payment systems among millennials (Hidayat, 2023). This framework is also relevant for interpreting the rapid adoption of QRIS, because the removal of the merchant discount rate for certain sectors increased the relative advantage of QRIS over cash, while QRIS Tap based on near field communication reduced the complexity of its use (Bank Indonesia, 2025). At the organizational level, AI adoption is also a question of digital transformation, because organizations must adjust their structures, capabilities, and processes to create value from new technologies (Vial, 2019), and public institutions such as regulators face specific challenges in applying AI (Wirtz et al., 2019).

3. Research Methodology

3.1 Research Design

This study used a systematic literature review. A systematic literature review was chosen because it allows researchers to identify, evaluate, and synthesize previous research in a transparent and replicable way and to develop conceptual insights from a broad body of evidence (Snyder, 2019). The review followed general guidelines for systematic reviews in business and management research Paul and Criado (2020) and Kraus, Breier, and Dasí-Rodríguez (2020), and the reporting principles of the PRISMA 2020 statement (Page, McKenzie, Bossuyt, Boutron, Hoffmann, Mulrow, Shamseer, Tetzlaff, Akl, Brennan, Chou, Glanville, Grimshaw, Hróbjartsson, Lalu, Li, Loder, Mayo-Wilson, & McDonald, 2021).

3.2 Data Sources and Selection

The review examined national and international scientific articles published between 2020 and 2025 in the Scopus, Google Scholar, Garuda, and SINTA databases, as well as official reports from Bank Indonesia, OJK, and the Indonesian Fintech Association (AFTECH) and relevant regulations. Keywords related to artificial intelligence, fraud detection, digital payments, cashless society, and trust were used in the search. The literature was selected based on inclusion and exclusion criteria. Articles were included when they discussed the application of AI or machine learning to fraud detection or the relationship between digital payment security and user trust, and they were excluded when they were not related to the research questions or were not available in full text.

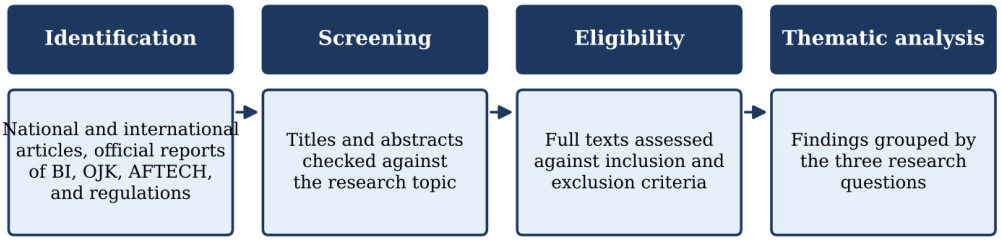


Figure 1. Stages of the systematic literature review process

Figure 1 shows the four stages of the review process. In the identification stage, national and international articles were collected together with official reports from Bank Indonesia, OJK, and AFTECH and relevant regulations. In the screening stage, titles and abstracts were checked against the research topic. In the eligibility stage, full texts were assessed against the inclusion and exclusion criteria. In the final stage, the selected literature was analyzed thematically and the findings were grouped according to the three research questions. The figure was prepared by the authors in 2026.

3.3 Data Analysis

The selected literature was analyzed thematically. Findings were coded and grouped into three themes that correspond to the research questions, namely the mechanisms through which AI improves the speed and accuracy of fraud detection, the contribution of AI to the security, transparency, and credibility of cashless payment systems, and the factors that influence synergy among stakeholders. Each theme was interpreted through the theoretical frameworks described in the literature review, namely the fraud pentagon theory, the technology acceptance model and trust theory, and diffusion of innovation theory. Data from official reports and credible news sources were used to illustrate the scale of digital transactions and fraud in Indonesia.

© 2026 Abnus | Open Access — CC BY 4.0 | abnus.co | DOI:10.52909/faeij.v1i4.464

4. Results and Discussions

4.1 The Scale of Digital Transactions and Fraud in Indonesia

Table 1. Key indicators of digital transactions and digital fraud in Indonesia

Indicator	Value	Period	Source
QRIS users	About 57 million	First half of 2025	(Bank, 2025)
QRIS merchants	39.3 million	First half of 2025	(Bank, 2025)
QRIS transactions	6.05 billion transactions worth IDR 579 trillion	First half of 2025	(Bank, 2025)
Growth in QRIS transaction volume	594% compared with the same period in 2023	First quarter of 2025	(Bank, 2025)
Fictitious online shopping fraud	64.933 cases, losses of IDR 1.14 trillion	November 2024 to November 2025	(CNBC, 2025)
Impersonation fraud	39.978 reports, losses of IDR 1.54 trillion	November 2024 to November 2025	(CNBC, 2025)
Digital fraud reports to IASC	More than 432,000 reports, losses of IDR 9.1 trillion	November 2024 to January 2026	(Babel, 2026)

Table 1 table shows the scale of digital transactions and digital fraud in Indonesia. By the first half of 2025, QRIS had about 57 million users and 39.3 million merchants and had processed 6.05 billion transactions worth IDR 579 trillion, and in the first quarter of 2025 its transaction volume grew by 594% compared with the same period in 2023. At the same time, fictitious online shopping fraud accounted for 64.933 cases with losses of IDR 1.14 trillion and impersonation fraud for 39.978 reports with losses of IDR 1.54 trillion between November 2024 and November 2025, while the IASC received more than 432.000 reports with total losses of IDR 9.1 trillion between November 2024 and January 2026. The table shows that rapid growth in digital payments has been accompanied by large and rising losses from fraud. The table was compiled by the authors from the cited reports.

4.2 Mechanisms of AI in Improving the Speed and Accuracy of Fraud Detection

The scale of digital transaction growth in Indonesia clearly shows why AI based fraud detection has become an urgent need. A scale and speed of growth of this magnitude make manual or static rule based detection systems inadequate. The authors therefore argue that the shift to adaptive AI based detection is not merely a strategic option but a structural necessity for the sustainability of the national digital payment ecosystem. Reviews of financial fraud detection confirm that static rules cannot keep pace with evolving fraud patterns and that intelligent methods based on data mining and machine learning are needed (West, & Bhattacharya, 2016; Abdallah, Maarof, & Zainal, 2016).

This argument becomes stronger when it is linked to the fraud pentagon theory. The two elements that distinguish modern fraud from conventional fraud are competence and arrogance. Perpetrators of digital fraud now have the technical ability to design phishing schemes, counterfeit QRIS codes, and break into OTP systems, and they act with strong confidence that they can escape detection. The large numbers of fictitious shopping and impersonation fraud cases illustrate the increasing technical competence and boldness of perpetrators (CNBC, 2025). Mitigation strategies that focus only on closing opportunities, for example through user education alone, will not be sufficient without detection systems that can technically match the competence of perpetrators.

At this point, behavioral profiling and machine learning based anomaly detection play an important role. The main strength of behavioral profiling lies in its ability to detect fraud that uses the valid credentials of victims, which is the scenario most difficult to recognize for ordinary login verification systems. This ability is highly relevant for addressing the high number of account takeover cases after

OTP theft in Indonesia, in which perpetrators technically enter the victim's account in a legitimate way but behave differently from the baseline of the original user, for example by transacting at unusual times, locations, or amounts (Correa, Aouada, Stojanovic, & Ottersten, 2016). Ensemble machine learning that combines supervised methods such as random forest and support vector machines with unsupervised methods such as isolation forest and k means allows the system to detect both known fraud patterns and new patterns that have not appeared in historical data (Pang, Shen, Cao, & van den Hengel, 2022). This capability is crucial because digital fraud patterns continuously change.

The authors therefore argue that effective fraud detection cannot rely on a single approach. The three approaches must be combined, namely real time anomaly detection for speed of response, behavioral profiling for detecting the misuse of valid credentials, and ensemble machine learning for accuracy and adaptability to new fraud patterns. This combination forms a layered defense that directly responds to the competence and arrogance of fraud perpetrators described in the fraud pentagon theory. Evidence from systematic reviews and real world deployments supports this view, showing that hybrid and ensemble models outperform single models and that practical constraints such as limited investigator capacity must be considered in system design (Kim, Lee, Shin, Yang, Cho, Nam, Song, Yoon, & Kim, 2019).

Table 2. AI mechanisms for fraud detection and their responses to the fraud pentagon

AI Mechanism	Main Function	Fraud Scenario Addressed	Fraud Pentagon Element Countered
Real time anomaly detection	Rapid identification of unusual transactions	Sudden spikes in transactions and abnormal amounts	Opportunity
Behavioral profiling	Comparison of current activity with the user's behavioral baseline	Account takeover after OTP theft and misuse of valid credentials	Competence
Ensemble machine learning	Combination of supervised and unsupervised algorithms for accuracy and adaptability	Known and new fraud patterns, including fake QRIS and phishing	Competence and arrogance

Table 2 summarizes how the three AI mechanisms respond to different fraud scenarios and to the elements of the fraud pentagon. Real time anomaly detection rapidly identifies unusual transactions such as sudden spikes and abnormal amounts, which narrows the opportunity for fraud. Behavioral profiling compares current activity with each user's behavioral baseline and therefore addresses account takeover after OTP theft and other misuse of valid credentials, which reflects the technical competence of perpetrators. Ensemble machine learning combines supervised and unsupervised algorithms to detect both known and new patterns, including fake QRIS codes and phishing, and therefore counters both the competence and the arrogance of perpetrators who believe they can avoid detection. The table was compiled by the authors from the synthesis of the reviewed literature.

4.3 Contribution of AI to the Security, Transparency, and Credibility of Cashless Payment Systems

Referring to the technology acceptance model and trust theory, the authors argue that the contribution of AI to the credibility of payment systems is not merely a technical issue but also a matter of perception that influences the public's decision to continue using digital payment services. When fraud detection systems work accurately and responsively, users' perceived risk decreases. Pavlou (2003) explains that this reduction in perceived risk directly encourages people to continue using electronic transactions, and studies of mobile payment confirm that trust and perceived security strongly predict adoption and continued use (Patil, Tamilmani, Rana, & Raghavan, 2020).

To assess the extent to which AI genuinely strengthens credibility, rather than merely being claimed to do so, the authors use the three dimensions of trust proposed by Mayer, Davis, and Schoorman (1995).

In terms of ability, the accuracy of the system in detecting anomalies in real time is concrete evidence of the technical competence of AI. In terms of integrity, the decisive factor is the consistency and transparency of the verification process itself, namely whether the system's decision to flag a transaction as suspicious can be reasonably explained or whether it is arbitrary and untraceable. In terms of benevolence, the benchmark is the extent to which the system and the institutions that operate it genuinely protect the interests of victims rather than merely fulfilling administrative obligations.

Empirical data are relevant to the benevolence dimension. The IASC recorded more than 432.000 reports of digital fraud with total losses of IDR 9.1 trillion from its establishment in November 2024 to January 2026, while the funds that were successfully recovered and returned to victims amounted to only hundreds of billions of rupiah (Babel, 2026). The wide gap between total losses and recovered funds may erode public perceptions of benevolence, even when detection systems technically work adequately.

Table 3. Trust dimensions and the contribution of AI to payment system credibility

Trust dimension	Meaning in AI Based Fraud Detection	Requirement for Credibility
Ability	Accuracy and speed of detecting anomalies in real time	High detection performance with few false alarms
Integrity	Consistency and transparency of verification decisions	Explainable decisions that can be traced by users and regulators
Benevolence	Orientation of systems and institutions toward protecting victims	Effective recovery of victims' funds and responsive complaint handling

Table 3 shows how each dimension of trust is reflected in AI based fraud detection and what is required for credibility. Ability refers to the accuracy and speed of detecting anomalies in real time and requires high detection performance with few false alarms. Integrity refers to the consistency and transparency of verification decisions and requires decisions that can be explained to and traced by users and regulators. Benevolence refers to the orientation of systems and institutions toward protecting victims and requires effective recovery of victims' funds and responsive complaint handling. The table indicates that technical accuracy addresses only the first dimension, so credibility also depends on explainability and victim protection. The table was compiled by the authors based on Mayer et al. (1995) and the reviewed literature.

This finding reveals something rarely discussed in the technical literature. The algorithmic accuracy of AI is a necessary but not sufficient condition for building the credibility of digital payment systems. A system may have statistically high fraud detection accuracy and still fail to build public trust if its operation is a black box that is not transparent and cannot be explained to users or regulators. Research on human trust in AI shows that transparency and reliability are key antecedents of trust (Glikson & Woolley, 2020), and experimental evidence shows that explainability increases trust and acceptance of AI systems (Shin, 2021). The authors therefore argue that the principle of explainable AI, namely the ability of a system to provide humanly understandable reasons for each fraud detection decision, deserves greater attention in the development of fraud detection systems in Indonesia (Barredo, Díaz-Rodríguez, Bennetot, Tabik, Barbado, Garcia, Gil-Lopez, Molina, Benjamins, Chatila, & Herrera, 2020; Adadi, & Berrada, 2018). Explainable AI is not merely an additional technical feature but an important instrument for building integrity and benevolence in the eyes of users. For high stakes decisions such as blocking accounts or transactions, interpretable models may be preferable to opaque models (Rudin, 2019), and explainable machine learning has already been applied successfully in other areas of financial risk management (Bussmann, Giudici, Marinelli, & Papenbrock, 2021).

4.4 Factors Influencing the Effectiveness of Synergy Among Stakeholders

To answer the third research question, the authors use diffusion of innovation theory as the analytical

framework (Rogers, 1962), because the success of national AI implementation is essentially a problem of diffusion across institutions rather than merely a matter of technological sophistication. The five attributes of diffusion can be mapped as follows.

Relative advantage should encourage payment providers to adopt AI as reductions in fraud losses are demonstrated. However, the continued increase in fraud losses, which reached IDR 9.1 trillion by early 2026 (Babel, 2026), indicates that the benefits of AI have not yet been realized evenly across all providers, which creates an implementation gap among institutions. Compatibility is reflected in Bank Indonesia's policy of removing the merchant discount rate for certain sectors and in the launch of QRIS Tap based on near field communication (Bank Indonesia, 2025), which align technical innovation with the transaction habits of the public and reduce resistance to new technology. Complexity is the main obstacle for small and medium providers. The simpler the process of adopting AI, for example through application programming interfaces or ready made models provided by regulators or fintech associations, the faster its diffusion, because limited resources are a structural constraint for these providers.

Trialability is reflected in the regulatory sandbox mechanisms facilitated by Bank Indonesia and OJK, which allow providers to test AI based fraud detection systems on a limited scale before national implementation and thereby reduce the risk of failure at an early stage. Observability is reflected in the regular publication of data on case handling by the IASC. This openness provides visible evidence of success for the public and for other providers, which encourages others to adopt similar standards. Research on digital transformation shows that organizations differ in their capabilities to adopt new technology and that these capability differences shape the outcomes of transformation (Vial, 2019).

Table 4. Diffusion of innovation attributes and stakeholder synergy in AI based fraud detection

Attribute	Evidence in the Indonesian Payment Ecosystem	Implication for Synergy
Relative advantage	Fraud losses continued to rise to IDR 9.1 trillion by early 2026	Benefits of AI are not yet evenly realized across providers
Compatibility	Removal of the merchant discount rate for certain sectors and launch of QRIS Tap	Innovation aligned with user habits reduces resistance
Complexity	Limited resources of small and medium providers	Shared tools and ready made models accelerate adoption
Trialability	Regulatory sandboxes facilitated by Bank Indonesia and OJK	Limited testing reduces the risk of failure
Observability	Regular publication of case data by the IASC	Visible results encourage adoption of common standards

Table 4 maps the five attributes of diffusion of innovation to evidence from the Indonesian payment ecosystem and their implications for stakeholder synergy. The continued rise in fraud losses shows that the relative advantage of AI is not yet evenly realized across providers. The removal of the merchant discount rate for certain sectors and the launch of QRIS Tap show compatibility with user habits, which reduces resistance. The limited resources of small and medium providers show that complexity is a major barrier that can be reduced through shared tools and ready made models. Regulatory sandboxes facilitated by Bank Indonesia and OJK provide trialability, and the regular publication of case data by the IASC provides observability that encourages the adoption of common standards. The table was compiled by the authors from the synthesis of the reviewed literature and official reports.

Based on this mapping, the authors argue that the main determinant of effective synergy lies not in the maturity of AI technology itself but in the gap in adoption capacity among stakeholders. Large banks and QRIS operators tend to be able to build AI systems independently, while small and medium providers depend more on common standards facilitated by regulators. Without mechanisms for data sharing and minimum standards set centrally by Bank Indonesia and OJK, fraud mitigation risks proceeding in a fragmented way. Perpetrators exploit this condition by moving between platforms to

avoid detection, as reflected in the increasingly organized and cross institutional nature of digital crime in Indonesia (Babel, 2026). Public sector research on AI also shows that regulators need appropriate capabilities, governance, and coordination mechanisms to support AI adoption effectively (Wirtz et al., 2019).

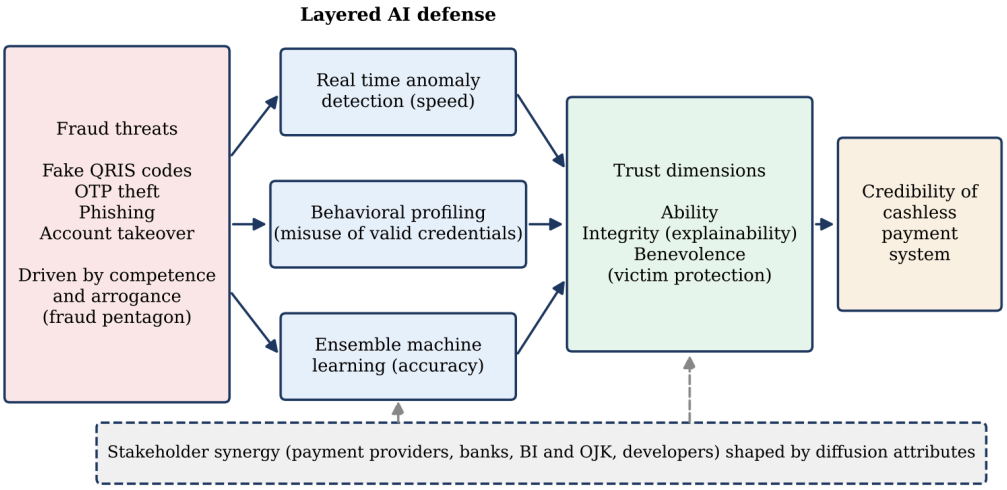


Figure 2. Conceptual framework of AI based fraud detection and the credibility of cashless payment systems

Figure 2 summarizes the conceptual framework that emerges from the review. On the left, fraud threats such as fake QRIS codes, OTP theft, phishing, and account takeover are driven by the competence and arrogance of perpetrators described in the fraud pentagon. These threats are countered by a layered AI defense in the middle, consisting of real time anomaly detection for speed, behavioral profiling for detecting the misuse of valid credentials, and ensemble machine learning for accuracy. The outputs of this defense shape the three trust dimensions of ability, integrity through explainability, and benevolence through victim protection, which in turn determine the credibility of the cashless payment system on the right. The dashed box at the bottom shows that stakeholder synergy among payment providers, banks, Bank Indonesia and OJK, and technology developers, shaped by the attributes of diffusion of innovation, supports both the AI defense and the formation of trust. The figure was prepared by the authors in 2026.

4.5 Discussion

The findings of this review show that AI is not merely a technical complement but a strategic foundation for the credibility of Indonesia's digital payment system. This conclusion is consistent with the broader fintech literature, which emphasizes that technological innovation in financial services must be accompanied by appropriate governance and risk management to sustain trust (Gomber, Kauffman, Parker, & Weber, 2018). It is also consistent with evidence that digital finance supports financial inclusion only when users trust the security of digital channels (Ozili, 2018; Mhlanga, 2020).

The review also contributes to theory by integrating three perspectives that are usually discussed separately. The fraud pentagon explains why fraud has become more sophisticated, trust theory explains why technical accuracy is not enough to build credibility, and diffusion of innovation theory explains why AI adoption is uneven across institutions. Together, these perspectives show that the effectiveness of AI depends on the technical design of detection systems, the transparency and fairness of their decisions, and the institutional arrangements that support their diffusion.

5. Conclusions

Indonesia's rapid adoption of QRIS, electronic wallets, and mobile banking has expanded financial inclusion but has also widened the opportunities for digital fraud. This review concludes that fraud is detected most effectively through a layered defense that combines real time anomaly detection, behavioral profiling, and ensemble machine learning. The credibility of cashless payment systems, however, depends not only on technical accuracy but also on trust built through ability, integrity, and benevolence, which requires explainable decisions and genuine protection of victims. The effectiveness of stakeholder synergy is determined mainly by gaps in adoption capacity between large and small providers rather than by technological maturity.

Bank Indonesia and OJK should therefore set binding standards for sharing fraud data and expand the IASC into a cross sector fraud database. Regulators should also provide shared AI infrastructure for smaller providers, and payment providers should adopt explainable AI. Because this study is a literature review, future research should empirically test how AI transparency and victim recovery influence user trust and the credibility of digital payment systems.

Acknowledgement

The authors thank Septian Bayu Kristanto and Febriani C. S. Magdalena for their guidance as supervisors, Universitas Kristen Krida Wacana for its support, and the organizers of the National Accounting Paper Competition for the opportunity to develop this work.

Author Contributions

EX contributed to the conceptualization of the study, literature search and selection, thematic analysis, and manuscript writing and revision. EA contributed to the literature search and selection, thematic analysis, and manuscript writing and revision. ALC contributed to the conceptualization of the study, thematic analysis, and manuscript writing and revision. All authors read and approved the final version of the 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

Astuti, D., & Geraldina, I. (2023). Peran fraud pentagon theory dalam mendeteksi fraud pada pelaporan keuangan (studi empiris pada badan usaha milik negara yang terdaftar di Bursa Efek Indonesia tahun 2015–2019) [The role of fraud pentagon theory in detecting fraud in financial reporting (an empirical study of state owned enterprises listed on the Indonesia Stock Exchange in 2015–2019)]. *Jurnal Keuangan dan Perbankan*, *18*(1), 13. https://doi.org/10.35384/jkp.v18i1.328

© 2026 Abnus | Open Access — CC BY 4.0 | abnus.co | DOI:10.52909/faeij.v1i4.464

Babel Insight. (2026). IASC OJK catat kerugian penipuan digital capai Rp9,1 triliun. Retrieved from https://www.babelinsight.id/kerugian-fraud-digital-ojk-triliun

Bank Indonesia,. (2025). QRIS Jelajah Indonesia 2025 dorong digitalisasi dengan wisata budaya (QRIS Jelajah Indonesia 2025 promotes digitalization through cultural tourism) (Press release). Retrieved from https://www.bi.go.id/id/publikasi/ruang-media/news-release/Pages/sp_2717025.aspx

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

Chen, L., & Aklikokou, A. K. (2020). Determinants of e-government adoption: Testing the mediating effects of perceived usefulness and perceived ease of use. *International Journal of Public Administration*, *43*(10), 850-865. https://doi.org/10.1080/01900692.2019.1660989

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

CNBC Indonesia. (2025). *10 modus penipuan paling banyak kuras rekening warga RI, waspada!*. Retrieved from https://www.cnbcindonesia.com/tech/20251227085126-37-697536/10-modus-penipuan-paling-banyak-kuras-rekening-warga-ri-waspada

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

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340. https://doi.org/10.2307/249008

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

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

Hidayat, A. R. (2023). Analisis adopsi penggunaan sistem pembayaran fintech pada generasi milenial menggunakan teori difusi inovasi [Analysis of the adoption of fintech payment systems among millennials using diffusion of innovation theory]. *Jurnal Ilmu Manajemen*, *13*(1), 117-132. https://doi.org/10.32502/jimn.v13i2.6974

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

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

Kraus, S., Breier, M., & Dasí-Rodríguez, S. (2020). The art of crafting a systematic literature review in entrepreneurship research. *International Entrepreneurship and Management Journal*, *16*(3), 1023-1042. https://doi.org/10.1007/s11365-020-00635-4

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709-734. https://doi.org/10.2307/258792

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

Nurqamarani, A. S., Fadilla, S., & Juliana, A. (2024). Revolutionizing payment systems: The integration of TRAM and trust in QRIS adoption for micro, small, and medium enterprises in Indonesia. *Journal of Information Systems Engineering and Business Intelligence*, *10*(3), 314-327. https://doi.org/10.20473/jisebi.10.3.314-327

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

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce*, *7*(3), 101-134. https://doi.org/10.1080/10864415.2003.11044275

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

Rogers, E. M. (1962). Diffusion of innovations. *Free Press*.

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

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2016). Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the Association for Information Systems*, *17*(5), 328-376. https://doi.org/10.17705/1jais.00428

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118-144. https://doi.org/10.1016/j.jsis.2019.01.003

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

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

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration*, *42*(7), 596-615. https://doi.org/10.1080/01900692.2018.1498103