Humanities and Social Sciences Communications

Article in Press

https://doi.org/10.1057/s41599-026-07406-6

AI-driven financial fraud detection in Pakistan's banking sector: bridging strategic intent and operational implementation

Received: 4 June 2025

Accepted: 17 April 2026

Cite this article as: Rahim, A., Ullah Jan, S., Ali, S. *et al.* AI-driven financial fraud detection in Pakistan's banking sector: bridging strategic intent and operational implementation. *Humanit Soc Sci Commun* (2026). https://doi.org/10.1057/s41599-026-07406-6

Adeel Rahim, Sharif Ullah Jan, Shujaat Ali, Dilawar Shah & Muhammad Tahir

AI-Driven Financial Fraud Detection in Pakistan's Banking Sector: Bridging Strategic Intent and Operational Implementation

Abstract

Financial fraud continues to pose a significant and persistent challenge for Pakistan's banking sector, particularly amid rapid digitalization and the expansion of online financial services. Artificial intelligence (AI) has emerged as a critical technological response to these evolving risks. This study examines how AI-based fraud detection and prevention systems are implemented in Pakistani banks and evaluates the extent to which strategic objectives align with operational practices. Adopting an exploratory qualitative design, the study integrates a systematic literature review with in-depth interviews involving branch managers, fraud officers, IT specialists, senior executives, and customers across multiple banking institutions. The findings reveal limited awareness and understanding of AI-driven fraud management among branch-level staff and customers, with responsibility for fraud monitoring remaining highly centralized among senior management. Though advanced techniques such as machine learning, anomaly detection, deep learning, and natural language processing are technically available, their operational utilization remains constrained. Grounded in the Technology Acceptance Model (TAM) and the Resource-Based View (RBV), the current study demonstrates that user acceptance, workforce capability, and organizational readiness critically shape AI effectiveness. The study highlights the urgent need for enhanced FinTech literacy, decentralized AI deployment, and stronger governance mechanisms to bridge the strategic-operational divide. These measures are essential for aligning operational practices with strategic intent and for unlocking AI's full potential to enhance fraud resilience, regulatory compliance, and customer trust.

Keywords: *Artificial Intelligence, Financial Institutions, Financial Frauds, FinTech.*

Practitioner Notes

What is already known about this topic

- Financial institutions globally face significant losses from fraud, and many have begun to employ AI techniques (e.g., machine learning, neural networks) to detect anomalous transactions. These technologies can reduce manual workload and improve early threat identification.
- Prior research on AI in banking has largely focused on algorithm development and theoretical performance, with evidence that advanced analytics can improve fraud detection accuracy (Hasham et al., 2019; Kaya et al., 2019). However, practical adoption is uneven and often challenged by data and resource limitations.
- Emerging-market banks often encounter barriers such as limited technical infrastructure, regulatory constraints, and workforce skill gaps that can slow down AI implementation in fraud prevention.

What this paper adds

- Qualitative examination of Pakistan's banking industry highlights the gap between branch operational knowledge and headquarters strategic AI objectives. Branch managers know nothing about AI tools, but upper management handles fraud detection.
- Highlights Pakistani bank enablers and challenges. AI-driven technologies, such as anomaly detection and NLP for transaction data, have potential; however, interviewees noted staff training, data quality issues, and compliance concerns.
- Includes AI-driven solutions, including supervised/unsupervised learning models and actionable methods, linking theory and practice. Improvements include data governance, FinTech understanding, and organisational improvements to deploy AI tools more widely in banks.

Implications for practice and/or policy

- Banks should teach all workers, especially branch staff, to use artificial intelligence fraud-detection systems. Clear communication of AI programs and cross-departmental cooperation helps align daily operations with strategic goals.

- Decentralising AI implementation, giving branch offices AI-powered monitoring, may improve fraud response. Productive model training requires high-quality, full data sets, including customer and transactional data. Legislators and authorities should create AI deployment guidelines that protect security and privacy. This includes banking AI guidelines, digital infrastructure investment, and public-private knowledge sharing and technology transfer.
- Banks, technology companies, and authorities communicate to adapt artificial intelligence fraud-detection systems to shifting risk profiles. Collaboration ensures that breakthrough ideas, like hybrid artificial intelligence models, are ethically used to boost industry resilience and consumer confidence.

1. Introduction

Financial frauds such as payment fraud, identity fraud, and cybercrime in banking represent increasing risks to banking stability, regulatory compliance, and customer confidence, especially in rapidly digitizing banking markets like Pakistan's. Cyber fraud volumes and sophistication have grown along with the rapid expansion of mobile and Internet banking. As promising tools to help improve fraud prevention, artificial intelligence (AI) technologies, which include computer systems that can perform tasks that usually require human intelligence, have emerged. Techniques based on AI, such as machine learning models for anomalous behavior detection and biometric analyses, can provide a way to quickly analyze large volumes of transactional data and detect and flag potentially fraudulent activity in real time (Bhatla et al., 2003; Jang-Jaccard & Nepal, 2014; Homer, 2020). For instance, in Pakistan, banks have started implementing AI-based solutions (e.g., BPC's SmartVista platform) to continuously monitor all transaction types across various channels, while also using link analysis to detect complex fraud networks (Bhatla et al., 2003). Industry analysts suggest that most of the AI-based fraud detection and prevention in Pakistan are focused primarily on two areas of application: fraud detection systems (including automated tools designed to identify and reject illicit transactions), and digital identity management platforms. These applications work together to provide additional support for risk management and customer protection (Mohanty & Mishra, 2023).

Although AI-based fraud detection has the potential to be very beneficial, a clear gap remains between the strategic intent to implement AI-based fraud detection and actual operational implementation at many banks. The State Bank of Pakistan (SBP) has identified this challenge as well. The SBP is developing policies to facilitate the responsible and transparent use of AI in financial services, and is doing so as part of its larger agenda to encourage innovation in financial technology (FinTech) (Schueffel, 2016). In its 2024 Financial Stability Review, the SBP reported that multiple Pakistani banks are beginning to integrate new and emerging technologies into their operations, such as robotics process automation, AI-powered virtual assistants, and machine learning algorithms for fraud detection and risk management (Vieira & Sehgal, 2018; Taherdoost, 2021; Adelakun, 2023). Governance research demonstrates that characteristics of boards of directors, oversight mechanisms, and other elements of governance are significant predictors of digital preparedness, cybersecurity disclosure, and the effective governance of advanced technologies such as AI (Alodat & Mansour, 2024). However, experts also indicate that there are more things that need to be done than just purchasing more advanced software in order to realize the potential benefits of AI-based fraud detection. Banks must also establish clear governance frameworks, develop sufficient human capital with the skills required to work with AI-based fraud detection systems, and establish the capacity to collect and manage quality data for the purposes of supporting AI systems. In short, if the underlying organizational capabilities do not exist (i.e., training programs and data management practices), the effectiveness of AI-based fraud detection systems will be severely limited.

Based on these considerations, the purpose of this study is to examine how Pakistani banks can bridge the gap between their strategic objectives (reducing fraud losses, regulatory compliance, and customer confidence) and the practical reality of implementing AI-based fraud detection systems. This study takes an exploratory approach by using in-depth interviews with senior bank managers and technology officers, and by conducting a review of relevant literature and regulatory documents (Kuvaas, 2006; Bozkus Kahyaoglu & Caliyurt, 2018; Lee & Bruvold, 2019). Our theoretical framework incorporates formal models

from both technology adoption and strategic management. Specifically, the Technology Acceptance Model (TAM) (Davis, 1989) is used to understand how users' perceptions of the usefulness and ease of use of AI-based fraud detection systems affect their willingness to accept and utilize those systems. Recent empirical studies suggest that the relationship between perceived usefulness/ease of use and acceptance is significantly conditioned by technological readiness and perceived AI usability, and that ethical and integrity considerations moderate these relationships (Al-Maaitah et al., 2025). The Resource-Based View (RBV) of the firm (Barney, 1991; Assensoh-Kodua, 2019) indicates that a bank's ability to successfully utilize AI-based fraud detection is dependent upon the availability of internal resources/capabilities that support the utilization of AI (e.g., high-quality data, competent personnel, and stable IT infrastructure). Utilizing these perspectives, the study seeks to understand the perceived value of AI-based fraud detection, identify the primary obstacles to implementation, and determine what organizational factors contribute to or inhibit the effective integration of AI into fraud prevention processes. By integrating these findings, we intend to inform practitioners and policy-makers regarding the optimal strategies for linking AI-based fraud detection initiatives to strategic objectives within the context of Pakistan's banking sector.

Despite increasing scholarly and regulatory attention toward AI applications in financial fraud management, existing research remains fragmented along organizational, technical, and governance dimensions. Prior research has primarily focused on either algorithmic performance or technical accuracy, while typically neglecting the relationship among organizational capabilities, user acceptance, and regulatory environments that collectively affect the effectiveness of AI-based fraud management, particularly in developing countries. This study bridges this gap by combining TAM and RBV to explore the extent to which strategic intentions surrounding AI adoption are converted (or not converted) into operational practices within Pakistan's banking industry.

This study makes three specific theoretical, empirical, and practical contributions. First, the study develops a new theory through analytical integration of TAM and RBV to explain why AI-based fraud detection systems may demonstrate high levels of performance at the strategic level but be poorly utilized at the operational level. Second, the study expands the emerging-market AI literature by presenting both qualitative and systematic evidence from Pakistan, a context underrepresented in the recent AI governance and digital banking research (2023-2025). Third, the study offers practical contributions through a distinction between AI-based fraud detection and fraud prevention, and provides actionable advice for banks and regulators who want to align their AI governance, workforce capabilities, and strategic objectives.

This study provides additional value through its focus on translating strategic intent into operational practice in an emerging-market banking context, as opposed to the majority of the previously published AI fraud literature that focuses on the technical aspects of AI fraud detection systems. As opposed to the majority of previous studies that have focused on the algorithmic performance or model accuracy, this research combines TAM and RBV to explain why AI-based fraud detection systems continue to be centralized and poorly utilized at the branch level in Pakistani banks. The study empirically provides rare qualitative evidence from Pakistan's banking industry, capturing perspectives from senior managers, IT specialists, fraud officers, and branch-level staff, a methodological approach missing from most of the previous studies. Practically, the study produces actionable knowledge for banks and regulators by separating AI-based fraud detection and fraud prevention and by indicating that governance, workforce capabilities, and organizational preparedness are significant drivers for resolving the strategic-operational gap. In doing so, the study provides context-specific knowledge to the developing body of literature on responsible and effective AI deployment in financial services.

The rest of the paper is structured as follows: Section 2 reviews the current body of literature on AI in banking and fraud detection, and identifies the research gaps. Section 3 outlines our research methodology. Sections 4 and 5 then describe the qualitative results and show how strategic intent can be more effectively linked to operational practices. Section 6 concludes with implications and possible future avenues of investigation.

2. Literature Review

2.1 AI in Banking and Fraud Detection

A progressing body of knowledge highlights that AI adoption in banking can substantially enhance security and efficiency. Global reviews note that AI tools such as machine learning, predictive analytics, and natural language processing (NLP) may improve personalization, automate processes, and, critically, strengthen risk management and fraud detection. For example, Dwivedi et al. (2021) identified enhanced fraud detection as a major benefit of AI adoption in financial services, while Bello et al. (2023) reported that AI-driven real-time transaction monitoring has significantly improved fraud detection rates, although they also highlighted operational challenges in deploying these tools at scale. In practice, banks deploy a range of AI techniques for fraud management, such as supervised models for transaction classification and unsupervised approaches for anomaly detection in unlabeled data (Perols, 2011; Kumar et al., 2022). Further advanced methods, including deep learning and graph-based analytics, have further enhanced detection accuracy, especially in identifying complex and coordinated fraud patterns (Vassio et al., 2022). Instead of algorithmic novelty, the literature increasingly emphasizes that the effectiveness of these techniques depends on organizational governance, integration, and user capability.

Contemporary research emphasizes that the effectiveness of AI in banking fraud management depends not only on algorithmic sophistication but also on governance mechanisms and strategic alignment. Evidence from corporate governance and disclosure studies exhibits that board effectiveness significantly enhances accountability, transparency, and technology-related disclosures, especially in a regulated environment (Khan et al., 2025; Mansour et al., 2025). Similarly, Papagiannidis (2025) argued that responsible AI adoption in financial institutions requires multi-level governance structures that align organizational strategy, operational implementation, and ethical oversight. Similarly, Vuković (2025), in a large-scale scientometric review of AI in banking, found and reported that AI-enabled fraud detection delivered sustained value only when embedded within organizational processes and supported by accountability and governance frameworks. These insights postulate that AI-driven fraud detection should be conceptualized as a strategic organizational capability rather than a standalone technological solution. This perspective aligns with innovation-performance research showing that eco-innovation and digital capability generate superior financial outcomes only when supported by firm-level resources and contextual factors, e.g., organizational size and absorptive capacity (Mansour et al., 2024a). More recent research on digital transformation further emphasizes that organizational culture plays a critical role in shaping how digital technologies are embedded into operational and accounting practices (Hasan et al., 2025). Furthermore, research shows that digital intelligence systems generate strategic value only when supported by effective knowledge sharing and ambidextrous capabilities, enabling organizations to balance control, learning, and innovation (Alsarayreh et al., 2025).

In a similar vein, research has identified significant challenges associated with AI deployment. Cross-market analyses point to ethical and governance issues such as algorithmic bias, lack of transparency in "black-box" models, technological constraints, and even cultural resistance as common barriers. In many emerging economies, the adoption of AI lags due to limited data infrastructure and shortages of specialized talent. Studies in Asian banking contexts emphasize that establishing clear governance frameworks is essential when implementing AI, in order to ensure consumer trust and regulatory compliance (Lim et al., 2021; Wu et al., 2021). These research works suggest that technological capability alone is not sufficient; effective use of AI also depends on organizational readiness and oversight. Recent studies further demonstrate that ownership structure and state involvement can moderate the relationship between board characteristics and ESG performance, influencing how governance mechanisms shape responsible technology use in emerging and Asia-Pacific markets (Al-Tahat et al., 2025). Bibliometric evidence on digital technologies and corporate sustainability further indicates that governance, accountability, and strategic alignment are central to realizing long-term value from AI adoption (Alshdaifat et al., 2024).

2.2 AI Techniques for Fraud Detection

In practice, Pakistani banks utilize AI-enabled fraud detection systems to monitor transactions in real time, applying a variety of fraud-detection techniques for risk-scoring and transaction monitoring. Although theoretically capable of rapidly identifying anomalies, interviews suggest that these systems' preventive capabilities remain limited by factors such as organizational preparedness, system integration, and workforce capacity (Butt & Aswani, 2021; Latif et al., 2021).

Together, these techniques form a layered fraud-prevention architecture; alerts generated by machine learning models might trigger higher-level reviews using graph network analysis or prompt additional biometric verification. The research also indicates that there is no one fraud detection model/technique that will prevent all types of fraud; therefore, robust fraud detection systems will include a combination of AI-based models, rule-based checks, and human-expert review to obtain both high detection rates and low false positives.

2.3 AI and Fraud Detection in the Pakistani Context

Pakistan's banking system offers a very specific context for fraud detection using AI, where there are both a number of strong drivers for innovation and significant operational constraints. On one hand, Pakistan has a young and tech-savvy population and a very high mobile phone penetration rate that creates a large opportunity for digital banking to grow and for AI to help. There are recently reported figures showing a 12.2 percent year over year increase in AI adoption within Pakistan's banking sector that is primarily focused on improving the risk management and fraud prevention capabilities of banks. Banks in Pakistan are experimenting with AI in a variety of areas, including mobile app security (for example, detecting fraudulent login attempts) and back-end risk analysis. For example, BPC's SmartVista platform has been used by multiple banks in Pakistan (including Muslim Commercial and Askar Banks Limited) to enable them to monitor transactions in real time and prevent fraud. This local enthusiasm reflects broader regional trends; for example, banks in neighboring India have also leveraged AI technologies to enhance their fraud prevention efforts (see Malali & Gopalakrishnan, 2020). The overall development demonstrates that Pakistan's financial industry is embracing FinTech innovations, including AI-driven fraud detection tools, to respond to the evolving threat environment and increasing demand from consumers for safe digital banking.

However, research also indicates that significant gaps in organizational preparedness exist in order to effectively implement AI in fraud detection. Butt and Aswani (2021) reported that significant deficiencies in IT infrastructure, data quality, and digital skills exist among Pakistani banks. Likewise, Latif et al. (2021) reported that while Pakistani bank managers exhibit moderate confidence in AI, they express concerns about cultural and organizational preparedness (i.e., resistance to change or lack of employee training). The State Bank of Pakistan (2018, 2021) has articulated a visionary agenda for the digital transformation of the banking sector and the deployment of emerging technologies. However, it has also acknowledged that the actual implementation has been slower than anticipated. Therefore, although there is clear regulatory support for the use of AI and a growing demand from customers for AI-enabled services, on-the-ground implementation continues to be constrained by legacy systems, skills gaps in staff, and data integration challenges.

The regulatory authorities are proactively addressing these issues. The SBP's draft guidelines on responsible AI use specifically emphasize applications in fraud detection and risk management and aim to ensure that banks utilize AI ethically and effectively. Iqbal (2025) reports that according to surveys conducted by the SBP, approximately half of Pakistan's regulated financial entities have either developed or are in the process of developing AI tools for fraud detection, customer service, credit assessment, and other purposes. The new regulations are intended to require banks to define a clear risk appetite for AI systems and address compliance, security, and the environmental impact of AI. The evolving governance landscape is encouraging banks to formally incorporate AI into their fraud prevention strategies.

In spite of the positive developments, the existing literature and practice leave many questions unanswered concerning how Pakistani banks can overcome the identified operational obstacles and fully capitalize on the potential of AI in fraud detection. Notably, few studies have investigated the intersection

of technological factors (i.e., the efficacy of algorithms and user acceptance of AI tools) with organizational factors (i.e., availability of skilled employees, manager support, etc.) related to the deployment of AI for fraud detection. Therefore, in order to address these knowledge gaps, the present study utilizes the theoretical perspectives discussed above. According to TAM, the adoption of AI-based fraud detection systems will be influenced, at least in part, by the perceptions of bankers of the utility and usability of the technology. In contrast, according to the RBV, a bank's internal resources will play a critical role in determining the bank's ability to develop and maintain such systems. Therefore, this study proposes a conceptual framework (Figure 1) which integrates insights from the systematic literature review and interview findings. It illustrates how different AI techniques (machine learning (ML), deep learning (DL), natural language processing (NLP), and hybrid approaches), feed into two separate but interrelated outcomes, namely fraud detection (post-transaction monitoring and anomaly identification), and fraud prevention (real-time blocking, biometric authentication, and predictive protection).

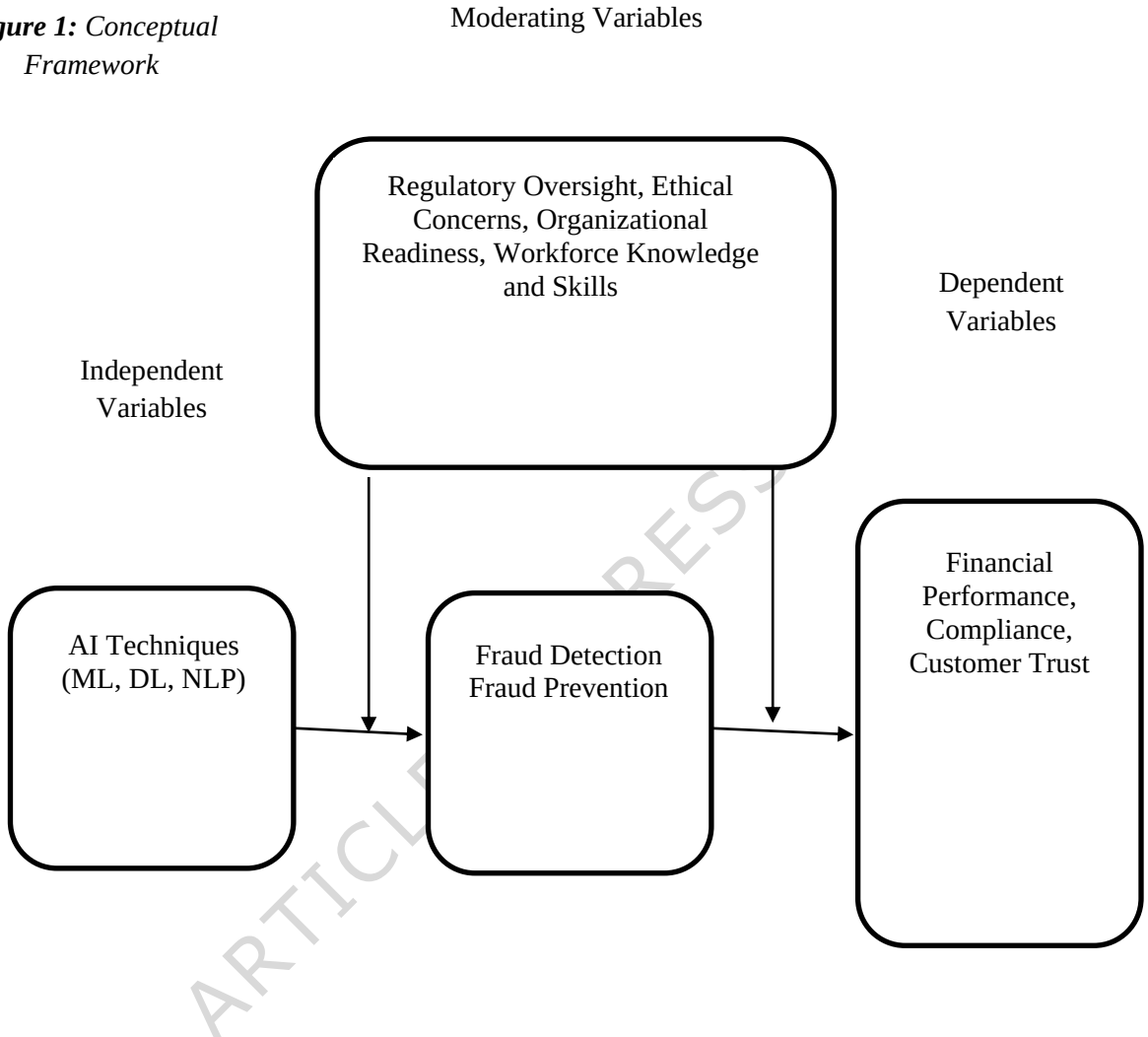
While the technical literature highlights rapid progress in machine learning, deep learning and hybrid AI models for fraud detection, recent governance-oriented research advises caution when adopting technology-centric approaches to AI. Nwachukwu, Chima, and Okolo (2025) argue that, in a regulated financial environment, AI systems should be designed with built-in controls, transparency, and auditability to ensure alignment with organizational risk strategies and regulatory requirements. If such "governance by design" principles are not followed, technologically sophisticated fraud detection systems may face limited adoption, mistrust, and suboptimal utilization. Thus, this perspective further underscores the importance of investigating AI techniques in combination with organizational preparedness and governance structures.

The above literature review further emphasizes the need to ensure strategic and organizational alignment when implementing artificial intelligence (AI) technology in emerging market financial systems. Developing economy evidence, as presented by Moraa (2025), shows that many AI-based projects in the financial services industry fail to perform adequately, resulting in a lack of workforce skills to implement these technologies, inadequate digital platforms to support their deployment, and an inability to effectively integrate strategic intent and operational execution. Findings in Pakistan's banking sector further reinforce these results; while Pakistan's banks utilize AI-based fraud management tools, these tools are typically deployed in a central manner and do not have a direct link to fraud detection at the branch level. Overall, it can be concluded that there are several contextual variables (regulatory maturity, organizational capacity, and human capital) that significantly influence how an organization's AI strategy translates into operational outcomes in emerging markets.

In synthesizing the literature reviewed above, it was determined that successful AI-based fraud detection and prevention is the result of a combination of the technological capabilities of an organization, its organizational resources, its users' acceptance of the use of technology, and the governance structure(s) that govern the use of technology in an organization. These conclusions provide significant direction for the development of the research study's conceptual framework, which utilizes the Technology Acceptance Model (TAM) and the Resource-Based View (RBV) to describe how organizations' strategic intentions regarding the use of AI are implemented operationally in the form of fraud detection and prevention practices in Pakistan's banking sector.

Figure 1 represents the study's conceptual framework as an integrated application of existing theoretical models of digital governance and technology adoption, and not as an original construct. The conceptual framework was developed through the use of the TAM, the RBV, and other relevant literature from the areas of institutional and AI governance. The conceptual framework illustrates how various forms of AI are utilized for the purposes of fraud detection and prevention within the boundaries of an organization's available resources and the regulatory environment in which they operate. The conceptual framework is intended to provide a basis for applying general principles of digital governance and technology adoption to the specific context of Pakistan's banking sector.

Figure 1: *Conceptual Framework*



The framework also presents the impact of moderating factors on the success of AI-based fraud prevention systems in banks. Moderating factors include: regulatory oversight, ethical issues (e.g., data privacy, algorithmic bias), organizationally-related issues (such as infrastructure, systems integration), and employee knowledge/skill levels. These moderating factors have a significant impact on how successful AI-based fraud prevention systems will be. In addition to addressing fraud prevention, the framework illustrates how fraud prevention is connected to other bank-wide outcomes. Examples of these outcomes include: improved financial results, better compliance, and increased customer trust. The framework's ability to illustrate the relationship between technical capabilities and organizational and regulatory enablers provides a comprehensive way of understanding the implementation of AI in fraud prevention within banking institutions.

Using TAM and RBV as a theoretical foundation, the proposed framework views AI-based fraud prevention as an output of both individual-level acceptance and firm-level capability deployment. From a TAM perspective, the degree to which employees perceive AI-based fraud detection as useful and the ease of using AI-based fraud detection systems directly relate to the usage and acceptance of AI-based fraud detection by employees. From an RBV perspective, the long-term sustainability of an organization's ability to prevent fraud effectively is dependent on complementary organizational resources (such as quality data, knowledgeable employees, governing policies and practices, and integrated systems).

This study does not utilize statistical analysis to test causal relationships between variables, but instead utilizes TAM and RBV as lenses for qualitative analysis, and to provide a structured method for interpreting themes identified through the literature review and interview methods. Therefore, the hypotheses are viewed as theory-informed analytical expectations that guide the analysis and interpretation of the research findings.

Guided by the framework, this study formulates the following hypotheses to steer our inquiry:

H_1: Higher perceived ease of use and perceived usefulness of AI technologies are positively associated with the effectiveness of AI-driven financial fraud detection systems.

H_2: The implementation of AI significantly reduces the opportunity for financial fraud by identifying and flagging unusual transaction patterns.

H_3: Institutional and cultural factors significantly moderate the adoption and success of AI in financial fraud detection across different regulatory environments.

H_4: The availability of key organizational resources, such as high-quality data, skilled personnel, and robust technological infrastructure, enhances the effectiveness of AI-based fraud detection systems.

H_5: AI implementation improves the efficiency of risk-based fraud detection by enabling the prioritization of high-risk transactions.

These hypotheses, which emerge from the gaps identified in the literature, form the basis of our investigation. Together, they facilitate a focused examination of how Pakistani banks can more effectively align their strategic objectives with operational practices in implementing AI-driven fraud prevention measures.

2.4 Theoretical Integration and Research Model Development

Beginning with the study's methodology, it utilizes an integrated theoretical framework, drawing upon the TAM and the RBV to identify the factors affecting the effectiveness of financial fraud detection using AI. Although previous studies have utilized each theory individually, the interplay between them will provide a more comprehensive understanding of why some organizations successfully utilize AI-based fraud detection systems while others do not, especially in developing countries like Pakistan. From a TAM viewpoint, the adoption and utilization of AI-based fraud detection systems is dependent on employee attitudes toward the perceived usefulness and perceived ease of use of those systems (Davis, 1989). The perceived usefulness of AI systems impacts how much branch managers and front-line staff believe that the system improves fraud detection, reduces manual workloads, and supports timely decision making, thus promoting the active involvement of employees with AI-generated alerts. The perceived ease of use of AI systems affects how employees perceive their ability to interpret AI output and integrate it into their normal operational processes. If employees find AI systems too difficult to understand or too opaque, then employee engagement with the systems will decrease, resulting in the underutilization of the systems despite significant strategic investments.

The RBV builds upon TAM by indicating that the effectiveness of perceived usefulness of AI systems is contingent upon the organization having the appropriate internal resources and capabilities. For example, in the case of AI driven fraud detection, these may include: (i) a strong data architecture that supports high-quality data, and that enables the real time collection and integration of data; (ii) an analytical capability that includes the presence of highly skilled data scientists, fraud analysts and IT professionals who can maintain, interpret and update the AI models; and (iii) governance quality that reflects the presence of clear and defined accountability structures, oversight of AI models, and ethical standards and policies for the use of AI systems. As a whole, these resources determine if an AI system will act as a strategic asset or simply as a technical tool. Additionally, empirical evidence indicates that performance-oriented governance mechanisms (such as performance-based budgeting) increase long-term organizational outcomes through improved innovation capability and increased quality of information, which are important for the successful implementation of AI (Alhasnawi et al., 2025).

This study asserts that TAM and RBV are interactive as opposed to independent. Perceived usefulness increases a manager's willingness to make an investment in an AI system; however, sustained adoption and

operational success of the AI system require resource mobilization and the development of capabilities. Conversely, organizations with high levels of resource and capability endowments can amplify the perceived value of an AI system by providing reliable outputs, decreasing false positives, and increasing trust in the use of automated decision-making (Daruwala et al., 2025). This interaction explains the large gap identified in Pakistani banks, where AI systems are typically located at the headquarters of the bank and are underutilized at the branch level due to skill shortages and limited infrastructure.

Finally, this paper uses Institutional Theory to further contextualize the adoption of AI, which states that organizational behavior is influenced by regulatory, normative, and ethical pressures (DiMaggio & Powell, 1983). Regulatory oversight from the State Bank of Pakistan (SBP) is an important factor in the adoption of AI systems in Pakistan's banking industry. The SBP's emerging AI and digital risk governance guidelines place emphasis on transparency, accountability, data privacy, and ethical use of automated systems. These institutional pressures affect not only whether a bank will adopt AI systems, but also how they will design, govern, and implement AI-based fraud detection systems. Therefore, the proposed research model (see Figure 1) views AI adoption as a function of technological perceptions (TAM), organizational resources and capabilities (RBV), and institutional constraints and enablers (Institutional Theory). This positioning is consistent with recent research on AI governance, board oversight, and digital readiness that emphasizes that successful AI deployment in financial institutions is dependent upon strategic oversight, accountability mechanisms, and the preparedness of organizations. Recent studies demonstrate the role of governance structures and leadership in the responsible use of AI, supporting the assertion that AI-driven fraud detection needs to be incorporated into larger digital governance frameworks as opposed to being viewed solely as a technical endeavor.

3. Methodology

The current research utilizes a systematic literature review (SLR) methodology, coupled with qualitative interview data, to provide a complete analysis of artificial intelligence (AI)-driven fraud detection in Pakistan's Banking Sector. Data from both literature reviews and qualitative interviews were deliberately triangulated so as to compare data from the qualitative interviews to literature reviews to enhance analytical rigour, validity and contextualization.

3.1 Search Strategy

A systematic literature search was undertaken using several databases, including: Academic and Industry Databases; Google Scholar; ScienceDirect; SpringerLink; Scopus; Web of Science; as well as Reports from Regulators/Professional Bodies (i.e., the State Bank of Pakistan). The searches used boolean operators (and/or) to combine the search terms: (“Artificial Intelligence” or “AI” or “Natural Language Processing” or “Deep Learning”) and (“Bank Fraud” or “Fraud Detection” or “Financial Fraud”) and (“Emerging Markets” or “Developing Countries” or “Pakistan”); and only articles which had been published between 2003-2025 were considered. This time frame was selected because it reflects the period of significant developments in AI applications in the financial sector over the past 20 years.

3.2 Inclusion and Exclusion Criteria

Studies were included if they: (1) used an artificial intelligence or machine learning based approach to detect fraud and were applied in the banking or financial services sector; (2) were review studies, regulatory/industry reports, or employed an evidence-based methodology; (3) were written in English.

The following criteria were used to exclude studies from the analysis: (1) technical papers on algorithms with no application in finance; (2) duplicate studies; (3) grey literature without clear methodological clarity; (4) studies outside of the domain of banking/financial fraud.

3.3 Study Selection and PRISMA Flow

Beginning with an initial search of 1,246 records from the electronic databases, duplicates were removed and the remaining records underwent a two-step screening process to eliminate irrelevant studies.

A total of 92 studies satisfied the inclusion criteria and thus comprised the studies included in this systematic review; this number was arrived at after a comprehensive evaluation of each study's full text. A detailed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flow diagram can be found in Appendix C.

3.4 Quality Assessment

The quality of all included studies was appraised by a modified version of the Critical Appraisal Skills Programme (CASP) checklist. Quality assessment criteria included the clearness of the research objective(s), the methodological rigor of the study, the adequacy of the data collected, the validity of the findings, and the relevance of the results to the research question(s). All studies were individually reviewed and rated high, medium, or low quality.

To ensure that there was consistency in the review process and to minimize subjectivity bias, quality assessments were cross-checked among reviewers, and any disagreements were resolved through discussion until a consensus was reached. Only studies that were assessed as either high or medium quality were used for the final synthesis.

3.5 Data Extraction and Synthesis

The most important information that was collected from each of the articles selected included the authors of the article, what type of AI technology was used, where the research was focused geographically, how the researchers obtained their data, what methodologies they used to evaluate the data, and what percentage of fraudulent activity AI detected.

The qualitative portion of this study involved analyzing the interview transcripts utilizing a thematic analysis method. The first step in this thematic analysis process was an open coding method to find common ideas or "concepts" associated with the implementation of AI technology for detecting fraudulent activity, along with concepts associated with the use of AI for fraud prevention, concepts associated with governance, and concepts associated with an organization's capabilities. Once the initial round of open coding was completed, the researcher utilized an axial coding method to group similar “codes” (i.e., concepts) into larger “themes.”

In addition to conducting an axial coding method, the researcher attempted to increase the overall rigor of this study by comparing the thematic structure of the study's findings to the findings of the Systematic Literature Review (SLR). This comparison of the two sets of findings was accomplished by creating a map of the thematic structure of the interviews to the expected theoretical and empirical patterns found within the literature on AI for fraud detection. By doing so, the researcher was able to determine whether there was a convergence of the findings from both the SLR and the thematic analysis of the interviews; whether the findings from the interviews were complementary to those found in the literature; and/or whether the findings from the interviews provided context to the findings from the literature. When discrepancies arose between the expected theoretical and empirical patterns found in the literature and the actual findings from the interviews, the researcher chose to analyze these discrepancies rather than treat them as anomalous, thereby increasing the overall interpretive validity of the findings from the study.

3.6 Complementary Interviews

The primary data for this qualitative research project was collected using semi-structured in-depth interviews. This is one of the most commonly used methods in qualitative financial research (Yin, 2011) and has been shown to be effective for collecting rich contextual information. Semi-

structured interviews provide flexibility for the researcher while also ensuring that all relevant questions are asked. Twenty-six interviews were completed with participants selected from a variety of sources within Pakistan's banking system. The sources included commercial banks, Islamic banks, microfinance banks, foreign banks and government owned banks. The participants selected for this study represented various departments within the bank such as: branch managers (n = 9), IT and data analytics staff (n = 6), fraud and compliance officers (n = 5), senior executives (n = 4) and retail banking customers (n = 2). Interviews lasted on average 45 to 75 minutes. All of the interviews followed a flexible interview guide which was developed based on the theoretical model that guided the study. Data collection occurred on an iterative basis. Data collection continued until thematic saturation had been reached. Thematic saturation was defined as the point where successive interviews provided no additional substantially new codes or conceptual insights.

4. Findings and Discussion

4.1 Fraud Detection vs. Fraud Prevention: Clarifying the Distinction

Fraud Detection is used to spot fraudulent behavior at the time of occurrence (or shortly thereafter) through techniques like Anomaly Detection, Supervised Classification, Transaction Monitoring, etc. Fraud Prevention uses techniques to proactively prevent fraudulent behavior from occurring in the first place using techniques like Real-Time Blocking of Suspicious Accounts, Biometric Authentication, Predictive Modeling, etc.

The appendix includes two figures illustrating the types of Machine Learning algorithms and their sub-classes. There are three primary types of Machine Learning: Supervised, Unsupervised, Reinforcement Learning. Each type has several different sub-classifications, which allows for many different models to be used for finding anomalous behavior. In addition to the figures, the appendix also contains tables illustrating common metrics used to measure the performance of Machine Learning Algorithms.

For clarity, Appendix Table A1 presents comparative performance metrics for supervised learning algorithms, Appendix Table A2 reports the results of unsupervised learning methods, and Appendix Table A3 summarizes the performance of deep learning approaches. These appendix tables are placed in Appendix A immediately after the appendix figures.

Table 1 below summarizes the most common techniques used to detect fraud vs. prevent fraud; this table clearly illustrates where each technique fits into the overall process of fraud detection and fraud prevention.

Table 1: AI Techniques in Fraud Detection vs. Fraud Prevention

AI Technique	Application in Fraud Detection	Application in Fraud Prevention	Strengths and Limitations
Supervised ML (Logistic Regression, Decision Trees, Random Forests, SVM, GBM/XGBoost) **Sources:** *Perols (2011); Kumare et al. (2022)*	Classifies transactions as fraudulent/non-fraudulent using historical data	Limited role; may be used for predictive blocking of suspicious accounts	High accuracy; interpretable (some models); however, Needs labeled datasets; risk of overfitting
Unsupervised ML (Clustering, Isolation Forests, Autoencoders, PCA) **Sources:** *Hassija et al. (2024); Kamuangu (2024)*	Detects anomalous/unusual transaction patterns without labels	Early anomaly detection can prevent fraud before losses occur	Works with unlabeled data; suitable for emerging fraud patterns; however, High false positives; less interpretable

Deep Learning (CNNs, RNNs, LSTMs) **Sources:** *Khan et al. (2023); Xie & Li (2023)*	Identifies complex sequential/temporal patterns in transaction streams	Predicts suspicious user behavior to block/prevent future fraud	Handles large/high-dimensional data; adaptive; however, Black-box; resource-intensive
Natural Language Processing (NLP) **Sources:** *Babu (2024); Bernstein (2009); Tounsi and Rais (2018).*	Analyzes customer communications for fraud-related patterns	Prevents fraud by detecting phishing/fake communications in real-time	Processes unstructured text; detects sentiment anomalies; however, Processing large volumes in real-time is challenging
Hybrid Models (Ensemble + Deep Learning + NLP) **Sources:** *Olaoye & Luz (2024); Buhrmester et al. (2021)*	Combine detection methods for higher accuracy.	Prevents fraud through integrated, proactive multi-layered checks	Robust, reduces false positives; however, Complexity, implementation cost

Source: *From the reviewed literature, compiled by the authors*

While there is a wide array of AI tools to support fraud detection and fraud prevention, as shown in Table 1, managerial relevance is dependent less on the complexity or richness of the algorithms used, but more on the organization's preparedness. Similar to previous research (Hassija et al., 2024; Kamuangu, 2024), both unsupervised and hybrid models may have particular appeal for developing countries and emerging markets due to the lack of fraud data with labels. However, even if an organization invests in unsupervised and hybrid models, it will be difficult for the organization to use the models proactively to prevent fraud if the organization does not also invest in training its workforce in how to use the models, integrating the systems necessary to support the models, and providing access to the models at the departmental level (Butt & Aswani, 2021; Moraa, 2025). The most important implication for managers is that they cannot rely solely on AI investment to achieve strategic benefits, they need to make complementary investments in human resources, governance structures, and operational processes to convert technical capabilities into strategic performance (Mikalef et al., 2020; Barney, 1991).

4.2 Case Studies of AI Applications in Banking

Technical literature provides a wealth of information on both machine learning, deep learning and natural language processing as well as hybrid models to be used in fraud management (See Appendix A). However, it is clear that the ways in which each of the above methods are actually being applied practically by banks can vary significantly depending on their specific banking context. As a way of illustrating the variations that exist in the use of AI tools in fraud detection versus fraud prevention, Table 2 lists several case studies from different countries including Pakistan, which provide an overview of the ways in which banks have used AI tools in the areas of fraud detection and/or fraud prevention.

Table 2: Case Studies on AI in Banking Fraud Detection and Prevention

Country/Bank	AI Tool/Approach	Focus (Detection/Prevention)	Key Findings/Outcomes
Pakistan (MCB, Askar Bank)	SmartVista real-time transaction monitoring, anomaly detection	Detection & limited prevention	Improved fraud detection accuracy; prevention constrained by lack of branch-level training (BPC Reports, 2023; Butt & Aswani, 2021)
India (Private Banks)	AI-driven behavioral biometrics, anomaly detection	Prevention (real-time blocking of suspicious logins)	Reduced online/mobile banking fraud cases (Malali & Gopalakrishnan, 2020)

US (Multiple Banks)	ML-based anomaly detection + neural networks	Detection (post-transaction monitoring)	Significant improvement in reducing financial fraud losses (Bello et al., 2023; Hasham et al., 2019)
UK (Retail Banks)	Hybrid AI models integrating ML, NLP, and biometric verification	Prevention (predictive and proactive)	Reduced phishing/identity fraud; better customer trust (Alsayed & Bilgrami,,2017; Dwivedi et al., 2021)
China (FinTech Sector)	Deep learning & graph neural networks	Detection & Prevention	Detected fraud rings via network analysis; proactive fraud blocking (Vassio et al., 2022)

In accordance with Table 2, Anomaly Detection is used by Pakistan-based Banks (e.g., Askari, MCB), in addition to Fraud Detection, however these are rare examples of how Pakistani-based Banks use A.I. for Preventative Measures because they lack Branch Level Expertise. In comparison, India's Banks have employed Behavioural Biometric methods for fraud prevention; whereas, as can be seen from the West, there has been an increase in Hybrid Models being used, which incorporate both Behavioural Biometric and Anomaly Detection Methods. These contrasts show that Organisational Readiness and Regulatory Frameworks greatly influence the Success of Artificial Intelligence Adoption.

4.3 Ethical and Regulatory Challenges

The barriers to implementing an AI-based system for fraud detection also include ethical and legal issues in addition to the technical ones. A number of recent bibliometric analysis have found that there is increasing overlap between AI, corporate social responsibility (ESG) and governance accountability (Alhasnawi et al., 2024; Alshdaifat et al., 2024). As a result, the need to implement formal governance structures regarding the use of AI in high-risk activities has been codified by numerous countries globally through legislation such as the OECD AI Principles and the European Union's AI Act (2024). These are based on several principles, including transparency, accountability, fairness, data protection and human oversight in high-risk AI applications, such as those in financial services. These principles are supported by emerging research suggesting that AI systems used in banking should be implemented using robust governance and oversight mechanisms to facilitate ethical implementation and long-term stakeholder trust (Papagiannidis, 2025; Vukovic, 2025).

Table 3 provides a summary of some of the barriers to implementing an AI-based system for fraud detection and prevention, including data privacy, algorithmic bias, transparency, compliance, and workforce skills. Table 3 also identifies the detection and prevention components of these barriers. These barriers align well with the international governance standards for AI, such as the OECD AI Principles and the EU AI Act (2024), which classify AI systems used in financial risk management as high-risk and therefore require greater levels of explainability, robustness and accountability. Consistent with previous studies, successful AI governance will require bias mitigation mechanisms, transparent model interpretation and secure data management practices to protect consumer confidence (Lim et al., 2021; Wu et al., 2021; Rajasekar et al., 2024).

Table 3: Ethical and Regulatory Challenges in AI-Driven Fraud Management

Challenge	Impact on Fraud Detection	Impact on Fraud Prevention	Regulatory/Ethical Guidance	Key Sources
Data Privacy & Security	Access restrictions limit training data quality	Preventive systems require sensitive customer data	GDPR (EU), SBP AI Guidelines (2024) emphasize anonymization, secure handling	Rajasegar et al. (2024); SBP Reports

Algorithmic Bias	May produce skewed detection results	Preventive systems may unfairly block customers	Calls for bias audits, fairness checks in AI	Akinrinola et al. (2024); Lim et al. (2021)
Transparency & Explainability	Black-box ML/DL models hinder trust	Preventive measures may be contested if the rationale is unclear	SHAP, LIME recommended for explainability	Buhrmester et al. (2021); Wu et al. (2021)
Compliance & Governance	Weak oversight leads to poor detection accuracy	Prevention fails without clear accountability	SBP draft AI guidelines (2025) mandate governance frameworks	State Bank of Pakistan (2024–25)
Workforce Skills Gap	Staff are unable to interpret AI alerts	Prevention is hindered by a lack of AI literacy among branch staff	Training & capacity-building stressed by regulators	Butt & Aswani (2021); Rožman et al. (2017)

In addition to those outlined above (i.e., fraud detection and prevention) Table 3 indicates that Algorithmic Opacity, Data Privacy and Fairness are significant challenges for fraud detection and fraud prevention alike. The results from this study support prior studies demonstrating that organizations will lose trust in "black box" AI systems until they implement explainability mechanisms and auditability controls (Buhrmester et al., 2021; Hassija et al., 2024). As a result, the State Bank of Pakistan's Draft AI Guidelines (2025) address the aforementioned issues by requiring responsible AI governance, model transparency, data protection and accountability within financial institutions (State Bank of Pakistan, 2024-25). The interview respondents also provided insight into how the use of explainable AI tools (e.g., SHAP or LIME) in conjunction with conducting regular fairness and bias audits and strengthening internal data governance would positively impact stakeholders' confidence (regulators, employees, customers) in the organization. These results are consistent with broader institutional research which indicate that an organization's governance maturity and level of ethical oversight has a direct effect on the sustainability of their digital transformation initiatives in highly regulated industries (DiMaggio & Powell, 1983; Hasan et al., 2025).

4.4 Thematic Analysis of Interview Findings

The interview findings were organized into four dominant themes, which also align with insights from the literature.

Theme 1 (Centralization of AI usage): Branch Managers and Compliance Officers have repeatedly indicated that AI Systems are located centrally within Head Offices, thus excluding Branches from being involved with them. This result is consistent with previous studies on how banks in developing countries usually put advanced fraud detection tools in Centralized Units (Chen & Zhang, 2014). From a TAM standpoint, limited branch level interaction with an AI system will reduce perceptions of both usefulness and usability (perceived ease of use), which are two of the primary drivers for technology adoption (Davis, 1989; Venkatesh et al., 2016). From a RBV view, the concentration of these AI systems centrally represents the wasted potential of the organizations' important internal assets- specifically their frontline human capital, as well as their local operational knowledge. The failure to utilize these organizational resources represents a constraint to the organization's ability to develop AI as a strategic capability throughout the entire organization (Barney, 1991; Assensoh-Kodua, 2019).

Theme 2 (Knowledge and training gaps): Interviewees reported that they had little or no knowledge of technology and were inadequately trained on the technical capabilities of AI. IT officials stated that without additional training for their employees at the branch level, they could not effectively use all the technological features available within the AI system. The literature has identified "readiness" as a key barrier to the implementation of AI (Alarfaj et al., 2022). Inadequate training directly impacts the "perceived ease of use" factor of the Technology Acceptance Model (TAM), which impacts employee's confidence in utilizing AI output and applying it to make decisions in their day-to-day work activities (Davis, 1989; Taherdoost, 2021). In addition, from a Resource-Based View (RBV) perspective, limited investment by

banks in the education and skill building of their employees represents a limitation to strategic human capital, thereby hindering a bank's ability to develop sustainable organizational value through the conversion of its AI technological investments (Barney, 1991; Lee & Bruvold, 2019). Structured learning environment and relationships have been shown to be significant moderators of the impact of both digital and AI capability development (Al-Hazaima et al., 2025).

Theme 3 (Regulatory and Ethical Concerns): Concerns from regulatory bodies over data privacy and the lack of a defined regulatory framework also expressed concerns about potential liability resulting from false positives. As discussed by Rezaee (2004), Peterson et al. (2005) in the literature there has been concern over explainability, transparency and clarity within regulatory frameworks. Regulatory ambiguity and uncertainty weaken the perceived usefulness of the system as employees and management are hesitant to rely upon AI generated decisions when they are unable to determine who is accountable for those decisions (Davis, 1989; Wu et al., 2021). The absence of governance frameworks, compliance mechanisms, and ethical oversight is an example of an intangible organizational resource that enables organizations to deploy AI responsibly and at scale (Lim et al., 2021; Papagiannidis, 2025). This would limit both detection and prevention effectiveness.

Theme 4 (Technology–Organization Misalignment): Three themes emerge from this analysis that influence banks' adoption of AI technology in Pakistan. The first theme is centralized control. While bank managers are willing to use and develop AI technology, they rely on it mostly for reporting and compliance purposes rather than as a tool to enhance their day-to-day operations. The second theme is limited training. Bank employees do not have sufficient knowledge or skills to effectively design, develop, deploy, and maintain AI systems. The third theme is regulatory ambiguity. Both bank management and staff are concerned about how to deal with government regulators who have rules and guidelines for using AI technology but there are no specific rules or guidelines available yet. These three themes represent contextual constraints that prevent banks from fully realizing the potential of AI technology.

4.5 Linking Findings to Theoretical Expectations

This subsection explicitly links the empirical themes emerging from the interviews to the study's theory-driven hypotheses and conceptual framework grounded in the Technology Acceptance Model (TAM) and the Resource-Based View (RBV).

Theme 1: Centralization of AI Usage and Hypotheses H1 & H5

The concentration of AI-driven fraud detection at a head office, with little to no involvement from branches, supports H1 and H5. In terms of TAM, limited exposure to an organization's AI system limits the perceived usefulness and ease of use by the front line staff which limits operational usage. As such, the ability of AI to enhance risk-based priority on high-risk transactions (H5) has yet to be realized at the branch level. These results support the theoretical expectation that user acceptance is required for successful utilization of an AI system.

Theme 2: Knowledge and Training Gaps and Hypotheses H1 & H4

Branch staff's lack of knowledge about AI is substantial evidence to support H4 and its emphasis on how an organization's resources influence the success of AI. The lack of trained personnel (human capital) limits the ability to translate AI investment into operational performance as described by the RBV. Additionally, a lack of training negatively influences the ease of use perception (perceived ease of use), and reinforces the interdependence of acceptance at the individual level and capabilities at the firm level, as indicated in the research framework.

Theme 3: Regulatory and Ethical Concerns and Hypothesis H3

The data privacy, explainability, and regulatory accountability concerns are also very close to the H3 as it is stated that institutional and cultural elements can limit or enhance the use of AI (H3). Although the study demonstrates that there is a negative impact on the ability to rely on AI output and thus on its ability to detect and prevent criminal activities due to lack of clarity and uncertainty as to what regulatory responsibilities exist for the development and implementation of AI; this is consistent with the institutional element of the conceptual model that hypothesizes how regulation impacts the deployment of AI.

Theme 4: Technology–Organization Misalignment and Hypotheses H2 & H4

The fragmentation of infrastructure and poor integration among the systems in banking directly relates to H2 and H4. Although AI systems can provide technical support to an organization for detecting anomalies (thus supporting H2), the lack of cross-platform integration of these systems limits the ability of organizations to proactively prevent fraud. Thus, these results illustrate the expectation based on RBV that it is not sufficient to use AI as a single source for fraud prevention. In addition, it will be necessary to have complementary organizational resources and integration capabilities to sustainably reduce fraud.

Synthesis of Theory and Evidence

Pakistani banks remain more centralized and detection-oriented compared to those in other emerging markets with more advanced digital infrastructures and regulations. This is due to differences in regulatory maturity and organization-wide readiness that ultimately impact how AI performs in each respective market (Lim et al., 2021; Dwivedi et al., 2021; Malali & Gopalakrishnan, 2020; Papagiannidis, 2025). Thus, comparative evidence from other emerging markets provides context for the current study's findings. 1) Indian banks use AI to implement both behavioral biometric monitoring and real-time fraud prevention at the customer interface level and are able to do so because of better digital infrastructure and employee training (Malali & Gopalakrishnan, 2020). 2) Banks in Malaysia operate within a more developed and structured governance framework that has created an alignment between AI systems used by the bank and their internal processes (Lim et al., 2021). 3) Banks in Gulf Cooperation Council (GCC) countries are capable of using AI for proactive fraud prevention as well as post-transactional detection because they have invested heavily in digital infrastructure and have received clear guidelines on how to use AI effectively through regulatory bodies (Dwivedi et al., 2021; Papagiannidis, 2025).

4.6 Strategic Alignment Between AI Capabilities and Banking Operations

These results show there is a systemic disconnect between the strategic intentions regarding the implementation of AI for fraud detection and its actual operational execution. Senior managers view AI as a means to implement a strategic control mechanism, with respect to compliance and risk management, while the operational staff at branches have limited opportunity to engage, receive sufficient training or to make decisions using AI-based fraud detection systems. From the perspective of the Resource-Based View (RBV), this disconnection can be seen as an example of an organization not leveraging critical organizational resources (i.e., human capital and embedded knowledge and expertise related to their operational processes). Simultaneously, the Technology Acceptance Model (TAM) provides insight into why the operational staff resist the adoption of AI-based fraud detection, based on the fact that their limited exposure to AI-based fraud detection systems leads them to perceive such systems as being neither useful nor easy to use. Therefore, the strategic commitments made by Headquarters regarding AI-based fraud detection are ultimately not translated into meaningful, day-to-day fraud prevention practices at Branch Operations.

In summary, these results support that AI-based fraud detection should be seen as more than a technology implementation challenge, but rather as a strategic and organizational transformation challenge concerning governance, human capital development and regulatory alignment.

5. Conclusion and Implications

Pakistani Banks can better serve their customers by reducing the amount of time spent on fraud management with AI, improving the security of their systems and building customer trust. Although AI is being used to detect fraudulent transactions through anomaly monitoring and classification; the study shows that the use of AI to prevent fraud is still in its infancy. Therefore, bridging this gap will require more than just an upgrade in technology, but also changes in governance, organizational preparedness, and the training of employees.

This research includes three major contributions. The first contribution of the research is to clarify the difference between fraud detection and fraud prevention which is often overlooked by researchers and

practitioners, providing a more exact analytical tool for those interested in studying and addressing fraud issues. The second contribution of the research is to provide a conceptual framework (Figure 1) that brings together the technical, organizational, and regulatory dimensions of using AI as a fraud management tool, illustrating how moderating variables affect the efficacy of AI-based tools. The third contribution of the research is to provide empirical data collected in Pakistan, allowing for the study to be contextualized in terms of the impact that centralizing the fraud management function, skill deficits and changing regulatory oversight have on the adoption of AI-based fraud management tools.

Strategic Implications: AI fraud management is a strategic investment that needs to be integrated in to your banking organization's overall strategy (not a siloed technology investment). It should also have alignment with senior leadership on an enterprise-wide level so it can be aligned to the broader corporate objective of risk management and customer trust. If AI is not strategically integrated into the banking organizations strategy it will be difficult for the AI to be utilized by the banking organizations employees and will likely remain an unused technology system.

Managerial implications include a decentralization of AI supported monitoring and investing in explanations for the AI systems as well as the creation of cross-functional teams (IT/compliance/branch) that will foster interdepartmental collaboration. Banks also need to provide formalized, cross departmental education/training and literacy programs to increase awareness on the part of frontline staff regarding the AI based tools used for fraud detection and how they may better utilize the outputs from these tools. This will help to build an organization wide sense of the usefulness and ease of use of the technology being adopted.

Policy Implications: Regulators should develop a new generation of clearly defined, comprehensive and internationally compatible regulatory frameworks for AI governance, such as the OECD AI Principles and the EU AI Act. A high degree of coordination among regulators and financial institutions will be necessary to create uniform ethics-related standards for the use of AI (e.g., AI-ethics), for the sharing of data related to the use of AI, and for the accountability associated with AI. Such cooperation can enable banks to have responsibility for data sharing in order to prevent fraud, while maintaining customer confidentiality and their trust in both the banking institution and its technological applications (i.e., AI). The State Bank of Pakistan's draft guidelines for the use of AI in the banking industry (2025) are a very positive first step toward achieving the goal of developing a set of guidelines for the use of AI in the banking industry in Pakistan (State Bank of Pakistan, 2024–25). In addition to creating a framework for the use of AI in the banking industry, it would also be beneficial to require each bank to conduct periodic fairness audits, monitor for bias in AI models, document how each AI model was developed in a manner that is transparent to all stakeholders in the organization, and document the performance metrics used to evaluate the effectiveness of each AI model (Papagiannidis, 2025). As stated previously, these are two factors that influence the long-term benefits that digital technologies (including AI) may provide to a bank's customers and employees (Alshdaifat et al., 2024; Mikalef et al., 2020; Papagiannidis, 2025).

Limitations of this study are: (a) focus on selected banks in Pakistan and (b) reliance on qualitative/secondary literature for empirical data collection (via interviews) rather than qualitative/primary transactional data. While providing rich qualitative/contextual insights, the study did not measure the performance of the models in the actual local banking environment.

Thus, future research directions will be: (1) Testing AI models with real-world transaction data in Pakistan to assess the models' performance within local banking environments; (2) Investigating the impact of employee training, organizational culture and change management practices on the effectiveness of AI systems in detecting fraudulent transactions; (3) Conducting comparative studies across banking markets in South Asia (India, Bangladesh, Sri Lanka etc.) to determine if the results can be generalized to other regulatory and institutional environments; and, (4) Studying customer trust and acceptance of AI-based systems used for fraud detection and prevention – an area of research that is currently underdeveloped.

In summary, AI-based fraud detection and prevention has enormous potential to revolutionize banking services in Pakistan, but achieving this goal will require a balanced approach that incorporates robust governance, regulatory support, education/training for banking employees and collaborative relationships

with stakeholders. In addition, as this study distinguishes clearly between detection and prevention while also including contextual issues in a conceptual framework, it contributes both to the theoretical understanding of the application of AI-based fraud detection and prevention systems in banking services and to the development of practical approaches for developing sustainable and trustworthy banking systems.

References

Adelakun, B. O. (2023). How Technology Can Aid Tax Compliance in the US Economy. *Journal of Knowledge Learning and Science Technology*, *2*(2), 491-499. ISSN: 2959-6386 (online).

Akinrinola, O., Okoye, C. C., Ofodile, O. C., & Ugochukwu, C. E. (2024). Navigating and reviewing ethical dilemmas in AI development: Strategies for transpwill bency, fairness, and accountability. *GSC Advanced Research and Reviews*, *18*(3), 050-058.

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. *IEEE Access*, *10*, 39700-39715.

Alhasnawi, M. Y., Alshdaifat, S. M., Aziz, N. H. A., & Almasoodi, M. F. (2024). Artificial Intelligence and Environmental, Social and Governance: A Bibliometric Analysis Review. In International Conference on Explainable Artificial Intelligence in the Digital Sustainability (pp. 123-143). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-63717-9_8

Alhasnawi, M. Y., Alshdaifat, S. M., Mansour, M., Saleh, M. W., & Hu, G. (2025). How does performance-based budgeting enhance sustainable performance? A mediated-moderated model of innovation and information quality. *International Journal of Innovation Science*, 1-19.

Al-Hazaima, H., Alduneibat, K. A., Alhasnawi, M. Y., Alshdaifat, S. M., Hu, G., & Al-Shbiel, S. O. (2025). Exploring sustainability accounting education in the digital era with sustainability leadership as a moderating factor using a two-stage approach PLS-SEM–ANN. *Discover Sustainability*, *6*(1), 600.

Al-Maaitah, T. A., Alduneibat, K. A., Alshdaifat, S. M., Alsarayreh, R., Bani Ahmad, A. Y. A., & Hijazin, A. (2025). AI adoption, technological readiness, and AI usability in sustainability accounting education: The moderating role of academic integrity. *Heritage and Sustainable Development, 7*(1), 611–628. https://doi.org/10.37868/hsd.v7i1.1176

Alodat, A., & Mansour, M. (2024). Board characteristics and cybersecurity disclosure: evidence from the UK. *Available at SSRN 5137138*.

Alsarayreh, R., Buraik, O., Hijazin, A., & Alshdaifat, S. M. (2025). The Impact of Business Intelligence on Strategic Ambidexterity: The Mediating Role of Knowledge Sharing. *Electronic Journal of Business Research Methods, 23*(1), 20-34. https://doi.org/10.34190/ejbrm.23.1.3719

Alsarayreh, R., Buraik, O., Hijazin, A., & Alshdaifat, S. M. (2025). The Impact of Business Intelligence on Strategic Ambidexterity: The Mediating Role of Knowledge Sharing. *Electronic Journal of Business Research Methods, 23*(1), 20-34. https://doi.org/10.34190/ejbrm.23.1.3719

Alsayed, A., & Bilgrami, A. (2017). E-banking security: Internet hacking, phishing attacks, analysis and prevention of fraudulent activities. *International Journal of Emerging Technology and advanced engineering*, *7*(1), 109-115.

Alshdaifat, S. M., Aziz, N. H. A., Alhasnawi, M. Y., Alharasis, E. E., Al Qadi, F., & Al Amosh, H. (2024). The Role of Digital Technologies in Corporate Sustainability: A Bibliometric Review and Future Research Agenda. *Journal of Risk and Financial Management, 17*(11), 509. https://doi.org/10.3390/jrfm17110509

Al-Tahat, S., Bani-Khaled, S., Jaradat, Z., Mansour, M., & Al-zoubi, A. M. (2025). State ownership as a moderator in the relationship between board characteristics and ESG performance: Evidence from Asia-Pacific markets. *Journal of Business and Socio-economic Development*, 1-22.

Assensoh-Kodua, A. (2019). The resource-based view: A tool of key competency for competitive advantage. *Problems and Perspectives in Management*, *17*(3), 143.

Babu, C. S. (2024). Adaptive AI for dynamic cybersecurity systems: Enhancing protection in a rapidly evolving digital landscape. In *Principles and Applications of Adaptive Artificial Intelligence* (pp. 52-72). IGI Global.

Bello, O. A., Ogundipe, A., Mohammed, D., Adebola, F., & Alonge, O. A. (2023). AI-Driven Approaches for Real-Time Fraud Detection in US Financial Transactions: Challenges and Opportunities. *European Journal of Computer Science and Information Technology*, *11*(6), 84-102.

Bernstein, D. J. (2009). Introduction to post-quantum cryptography. In *Post-quantum cryptography* (pp. 1-14). Springer Berlin Heidelberg.

Bhatla, T. P., Prabhu, V., & Dua, A. (2003). Understanding credit card frauds. *Cards Business Review*, *1*(6), 1-15.

Bozkus Kahyaoglu, S., & Caliyurt, K. (2018). Cyber security assurance process from the internal audit perspective. *Managerial Auditing Journal*, *33*(4), 360-376.

Buhrmester, V., Münch, D., & Will bens, M. (2021). Analysis of explainers of black box deep neural networks for computer vision: A survey. *Machine Learning and Knowledge Extraction*, *3*(4), 966-989.

Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, *275*, 314-347.

Daruwala, Z., Khan, F., & Ullah Jan, S. (2025). Unpacking leverage and lagged effects: do digitally mature and complex firms outpace their rivals?. *Cogent Economics & Finance*, *13*(1), 2577909.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160.

Hasan, E. F., Alzuod, M. A., Al Jasimee, K. H., Alshdaifat, S. M., Hijazin, A. F., & Khrais, L. T. (2025). The Role of Organizational Culture in Digital Transformation and Modern Accounting Practices Among Jordanian SMEs. *Journal of Risk and Financial Management*, *18*(3), 147. https://doi.org/10.3390/jrfm18030147

Hasham, S., Joshi, S., & Mikkelsen, D. (2019). Financial crime and fraud in the age of cybersecurity. *McKinsey & Company*, 1-11.

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., & Hussain, A. (2024). Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation*, *16*(1), 45-74.

Homer, E. M. (2020). Testing the fraud triangle: a systematic review. *Journal of Financial Crime*, *27*(1), 172-187.

Jang-Jaccard, J., & Nepal, S. (2014). A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences*, *80*(5), 973-993.

Kamuangu, P. (2024). A review on financial fraud detection using AI and machine learning. *Journal of Economics, Finance and Accounting Studies*, *6*(1), 67-77.

Kaya, O., Schildbach, J., AG, D.B., & Schneider, S. (2019). Artificial intelligence in banking. *Artificial intelligence*.

Khan, F., Ullah Jan, S., & Zia-ul-haq, H. M. (2025). Artificial intelligence adoption, audit quality and integrated financial reporting in GCC markets. *Asian Review of Accounting*, *33*(3), 464-495.

Khan, W. Z., Raza, M., & Imran, M. (2023). Quantum cryptography a real threat to classical blockchain: Requirements and challenges. *Authorea Preprints*.

Kumar, G., Muckley, C. B., Pham, L., & Ryan, D. (2018). *Can Alert Models for Fraud Protect the Elderly Clients of a Financial Institution? Michael J* (No. 18-16). Brennan Irish Finance Working Paper Series Research

Kumar, S., Ahmed, R., Bharany, S., Shuaib, M., Ahmad, T., Tag Eldin, E., ... & Shafiq, M. (2022). Exploitation of machine learning algorithms for detecting financial crimes based on customers' behavior. *Sustainability*, *14*(21), 13875.

Kuvaas, B. (2006). Performance appraisal satisfaction and employee outcomes: mediating and moderating roles of work motivation. *The International Journal of Human Resource Management*, *17*(3), 504-522.

Lee, C. H., & Bruvold, N. T. (2019). Creating Value for Employees: Investment in Employee Development. International Journal of Human Resource Management, 14(6), 981-1000.

Malali, A. B., & Gopalakrishnan, S. (2020). Application of Artificial Intelligence and Its Powered Technologies in the Indian Banking and Financial Institutions: An Overview. *IOSR Journal Of Humanities And Social Science*, *25*(4), 55-60.

Mansour, M., Abu-Allan, A. J., Alshdaifat, S. M., E'leimat, D. A., & Saleh, M. W. (2025). Board effectiveness and carbon emission disclosure: evidence from ASEAN countries. *Discover Sustainability*, *6*(1), 604.

Mansour, M., Saleh, M. W., Marashdeh, Z., Marei, A., Alkhodary, D., Al-Nohood, S., & Lutfi, A. (2024a). Eco-innovation and financial performance nexus: does company size matter?. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(1), 100244.

Michie, S., & Williams, S. (2019). Reducing Psychological Ill Health and Associated Sickness Absence: A Systematic Literature Review. Occupational and Environmental Medicine,60,3-9.

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. A. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management*, *57*(2), 103169.

Mohanty, B., & Mishra, S. (2023). Role of Artificial Intelligence in Financial Fraud Detection. *Academy of Marketing Studies Journal*, *27*(S4).

Olaoye, G., & Luz, A. (2024). Hybrid Models for Medical Data Analysis. *Available at SSRN 4742530*.

Perols, J. L. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice & Theory*, 30(2), 19-50.

Rajasegar, R. S., Gouthaman, P., Ponnusamy, V., Arivazhagan, N., & Nallarasan, V. (2024). Data Privacy and Ethics in Data Analytics. In *Data Analytics and Machine Learning: Navigating the Big Data Landscape* (pp. 195-213). Singapore: Springer Nature Singapore.

Rezaee, Z. (2004). Restoring public trust in the accounting profession by developing anti-fraud education, programs, and auditing. *Managerial Auditing Journal*, *19*(1), 134-148.

Rožman, M., Treven, S., & Čančer, V. (2017). Motivation and Satisfaction of Employees in the Workplace. Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy, 8(2), 14-25.

State Bank of Pakistan. (2024). Digital Financial Services and Responsible AI: Emerging Regulatory Considerations. SBP Policy Review Report.

Taherdoost, H. (2021). A review on risk management in information systems: Risk policy, control and fraud detection. *Electronics*, *10*(24), 3065.

Tounsi, W., & Rais, H. (2018). A survey on phishing detection techniques. *Journal of Information Security and Applications*, 41, 1-11.

Vassio, L., Garetto, M., Leonardi, E., & Chiasserini, C. F. (2022). Mining and modelling temporal dynamics of followers' engagement on online social networks. *Social Network Analysis and Mining*, *12*(1), 96.

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.

Vieira, A., & Sehgal, A. (2018). How banks can better serve their customers through artificial techniques. In *Digital marketplaces unleashed* (pp. 311-326). Springer, Berlin, Heidelberg.

Wells, J. T. (2004). New approaches to fraud deterrence. *JOURNAL OF ACCOUNTANCY-NEW YORK-*, *197*(2), 72-76.

Xie, B., & Li, Q. (2023). Detecting fake news by RNN-based gatekeeping behavior model on social networks. *Expert Systems with Applications*, *231*, 120716.

Zhang, W., Zhao, S., Wan, X., & Yao, Y. (2021). Study on the effect of digital economy on high-quality economic development in China. *PloS one*, *16*(9), e0257365.

Ethical Approval: This study involved human participants through semi-structured interviews. Ethical approval was obtained from the relevant institutional ethics committee, approval number IRB/2025/123, dated 30 March 2025. All procedures were conducted in accordance with the relevant institutional ethical requirements and applicable national guidelines and regulations.

Informed Consent: Informed consent was obtained from all participants before the interviews. Participants were informed about the purpose of the study, the voluntary nature of participation, confidentiality protections, and their right to withdraw at any stage without penalty.

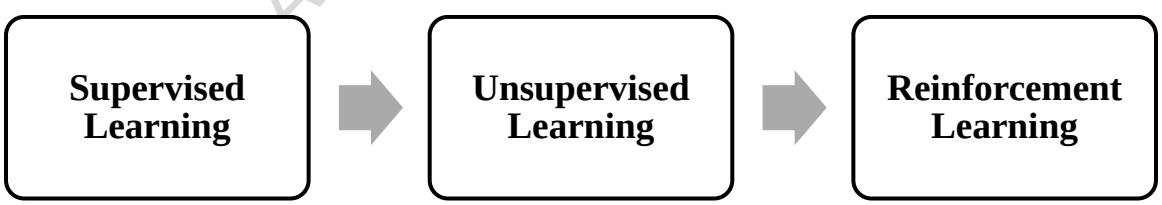
Data Availability: This study draws on two sources: a systematic literature review and qualitative interviews. The literature review search strategy, screening criteria, study-selection approach, and data-extraction framework can be made available to the editors and reviewers on request. Interview transcripts and institution-identifying materials are not publicly available because they contain confidential information from participating banking professionals and institutions. De-identified supporting materials may be shared for editorial review on reasonable request, subject to confidentiality obligations and institutional approval.

Funding: This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests: The authors declare no competing interests, financial or non-financial, that could have influenced the work reported in this manuscript.

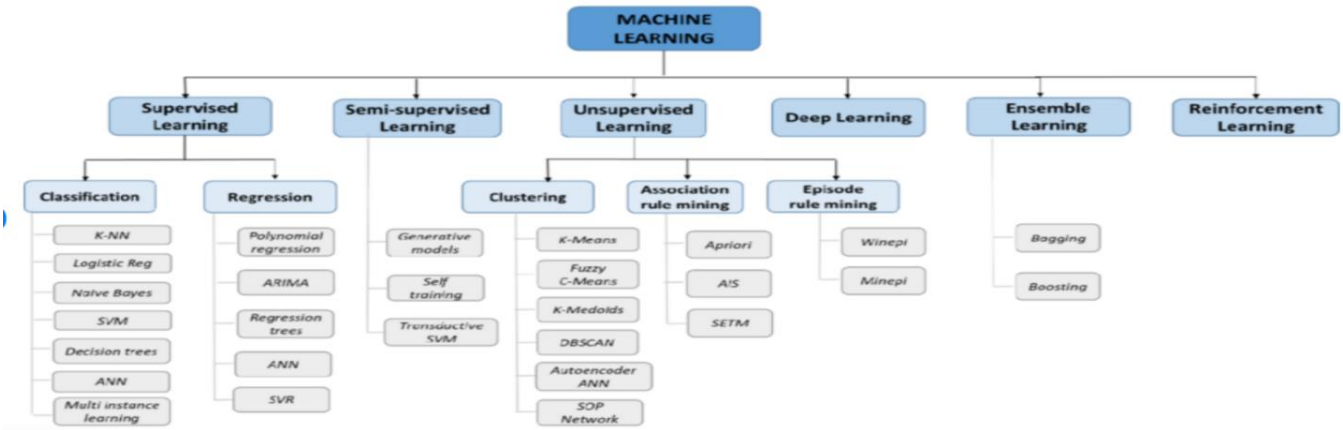
Appendix A: Technical Details on AI-Algorithms

Appendix Figure A1: *Machine Learning Algorithms*



Description: Taxonomy of supervised, unsupervised, and reinforcement learning models commonly applied in fraud detection and prevention.

Appendix Figure A2: *Machine Learning Algorithms with Sub-Classifications*



Description: Detailed breakdown of supervised (e.g., decision trees, logistic regression), unsupervised (e.g., clustering, isolation forests), and deep learning models (e.g., CNNs, RNNs, autoencoders) relevant to fraud management.

Appendix Table A1 appears here and reports the comparative performance of supervised learning algorithms used in fraud detection.

Table A1: Results for Supervised Learning Algorithms (Kamuangu, 2024)

Supervised Algorithm	Accuracy	Precision	Recall	F1 Score	AUC ROC
Logistic regression	0.92	0.89	0.85	0.87	0.94
Decision trees	0.94	0.91	0.88	0.89	0.96
SVM	0.93	0.90	0.87	0.88	0.95
GBM	0.95	0.93	0.91	0.92	0.97

Appendix Table A2 appears here and summarizes the results of unsupervised learning methods relevant to fraud detection.

Table A2: Results for Unsupervised Learning Methods (Kamuangu, 2024)

Unsupervised Method	Accuracy	Slhouette Score	Auc Roc
K-Means Clustering	0.85	0.60	0.88
Isolation Forests	N/A	N/A	0.92
DBSCAN	N/A	N/A	0.87
Autoencoders	N/A	N/A	0.94

Appendix Table A3 appears here and presents the reported performance of deep learning approaches used in fraud detection.

Table A3: Results for Deep Learning Approaches (Kamuangu, 2024)

Deep Learning Approaches	Accuracy	Precision	Recall	F1 Score	AUC ROC
Neural Networks	0.94	0.91	0.88	0.89	0.96
CNNS	0.95	0.93	0.90	0.91	0.97
RNNS/LSTMS	0.93	0.90	0.87	0.88	0.95
Autoencoders	0.95	0.94	0.92	0.93	0.98

Appendix B: Additional Technical Challenges in AI Fraud Detection

In addition to the algorithmic details and performance metrics presented above, the literature and interview findings highlight several technical challenges that complicate the implementation of AI-driven fraud detection in banking:

- **Data Quality and Availability:** Incomplete, imbalanced, or noisy datasets limit the effectiveness of machine learning models, especially in environments where fraudulent cases are under-reported.
- **Model Reliability and Accuracy:** Even high-performing algorithms produce false positives and false negatives, which can erode user confidence and increase operational costs.
- **System Integration Issues:** Many banks lack the IT infrastructure to integrate AI tools with legacy systems, creating bottlenecks in real-time fraud monitoring.
- **Interpretability and Transparency:** Deep learning and ensemble methods often act as "black boxes," making it difficult for managers to understand how decisions are generated.
- **Compliance and Governance:** Regulatory frameworks demand explainability, auditability, and accountability, which opaque AI models do not easily meet
- **Workforce Skills Gap:** Branch managers and frontline staff often lack training in AI systems, which prevents effective use and limits proactive fraud prevention.

These challenges overlap with the ethical and regulatory concerns summarized in Table 3 of the main text. While technical improvements (e.g., better feature engineering, hybrid models, real-time analytics) may address some of these issues, organizational readiness and regulatory clarity remain equally critical for effective fraud management.

Appendix C: PRISMA Flow Diagram

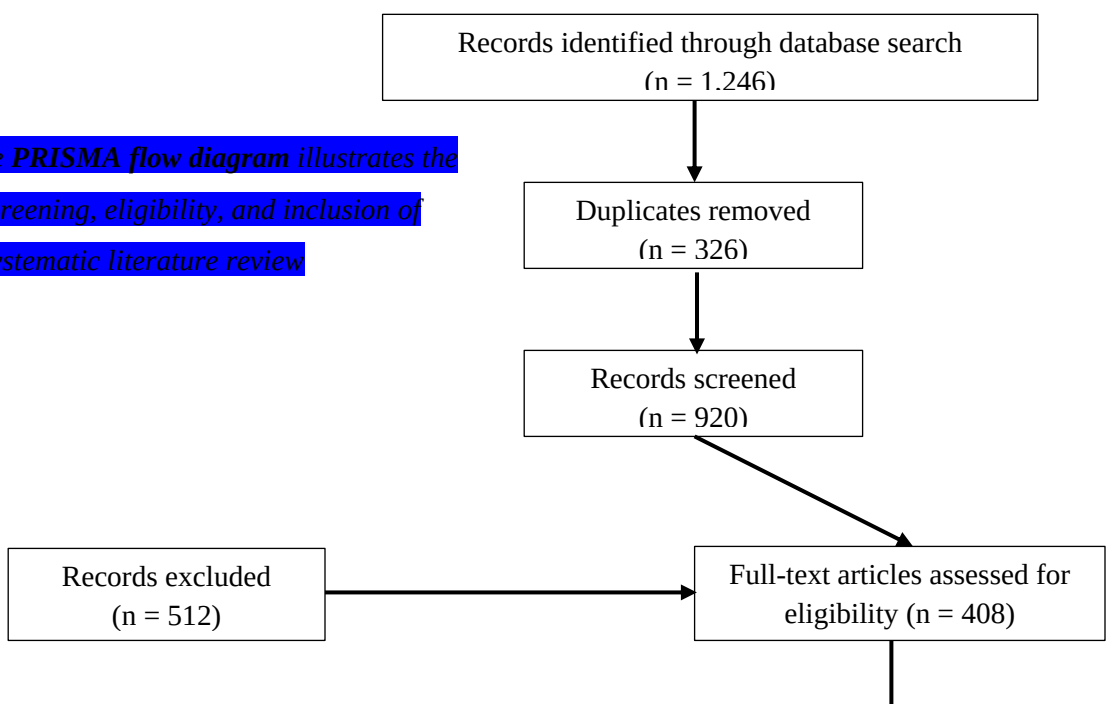


Figure C1: *The **PRISMA flow diagram** illustrates the identification, screening, eligibility, and inclusion of studies for the systematic literature review*

Appendix D: Summary of Interview Themes and Illustrative Quotations

This appendix provides a concise overview of the main themes derived from the qualitative interviews, along with illustrative quotations to enhance transparency and analytical credibility.

Theme	Description	Illustrative Quotation
Centralization of AI Usage	AI-based fraud detection systems are primarily managed at the head-office level, with limited branch-level autonomy.	"We rely on instructions from the central monitoring unit; AI tools are not something we directly operate at the branch." (Branch Manager)
Knowledge and Training Gaps	Limited AI literacy among operational staff constrains the effective use of AI outputs.	"Most staff do not fully understand how AI flags transactions, so they hesitate to rely on it." (Compliance Officer)
Regulatory and Ethical Concerns	Uncertainty around accountability, explainability, and data privacy reduces trust in AI systems.	"If an AI system blocks a customer, it is not always clear who is responsible." (Senior Manager)
Technology–Organization Misalignment	Weak system integration limits real-time fraud prevention.	"Detection works, but prevention is difficult because systems don't fully talk to each other." (IT Specialist)