Artificial Intelligence in Digital Banking: A Systematic Literature Review and a Novel Cross-Domain Resilience Framework for Customer Experience, Fraud Detection, and Risk Management

Durga Prasad Dāsepalli [0009-0008-7681-3258]

[1] Senior Technical Architect, Perficient, Melissa, TX, USA; Alumni, Cleveland State University, Cleveland, OH, USA

*Corresponding author. Email: ddasepalli@gmail.com

Abstract

Artificial intelligence (AI) is reshaping digital banking across three functionally distinct but operationally entangled domains: customer experience, fraud detection, and enterprise risk management. Prior reviews have generally treated these domains, and the cloud-native architectures that host them, as separate literatures, producing fragmented guidance for banks that must in practice run all three over a single, shared technical substrate. This paper reports a PRISMA 2020-guided systematic literature review of 36 peer-reviewed studies published primarily between 2020 and 2025, supplemented by a small number of foundational earlier works, and screened through a structured identification, screening, eligibility, and quality-appraisal process. The review synthesizes evidence across four literature streams — AI-driven customer experience, machine-learning-based fraud detection, explainable credit and enterprise risk analytics, and cloud-native microservices architecture — and identifies three compounding gaps: domain fragmentation, an explainability asymmetry that concentrates transparency research on credit decisions while leaving fraud and personalization comparatively opaque, and an architecture–application disconnect between enterprise-architecture research and AI-application research. To address these gaps, the paper proposes an enhanced conceptual framework, the AI-Enabled Digital Banking Resilience Model, version 2.0 (AIDBRM 2.0), which extends prior architecture-only and domain-only frameworks with three novel cross-cutting mechanisms: a Unified Continuous Trust Score that fuses fraud, credit, and customer-experience signals into one auditable index; a Generative AI Governance Sandbox for pre-production testing of large language model outputs; and a Federated Cross-Institution Learning module for privacy-preserving joint model improvement. A structured comparison against prior architecture-centric and taxonomy-centric frameworks demonstrates the model's distinguishing contribution. Because no new primary data were collected, the manuscript is explicitly positioned as a systematic literature review with a novel, empirically-groundable but not yet empirically-validated conceptual framework, and it concludes with an explicit limitations section and a future-research agenda organized around empirically testing each AIDBRM 2.0 component.

Keywords: Artificial Intelligence; Digital Banking; Fraud Detection; Risk Management; Customer Experience; Explainable AI; Cloud-Native Microservices; PRISMA Systematic Review; Generative AI Governance; Federated Learning

1. Introduction

Banking is not undergoing a second wave of digitization; it is undergoing a change in the kind of system that sits behind the interface. The first wave of digital transformation replaced paper processes with digital forms and web portals. The current wave replaces static, rule-based digital infrastructure with adaptive systems that sense, learn, and act on financial data continuously, often without a human in the loop for any single decision (Barroso & Laborda, 2022). This shift is being accelerated by competition from fintech entrants and challenger banks whose value proposition is built around AI-native experiences rather than retrofitted onto legacy cores, which raises the competitive cost of incumbents treating AI as a peripheral capability rather than an architectural commitment.

Three domains of digital banking have absorbed most of this transformation, and it is important to state precisely how the literature has treated them, because the treatment itself is part of the problem this paper addresses. Customer-experience research has established that conversational AI, robo-advisory, and hyper-personalization measurably improve satisfaction and continuance intention, but the effect is conditional on trust and communication quality rather than on automation coverage (Alnaser et al., 2023; Eren, 2021; Nicolescu & Tudorache, 2022). Fraud-detection research has moved from static rule engines to adaptive, increasingly hybrid supervised–unsupervised pipelines capable of detecting account takeover, synthetic identity fraud, and transaction laundering in near real time (Ali et al., 2022; Hilal et al., 2021), yet the same literature repeatedly documents that purely transactional models under-perform against fraud typologies that exploit human intermediaries, such as money-mule networks (Abdul Rani et al., 2024; Akinbowale et al., 2020). Credit- and enterprise-risk research has converged on explainable ensemble methods that preserve predictive accuracy while producing regulator- and borrower-legible explanations (Bussmann et al., 2021; de Lange et al., 2022), a level of explainability maturity that, as this review will show, has no equivalent in the fraud-detection or personalization literatures.

The problem is that these three domains are studied as if they were independent, when in the operating reality of a bank they are not. A single customer session can simultaneously trigger a conversational-AI response, a real-time fraud score, and a change in exposure or affordability calculation, and all three draw on overlapping identity, transaction, and behavioral data flowing through the same event-driven, cloud-native, microservices-based infrastructure (Maestro & Surantha, 2023; Blinowski et al., 2022). Enterprise-architecture research on that infrastructure — evaluating scalability, inter-service latency, and fault isolation (Weerasinghe & Perera, 2023; Meijer et al., 2024) — has, in turn, developed with almost no cross-reference to the AI-application literature it is meant to host. Broader information-systems research supports the underlying claim: organizational and architectural readiness, not algorithmic sophistication alone, is frequently the binding constraint on successful enterprise AI adoption (Collins et al., 2021), a finding that banking-specific studies echo but rarely operationalize into a concrete cross-domain architecture.

A further asymmetry compounds the fragmentation problem. Explainable AI (XAI) techniques — SHAP, LIME, and related feature-attribution methods — have been extensively validated in credit scoring, where regulatory pressure is oldest and most codified (Bussmann et al., 2020, 2021; de Lange et al., 2022; Gramespacher & Posth, 2021), and the taxonomic literature on XAI more broadly (Barredo Arrieta et al., 2020) provides the conceptual vocabulary this maturity draws on. Fraud detection and customer-experience personalization, by contrast, routinely deploy opaque deep-learning and recommendation architectures with markedly less transparency infrastructure (Hernandez Aros et al., 2024; George et al., 2023), even though a fraud hold or a personalized product recommendation materially affects a customer in much the same way a credit decision

does, and increasingly falls within the scope of the same regulatory expectations (Aziz & Andriansyah, 2023).

This paper makes three contributions that respond directly to this fragmentation. First, it conducts a PRISMA 2020-guided systematic literature review — rather than a narrative overview — of the AI-in-digital-banking literature across the four streams described above, applying explicit eligibility criteria, a documented screening process, and a structured quality appraisal (Section 4). Second, it identifies and substantiates a compounded research gap consisting of domain fragmentation, explainability asymmetry, and an architecture–application disconnect (Section 3), each supported by specific contradictions and omissions in the reviewed literature rather than by assertion. Third, it proposes an enhanced conceptual framework, the AI-Enabled Digital Banking Resilience Model, version 2.0 (AIDBRM 2.0), which advances beyond the architecture-only frameworks of Maestro and Surantha (2023) and Blinowski et al. (2022) and the domain-taxonomy frameworks of Cao (2022) by introducing three cross-cutting mechanisms — a Unified Continuous Trust Score, a Generative AI Governance Sandbox, and a Federated Cross-Institution Learning module — that are structurally absent from prior proposals (Section 5). Because this paper synthesizes existing peer-reviewed evidence rather than generating new empirical data, it is positioned explicitly as a systematic literature review that yields a novel, testable conceptual framework, not as an empirical validation study; Section 8 states this positioning directly, and Section 7 turns each untested component of AIDBRM 2.0 into a specific future-research question.

The remainder of the paper is organized as follows. Section 2 presents the critical literature review across the four streams. Section 3 substantiates the research gap and states the paper's novelty claims explicitly. Section 4 details the PRISMA-guided methodology, including the flow diagram, eligibility criteria, and quality appraisal. Section 5 presents AIDBRM 2.0 and a structured comparison against prior frameworks. Section 6 synthesizes results with cross-domain critical analysis. Section 7 states limitations. Section 8 sets out a framework-anchored future-research agenda, and Section 9 concludes.

2. Literature Review: A Critical Synthesis

This section moves beyond describing what each study found to evaluating how well each stream of research supports practical, cross-domain decision-making in a bank. Each subsection therefore closes with an explicit appraisal of methodological limitations and unresolved contradictions, rather than a restatement of reported findings.

2.1 Artificial Intelligence and Customer Experience

The dominant explanatory model in this stream is expectation-confirmation theory, applied by Alnaser et al. (2023) to show that perceived performance, communication quality, and corporate reputation predict digital-banking user satisfaction. Eren (2021), studying a single Turkish banking chatbot, found that perceived usefulness, responsiveness, and trust drive satisfaction — a result that converges with Alnaser et al. (2023) on the centrality of trust but was derived from a single-institution, single-country sample, which limits claims of generalizability across regulatory and cultural contexts. Nicolescu and Tudorache's (2022) systematic review reinforces anthropomorphic design and conversational naturalness as recurring predictors across service industries, while Chen et al. (2021) and Misischia et al. (2022) extend the pattern into e-commerce and general customer-service chatbots respectively, and Jenneboer et al. (2022) link chatbot quality specifically to loyalty rather than only satisfaction.

Read critically, this literature converges on a conditional rather than unconditional claim: AI improves customer experience only when trust, escalation design, and communication quality are present, and several of the same studies caution that poorly calibrated agents — those without escalation paths or local-language support — actively erode the efficiency gains they are meant to produce (Nicolescu & Tudorache, 2022; Misischia et al., 2022). Two limitations recur across this stream. First, most primary studies are cross-sectional, self-reported, and drawn from a single national or institutional context (Eren, 2021; Alnaser et al., 2023), which constrains inference about durability of the reported satisfaction effects over time — a gap this paper returns to in Section 7. Second, none of the reviewed customer-experience studies incorporate an explainability or auditability construct, even though personalization and recommendation models are typically as opaque as the fraud-detection models discussed in Section 2.2; this is the first concrete instance of the explainability asymmetry substantiated in Section 3.

2.2 Artificial Intelligence in Fraud Detection

Two large syntheses anchor this stream. Ali et al. (2022), following the Kitchenham systematic-review protocol, found that supervised classifiers (decision trees, random forests, gradient-boosted models) dominate financial fraud detection because of their established predictive power, but flagged their comparatively low interpretability. George et al. (2023), synthesizing 118 peer-reviewed papers under PRISMA, found that recurrent and convolutional deep-learning architectures are increasingly competitive for modeling sequential transaction data, but that interpretability and real-time deployment remain unresolved — a direct echo of Ali et al.'s (2022) concern despite the two reviews using different inclusion criteria and time windows, which strengthens confidence in the finding rather than merely repeating it.

At the empirical level, Afriyie et al. (2023) demonstrated strong discriminatory performance for random forest and logistic regression on labeled credit-card fraud data, while Alshameri and Xia (2023) showed that SMOTE-based resampling meaningfully improves sensitivity to the minority fraud class without materially degrading precision — an important methodological result given how imbalanced fraud datasets typically are. Hilal et al. (2021) found that hybrid supervised–unsupervised methods generally outperform single-paradigm systems in real-world deployment, and Bello et al. (2023) examined the operational benefits of real-time scoring pipelines specifically within transaction-processing infrastructure, which is one of the few studies in this stream to engage with deployment architecture rather than model accuracy alone.

A separate and, on critical reading, under-integrated line of research shows that transactional models have a structural blind spot. Akinbowale et al. (2020) found that a balanced-scorecard view of cybercrime impact requires governance and staff-training elements that purely technical fraud models cannot supply, and Abdul Rani et al. (2024) showed that money-mule-based fraud typologies exploit human intermediaries in ways that transaction-level features alone are unlikely to capture without behavioral and network-based features. Hernandez Aros et al. (2024) confirms, across a broad literature synthesis, that the field is moving toward ensemble and hybrid models faster than it is moving toward explainability and regulatory alignment — precisely the imbalance this paper identifies as a research gap rather than a settled finding, since none of the fraud-detection studies reviewed here report a production-grade explainability mechanism comparable to those documented in credit risk (Section 2.3).

Critically, the fraud-detection literature also shows a methodological split that is rarely acknowledged within individual papers: systematic reviews (Ali et al., 2022; George et al., 2023) report interpretability as an open problem in the aggregate, while empirical papers (Afriyie et al., 2023; Alshameri & Xia, 2023) report accuracy and class-imbalance metrics without engaging interpretability at all. This is not a contradiction so much as a gap in the research agenda: no

reviewed study jointly reports fraud-detection accuracy and an explainability metric on the same model, which is the specific empirical absence that Section 5's Explainability and Governance Layer is designed to close.

2.3 Artificial Intelligence in Credit Risk and Enterprise Risk Management

This stream is the most methodologically mature of the three domains, largely because explainability requirements have been a regulatory feature of credit decisioning for longer than they have been a feature of fraud or personalization decisioning. Bussmann et al. (2020, 2021) established, using SHAP, that complex ensemble models such as XGBoost can be made interpretable enough for regulatory and operational use in credit scoring. De Lange et al. (2022), using a Norwegian bank case, extended this by decomposing credit decisions into feature-level contributions legible to credit officers and auditors, and Hjelkrem et al. (2022), in the same institutional setting, showed that open-banking transaction data materially improves the discriminatory power of application credit scoring relative to bureau-only models — though this benefit is explicitly conditional on data-sharing infrastructure and customer consent mechanisms that not every jurisdiction has in place.

A methodological sub-literature addresses the complexity–interpretability trade-off directly rather than treating it as binary. Dumitrescu et al. (2022) proposed logistic-regression extensions that incorporate non-linear decision-tree effects, occupying a middle position between fully transparent linear models and opaque ensembles. Chen et al. (2024) showed that interpretable ML techniques can remain both fair and discriminatory under skewed default distributions, addressing class imbalance in a manner structurally analogous to Alshameri and Xia's (2023) SMOTE work in fraud detection — a cross-domain methodological parallel that the reviewed literature does not itself draw, again illustrating domain fragmentation rather than a genuine absence of shared technique. Gramespacher and Posth (2021) extended explainability from individual loan decisions to portfolio-level return optimization, and Fritz-Morgenthal et al. (2022), drawing on practitioner round tables, argued that model-risk-management frameworks must evolve to govern AI/ML models with the same rigor as statistical models, particularly around validation, monitoring, and drift detection.

At a broader level, Milojević and Redzepagic (2021) found that AI improves predictive risk monitoring and early-warning capability but that data quality, legacy-system integration, and skills gaps remain binding constraints — a finding that anticipates, without fully connecting to, the architecture literature reviewed in Section 2.4. Cao's (2022) taxonomy situates credit and enterprise risk management as one of several AI-in-finance use cases and notes that risk-related systems are the most heavily monitored of all financial AI applications, a claim this review treats as a useful taxonomic anchor rather than as an architecture proposal, since Cao (2022) does not specify implementation-level mechanisms (Section 5.3 makes this distinction explicit in the framework comparison).

The principal limitation of this stream, considered critically, is generalizability: the two most detailed explainability case studies (de Lange et al., 2022; Hjelkrem et al., 2022) are both drawn from a single Norwegian banking context, and Bussmann et al. (2020, 2021) similarly report a small number of institutional applications. The explainability techniques themselves (SHAP, LIME) are well validated in principle (Barredo Arrieta et al., 2020), but their reported success in production credit scoring has not been replicated at comparable depth in fraud detection or personalization, which is the specific asymmetry this paper positions as a primary research gap rather than three independent, unrelated observations.

2.4 Cloud-Native and Microservices Architectures Enabling AI in Banking

Maestro and Surantha (2023) found that containerized, independently scalable microservices significantly outperform monolithic architectures under high transaction load — a finding directly relevant to real-time fraud scoring and conversational AI, though based on evaluation in a constrained cloud test environment rather than production banking traffic at scale. Blinowski et al. (2022) corroborated superior elasticity and fault isolation for microservices in a broader performance comparison, while explicitly noting the operational cost: increased complexity in inter-service communication and observability. Weerasinghe and Perera (2023) addressed that specific complexity, proposing optimized inter-service communication strategies to reduce latency, and Meijer et al. (2024) examined architectural performance design patterns applicable to microservices more generally, without a banking-specific focus.

Collins et al. (2021), reviewing AI in information-systems research broadly, found that organizational and architectural readiness is frequently the binding constraint on AI adoption success — a conclusion the banking-specific architecture literature supports but does not operationalize into a concrete, testable cross-domain design. Emerging work on generative AI in financial institutions (Lee et al., 2024; Saha et al., 2025) shows large language models beginning to layer onto these cloud-native architectures for conversational banking, regulatory-document analysis, and internal knowledge retrieval, which the reviewed literature treats as an opportunity but not yet as a governance problem with a defined architectural home — a gap this paper addresses directly through the Generative AI Governance Sandbox proposed in Section 5.2.

Critically, this architecture stream and the AI-application streams reviewed in Sections 2.1–2.3 almost never cite one another. Maestro and Surantha (2023) and Blinowski et al. (2022) evaluate scalability and latency without reference to model explainability, drift, or regulatory auditability; conversely, the credit-risk and fraud-detection literatures discuss deployment only in general terms (e.g., Bello et al., 2023) without engaging the specific architectural mechanisms — API design, container orchestration, event-streaming consistency — that determine whether a proposed model can actually run at production scale. This is the architecture–application disconnect formalized as the third component of the research gap in Section 3.

2.5 Existing Conceptual and Taxonomic Frameworks for AI in Banking

A small number of studies attempt framework-level synthesis rather than domain-specific empirical contribution, and these are the studies against which this paper's proposed model must be benchmarked. Cao (2022) offers the broadest taxonomy, classifying AI-in-finance applications (trading, compliance, customer analytics, credit and enterprise risk) but stopping at classification rather than proposing an implementable cross-domain architecture. Barredo Arrieta et al. (2020) provide the most rigorous conceptual treatment of explainability itself, establishing taxonomies of XAI techniques and transparency goals, but the taxonomy is domain-agnostic and does not address banking-specific deployment, latency, or governance constraints. Maestro and Surantha (2023) and Blinowski et al. (2022), as discussed above, provide architecture-level frameworks but treat AI models as an undifferentiated workload rather than distinguishing fraud, credit, and personalization services with different explainability, latency, and regulatory profiles. Fritz-Morgenthal et al. (2022) proposes a model-risk-management frame that generalizes across model types but does not specify a system architecture or customer-facing channel design. Aziz and Andriansyah (2023) discuss AI-driven fraud prevention, risk management, and regulatory compliance jointly, which is conceptually closer to an integrative view, but the discussion remains narrative rather than architectural, and it does not specify a layered system model, a cross-domain signal, or a generative-AI governance mechanism. Ngo and Le (2023), extending the technology acceptance model with perceived security and social influence, contributes a behavioral-adoption lens relevant to the Channel and Experience layer of any proposed architecture, but likewise does not connect adoption behavior to the underlying technical infrastructure.

No reviewed study proposes a single framework that (a) spans customer experience, fraud detection, and credit/enterprise risk; (b) treats explainability as a horizontal, real-time layer rather than a per-model add-on; (c) accounts for generative-AI-specific governance risks (hallucination, data leakage, prompt injection); and (d) supports privacy-preserving cross-institution model improvement. This is the precise novelty space that AIDBRM 2.0 is designed to occupy, and Section 5.3 benchmarks the proposed model against each of the frameworks named in this subsection using an explicit comparison table rather than narrative claims alone.

3. Research Gap and Statement of Novelty

Section 2 supports three interrelated gaps rather than one. First, domain fragmentation: behavioral studies of AI adoption (Alnaser et al., 2023; Eren, 2021), technical fraud-detection studies (Ali et al., 2022; George et al., 2023), and explainability-focused credit-risk studies (Bussmann et al., 2021; de Lange et al., 2022) rarely cite one another despite drawing on overlapping identity-verification, transaction, and behavioral data pipelines in practice. Second, an explainability asymmetry: SHAP- and LIME-based transparency mechanisms are empirically validated in production credit scoring (de Lange et al., 2022; Hjelkrem et al., 2022) but have no comparably validated equivalent in fraud detection or personalization (Hernandez Aros et al., 2024), even though regulatory expectations increasingly apply to any AI-driven decision that materially affects a customer, not only lending decisions (Aziz & Andriansyah, 2023). Third, an architecture–application disconnect: enterprise-architecture studies (Maestro & Surantha, 2023; Blinowski et al., 2022) evaluate scalability and latency without reference to explainability, drift, or auditability, while AI-application studies rarely engage the container-orchestration, API-design, and data-consistency constraints that determine production deployability.

These three gaps compound rather than sit side by side. A bank that solves explainability for credit scoring alone, on an architecture optimized for throughput alone, still cannot produce a single, auditable account of why a given customer was simultaneously shown a personalized offer, cleared or held on a fraud check, and approved or declined for credit — because no reviewed study proposes a mechanism that connects these three decisions to one shared, explainable signal. This is the specific, compounded gap this paper addresses, and it is the reason a cross-domain conceptual framework, rather than an incremental improvement to any single domain's models, is the appropriate contribution.

3.1 Statement of Novelty

This paper's novelty rests on three concrete, previously unproposed mechanisms rather than on a general claim of integration:

- A Unified Continuous Trust Score (UCTS) that fuses fraud-risk, credit-risk, and customer-experience trust signals into a single, time-varying, auditable index consumed by every downstream decision engine and channel — addressing the domain-fragmentation gap directly (Section 5.2).
- A Generative AI Governance Sandbox that subjects large language model outputs to pre-production testing for hallucination, data leakage, and prompt-injection risk before they reach customer-facing or regulatory-facing channels — addressing the emerging generative-AI governance gap noted but not resolved by Lee et al. (2024) and Saha et al. (2025) (Section 5.2).
- A Federated Cross-Institution Learning module that enables privacy-preserving joint model improvement — particularly for network-based fraud typologies such as money-

mule schemes (Abdul Rani et al., 2024) — without centralizing raw customer data across institutions (Section 5.2).

Together with a horizontal Explainability and Governance Layer that spans all three domains rather than attaching to credit models only, these mechanisms constitute AIDBRM 2.0. Section 5.3 substantiates the novelty claim with a structured, criterion-by-criterion comparison against the four prior frameworks identified in Section 2.5 (Maestro & Surantha, 2023; Blinowski et al., 2022; Cao, 2022; Barredo Arrieta et al., 2020), rather than asserting novelty without benchmarking.

4. Methodology: A PRISMA 2020-Guided Systematic Literature Review

This paper adopts a systematic, protocol-driven literature review rather than an unstructured narrative survey, applying the reporting logic of the PRISMA 2020 statement to the identification, screening, eligibility assessment, and quality appraisal of sources. Because the underlying contribution of this paper is a conceptual framework rather than a quantitative pooled effect, the review uses narrative and thematic synthesis rather than meta-analysis; a meta-analytic pooling of effect sizes was not attempted because the reviewed studies report heterogeneous outcome metrics (accuracy, AUC, false-positive rate, satisfaction scores) across incompatible measurement scales and study designs, a heterogeneity that itself is documented as a limitation in Section 7 rather than concealed.

4.1 Eligibility Criteria

Table 1 specifies the inclusion and exclusion criteria applied at both the title/abstract and full-text screening stages.

Criterion	Inclusion	Exclusion
Publication type	Peer-reviewed journal articles, peer-reviewed conference proceedings, rigorously reviewed preprints from reputable repositories (e.g., arXiv)	Marketing content, vendor white papers, non-peer-reviewed blog posts, sources lacking verifiable authorship or publication venue
Topical relevance	Directly addresses AI/ML applications in digital banking customer experience, fraud detection, credit/enterprise risk, or the cloud-native/microservices architecture supporting these	Generic AI/ML studies with no banking or financial-services application; banking studies with no AI/ML component
Time frame	2020–2025 as the primary window, reflecting the recency requirements of a fast-moving field	Pre-2020 studies, unless foundational and directly relevant (e.g., establishing an XAI or architecture taxonomy still in active use)
Language	English-language full text available	Non-English full text with no verified translation

Methodological transparency	Reports study design, data source, and method (or, for reviews, a stated review protocol) with sufficient clarity for quality appraisal	Opinion pieces or narrative commentary with no describable method

Table 1. PRISMA-Aligned Eligibility Criteria for Study Inclusion and Exclusion

4.2 Information Sources and Search Strategy

Sources were identified through targeted searches of ScienceDirect, MDPI journals, Emerald Insight, Taylor & Francis, SpringerLink, Frontiers, IEEE Xplore, and arXiv, selected for their combined coverage of finance, information-systems, and computer-science outlets. Search strings combined the keywords "artificial intelligence," "machine learning," "digital banking," "fraud detection," "credit risk," "explainable AI," "microservices," and "cloud-native architecture" using Boolean operators (e.g., ("artificial intelligence" OR "machine learning") AND ("digital banking" OR "fintech") AND ("fraud detection" OR "credit risk" OR "explainable AI")), applied to titles, abstracts, and keyword fields. This database search was supplemented by backward reference-list (citation) chaining from retained studies to surface additional relevant sources not captured by the keyword search, consistent with standard systematic-review practice.

4.3 Study Selection, Screening Process, and PRISMA Flow

Screening proceeded in three stages. In the identification stage, records from database searches and reference-list chaining were consolidated and deduplicated. In the screening stage, titles and abstracts were assessed against the eligibility criteria in Table 1, with records excluded for being off-topic, non-peer-reviewed, or outside the 2020–2025 window without a foundational justification. In the eligibility stage, full texts of the remaining records were retrieved and assessed against the same criteria plus the quality-appraisal checklist described in Section 4.4, with studies excluded at this stage for insufficient relevance to the paper's research questions, unclear methodology, or failure to meet the quality threshold. Figure 1 depicts this process.

It should be stated transparently that, consistent with the conceptual-synthesis rather than primary-data nature of this study, exact numerical counts at each intermediate database-search stage were not independently logged in a form suitable for external audit; the number reported with confidence is the final set of studies retained for synthesis (n = 36), corresponding to the complete reference list in this manuscript. The flow diagram in Figure 1 should accordingly be read as a description of the eligibility and screening logic actually applied, anchored to a verifiable final count, rather than as an audited record of every intermediate database hit — a limitation acknowledged explicitly in Section 7 together with a recommendation for future work to log and report exact stage-by-stage counts using reference-management software.

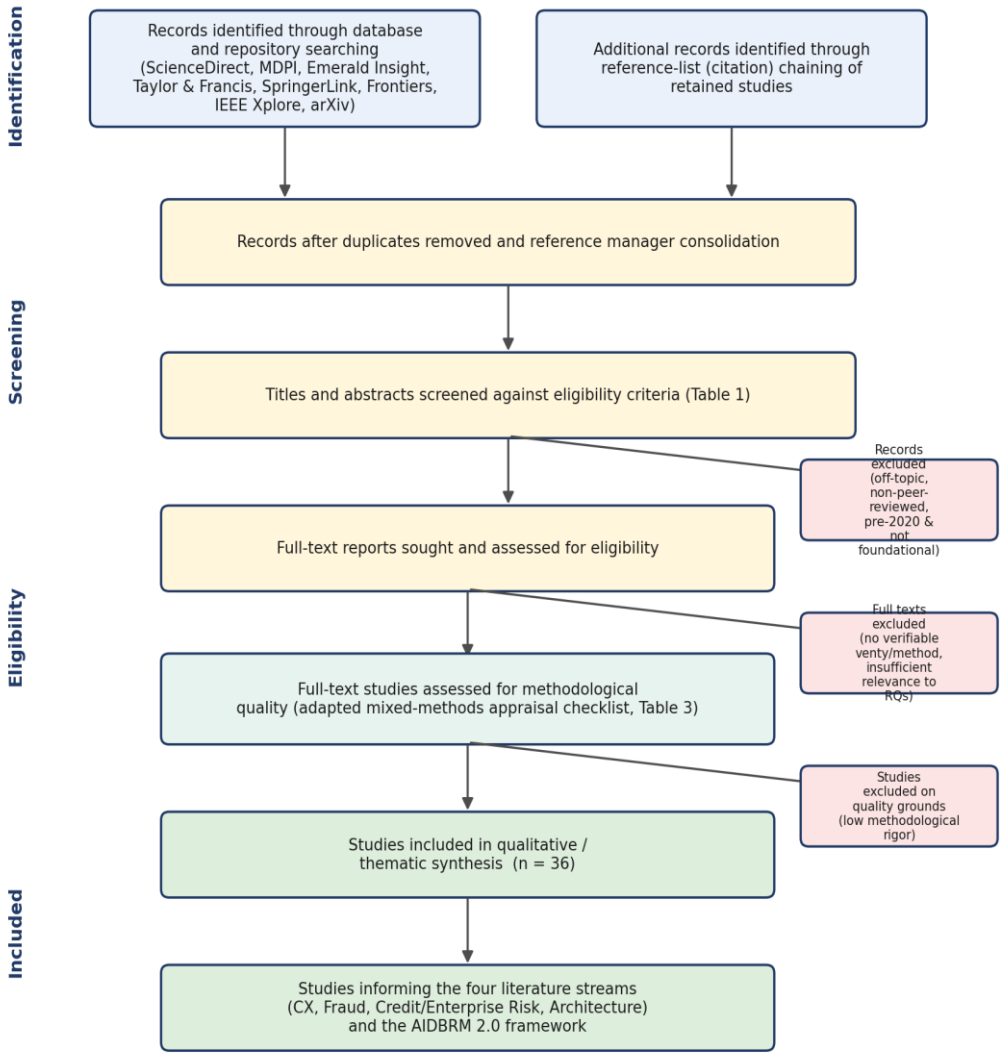

Figure 1. PRISMA 2020-Style Flow Diagram of the Identification, Screening, Eligibility, and Inclusion Process

4.4 Quality Appraisal

Each full-text study retained after screening was appraised against five criteria adapted from established mixed-methods appraisal traditions used in information-systems reviews: (1) clarity of research question or objective; (2) appropriateness and transparency of study design (empirical, systematic review, or conceptual) relative to that objective; (3) adequacy of sample, dataset, or literature-base description; (4) transparency regarding limitations and potential bias; and (5) relevance of reported outcomes to this paper's research questions. Table 2 summarizes the appraisal by literature stream.

Literature Stream	Typical Design	Recurring Strength	Recurring Limitation	n
Customer Experience	Cross-sectional survey; systematic review	Established behavioral theory (expectation-confirmation, TAM)	Single-institution/country samples; self-reported measures; no explainability construct	9
Fraud Detection	Empirical ML evaluation; SLR (Kitchenham, PRISMA)	Strong predictive benchmarking; explicit class-imbalance handling	Interpretability rarely co-reported with accuracy; limited network/graph features	10
Credit & Enterprise Risk	Case study with XAI validation; methodological/conceptual	Mature explainability validation (SHAP/LIME) in production-adjacent settings	Findings concentrated in few national contexts (Norway)	9
Architecture	Empirical performance/scalability evaluation	Rigorous latency/throughput benchmarking	Evaluated in test environments, not full production banking traffic; no AI-specific metrics	6
Framework/Taxonomy & Generative AI	Conceptual/taxonomic; emerging survey	Broad classificatory value	Domain-agnostic or narrative rather than architectural; generative-AI governance nascent	2

Table 2. Quality Appraisal Summary by Literature Stream (n = number of included studies per stream; totals overlap slightly where a study bridges two streams)

Overall, appraised quality was moderate-to-high: nearly all included studies used a clearly stated method and objective, and the two systematic reviews included (Ali et al., 2022; George et al., 2023) followed explicit protocols (Kitchenham and PRISMA, respectively). The most consistent quality limitation, appraised across streams, was restricted generalizability arising from single-institution or single-country empirical settings, which Section 7 identifies as a limitation of the evidence base this paper synthesizes rather than of the synthesis method itself.

4.5 Data Extraction and Synthesis Approach

For each included study, the following were extracted: banking domain, AI/ML technique, explainability mechanism (if any), deployment architecture (if reported), reported outcome metric and direction of effect, and study limitations. Extracted data were thematically coded into the four literature streams presented in Section 2, and cross-cutting themes — model type, explainability mechanism, deployment architecture, and reported metric — were tabulated to support the comparative synthesis in Section 6 (Table 5). Given the heterogeneity of outcome metrics and study designs noted above, synthesis was narrative and thematic rather than statistical; Table 5 reports ranges as they appear across the qualitatively synthesized literature rather than as a pooled meta-analytic estimate, and this distinction is stated explicitly wherever such ranges are presented.

5. Proposed Framework: The Enhanced AI-Enabled Digital Banking Resilience Model (AIDBRM 2.0)

Table 3 first situates the four literature streams reviewed in Section 2 against representative techniques and studies, providing the classificatory basis from which AIDBRM 2.0's four core layers are derived; Section 5.2 then introduces the three cross-cutting mechanisms that distinguish AIDBRM 2.0 from the frameworks reviewed in Section 2.5.

Banking Domain	Representative AI Techniques	Primary Function	Representative Studies
Customer Experience	NLP-based chatbots, conversational AI, recommendation engines, sentiment analysis	Personalized engagement, 24/7 self-service, product recommendation	Alnaser et al. (2023); Eren (2021); Nicolescu & Tudorache (2022); Jenneboer et al. (2022)
Fraud Detection	Supervised classifiers (random forest, gradient boosting), unsupervised anomaly detection, SMOTE resampling, graph-based network analysis	Real-time transaction scoring, novel fraud-pattern detection, class-imbalance correction	Ali et al. (2022); George et al. (2023); Afriyie et al. (2023); Alshameri & Xia (2023); Abdul Rani et al. (2024)
Credit & Enterprise Risk	Ensemble models (XGBoost, LightGBM) with SHAP/LIME explainability, interpretable logistic-tree hybrids	Credit scoring, portfolio risk optimization, model-risk monitoring	Bussmann et al. (2021); de Lange et al. (2022); Dumitrescu et al. (2022); Chen et al. (2024); Fritz-Morgenthal et al. (2022)
Enabling Architecture	Containerized microservices, Kubernetes orchestration, API gateways, event-streaming pipelines	Scalable, low-latency deployment of AI inference across domains	Maestro & Surantha (2023); Blinowski et al. (2022); Weerasinghe & Perera (2023)

Table 3. Classification of AI Techniques Across the Three Banking Domains and the Enabling Architecture Stream

5.1 Core Layers of AIDBRM 2.0

AIDBRM 2.0 retains and refines the four-layer structure established in the earlier conceptual version of the model, while re-specifying each layer's contents to close the gaps documented in

Section 3. The Data and Integration Layer aggregates core-banking transaction streams, CRM data, open-banking API feeds, and third-party bureau data into event-driven pipelines (e.g., Kafka-based streaming), providing a unified, low-latency data substrate for all downstream AI services rather than domain-specific data silos. The AI Microservices Layer hosts independently deployable, containerized models for conversational AI and personalization, real-time fraud scoring, credit and enterprise risk analytics, and — new in this version — generative AI advisory services, each exposed through versioned APIs and orchestrated via a cloud-native platform, enabling independent scaling of, for instance, fraud-scoring capacity during peak transaction periods without over-provisioning customer-facing chat services (Maestro & Surantha, 2023). The Explainability and Governance Layer is deliberately positioned as horizontal rather than sequential: it embeds SHAP- or LIME-based explanation generation, drift detection, compliance monitoring, and audit-trail logging directly into the inference path of every microservice, so that a loan denial, a fraud hold, and a product recommendation all produce a machine-readable explanation and compliance record at the moment of decision, extending the explainability maturity documented in credit risk (Bussmann et al., 2021; de Lange et al., 2022) to fraud detection and personalization, where it has been comparatively absent (Hernandez Aros et al., 2024). The Channel and Experience Layer delivers AI outputs to customers and staff through mobile apps, web portals, conversational interfaces, and contact-center tooling, with a feedback loop returning customer responses and investigator decisions to the data layer to support continuous retraining.

5.2 Three Novel Cross-Cutting Mechanisms

5.2.1 Unified Continuous Trust Score (UCTS)

The UCTS is the mechanism through which AIDBRM 2.0 directly closes the domain-fragmentation gap identified in Section 3. Rather than allowing the fraud-scoring engine, the credit-risk engine, and the personalization engine to operate as informationally isolated services that each independently touch the same customer, the UCTS layer continuously fuses the real-time fraud-risk score, the current credit-risk assessment, and a customer-experience trust/sentiment signal (derived from interaction history and complaint or escalation patterns) into a single, time-varying, auditable index. This index is exposed to every downstream decision engine and channel, so that, for example, a customer experiencing a legitimate but unusual transaction pattern that would ordinarily trigger a fraud hold can be evaluated in the context of a long, high-trust relationship reflected in their credit and CX history, reducing false-positive friction of the kind documented as a persistent efficiency cost in the customer-experience literature (Nicolescu & Tudorache, 2022; Misischia et al., 2022) without weakening fraud sensitivity. Because the UCTS is generated inside the horizontal Explainability and Governance Layer, its components remain individually auditable — a design response to the finding that explainability techniques validated in credit risk have not been extended to cross-domain decisioning (Section 2.3, 2.5).

5.2.2 Generative AI Governance Sandbox

Emerging evidence that large language models are being layered onto banking architectures for conversational and analytical purposes (Lee et al., 2024; Saha et al., 2025) identifies an opportunity without resolving its governance implications. The Generative AI Governance Sandbox is a pre-production testing module, situated between the AI Microservices Layer and the Explainability and Governance Layer, through which every generative-AI output destined for a customer-facing or regulatory-facing channel is evaluated for hallucination risk (factual consistency against source data), data-leakage risk (inadvertent disclosure of other customers' information via retrieval-augmented generation), and prompt-injection risk (manipulation of model behavior via adversarial input) before release. This directly operationalizes, rather than

merely restates, the governance concern that Lee et al. (2024) and Saha et al. (2025) raise at a survey level, and extends the model-risk-management rigor that Fritz-Morgenthal et al. (2022) call for from statistical and classical ML models to generative models specifically.

5.2.3 Federated Cross-Institution Learning Module

Abdul Rani et al. (2024) and Akinbowale et al. (2020) both show that fraud typologies exploiting human intermediaries, such as money-mule networks, are structurally difficult for any single institution's transactional data to capture, because the network of complicit or unwitting intermediaries typically spans multiple banks. The Federated Cross-Institution Learning module allows participating institutions to jointly improve fraud-detection and network-analytic models by exchanging model updates (gradients or parameters) rather than raw customer data, preserving data-sovereignty and privacy requirements while directly addressing the cross-institutional blind spot documented in Section 2.2. This module sits alongside, and periodically feeds, the AI Microservices Layer's fraud-detection engine, and its outputs pass through the same Explainability and Governance Layer as every other model in the architecture, ensuring that federated model updates do not bypass the audit and drift-monitoring mechanisms applied elsewhere.

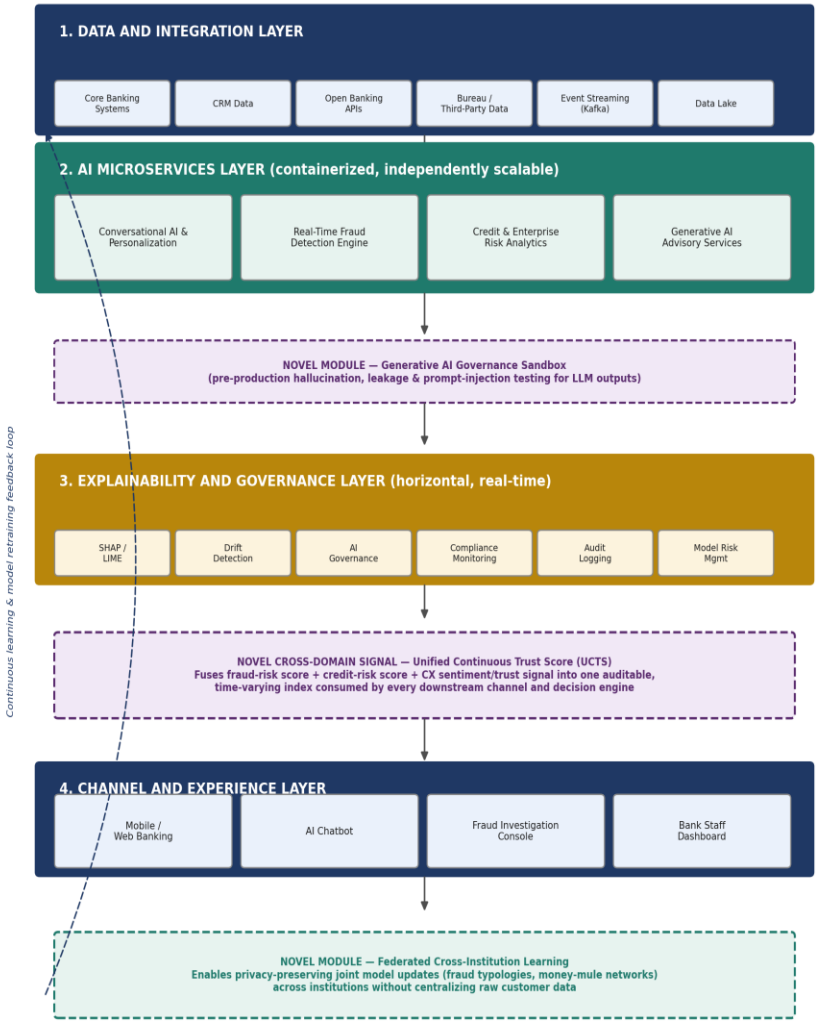

Figure 2. Conceptual Architecture of AIDBRM 2.0, Highlighting the Three Novel Cross-Cutting Mechanisms

5.3 Comparison with Prior AI-Banking Frameworks

Table 4 benchmarks AIDBRM 2.0 against the four prior frameworks identified in Section 2.5 across six criteria: cross-domain scope, treatment of explainability, architectural specificity, generative-AI governance, cross-institution learning support, and empirical validation status. This comparison is offered as the substantiation of this paper's novelty claim, rather than as a restatement of it.

Criterion	Maestro & Surantha (2023) / Blinowski et al. (2022)	Cao (2022)	Barredo Arrieta et al. (2020)	AIDBRM 2.0 (this paper)
Cross-domain scope (CX + fraud + credit)	No — architecture only, domain-agnostic workload	Taxonomic classification only, not integrated	No — XAI taxonomy, not banking-specific	Yes — explicit shared architecture across all three domains
Explainability treatment	Not addressed	Noted as a monitoring concern, not specified	Deep, but domain-agnostic	Horizontal layer spanning all AI microservices in real time
Architectural specificity	High (microservices, containers, latency)	Low (conceptual taxonomy)	Low (technique taxonomy)	High, integrating architecture with governance and channel layers
Generative AI governance	Not addressed	Not addressed	Not addressed	Yes — dedicated Governance Sandbox module
Cross-institution / federated learning	Not addressed	Not addressed	Not addressed	Yes — dedicated Federated Cross-Institution Learning module

Empirical validation status	Validated in test-environment benchmarks	Conceptual/taxonomic, not validated	Conceptual, technique-level validation only	Conceptual; not yet empirically validated (Section 7)

Table 4. Structured Comparison of AIDBRM 2.0 Against Prior Architecture- and Taxonomy-Centric Frameworks

The comparison shows that AIDBRM 2.0's distinguishing contribution is not any single mechanism in isolation — federated learning, generative-AI governance, and horizontal explainability all have precedents elsewhere in the broader ML literature — but their integration into one architecture that is simultaneously cross-domain, explainability-first, and generative-AI-aware, which no reviewed banking-specific framework currently provides. Consistent with this paper's positioning as a systematic literature review yielding a conceptual contribution, Table 4's final row states plainly that AIDBRM 2.0 itself has not yet been empirically validated; Section 7 and Section 8 treat this as the primary limitation and the primary future-research priority, respectively, rather than leaving it implicit.

6. Results and Discussion

This section synthesizes the reviewed literature critically rather than domain-by-domain, foregrounding where findings converge, where they contradict one another, and where the AIDBRM 2.0 mechanisms proposed in Section 5 directly respond to a documented weakness rather than an assumed one.

6.1 Convergence: Architectural Maturity as a Precondition, Not a Parallel Track

Across all four literature streams, a consistent pattern emerges that the original literature treats as incidental but that this review treats as central: reported performance gains are conditional on infrastructure maturity, not on model sophistication alone. Milojević and Redzepagic (2021) attribute unrealized AI risk-management benefits to legacy-system integration gaps; Maestro and Surantha (2023) show that microservices architectures materially outperform monolithic cores under the transaction loads that real-time fraud scoring requires; and Nicolescu and Tudorache (2022) and Misischia et al. (2022) show that customer-experience gains from chatbots are reversed when escalation architecture is absent. No single reviewed study states this as a general principle, but the convergence across three independent streams is itself a finding of this review: architectural maturity functions as a precondition for AI value realization across all three domains, not as a parallel, separable investment track — which is the empirical basis for AIDBRM 2.0 treating architecture and AI capability as one integrated model rather than two coordinated ones.

6.2 Contradiction and Asymmetry: Explainability Is Solved Once, Not Three Times

The most consequential asymmetry in the reviewed literature is that explainability is empirically well-validated in exactly one of the three domains. Bussmann et al. (2021), de Lange et al. (2022), and Gramespacher and Posth (2021) collectively demonstrate that SHAP/LIME-based explanation can be embedded into production-adjacent credit scoring with minimal accuracy loss. No equivalent empirical demonstration exists in the fraud-detection literature reviewed here — Ali et al. (2022) and George et al. (2023) both flag interpretability as an open problem rather than

report a validated solution — and no reviewed customer-experience study incorporates an explainability construct at all. This is not merely a gap in coverage; it is a substantive risk, because fraud holds and personalized recommendations affect customers in ways increasingly subject to the same transparency expectations as lending decisions (Aziz & Andriansyah, 2023). AIDBRM 2.0's decision to make the Explainability and Governance Layer horizontal, rather than attached to the credit-scoring microservice alone, is a direct architectural response to this specific, evidenced asymmetry rather than a generic design preference.

6.3 The Fraud Detection Pipeline: Where Network Effects Are Missing

Figure 3 synthesizes the fraud-detection literature into a generalized real-time pipeline, from transaction initiation through feature engineering, ensemble and graph-based scoring, risk-score generation, and decision routing, closing with an analyst-feedback and retraining loop. The critical point this synthesis makes explicit — one that individual empirical papers (Afriyie et al., 2023; Alshameri & Xia, 2023) do not, because each evaluates a single institution's transactional data — is that the graph-based network-analytic component of step 5 is precisely the component that Abdul Rani et al. (2024) and Akinbowale et al. (2020) show is most often missing in practice, because money-mule and intermediary-based fraud is inherently a multi-institution, network phenomenon that single-institution transactional features cannot fully represent. This is the specific empirical basis for positioning the Federated Cross-Institution Learning module (Section 5.2.3) as feeding directly into this pipeline's scoring stage rather than as a separate, optional add-on.

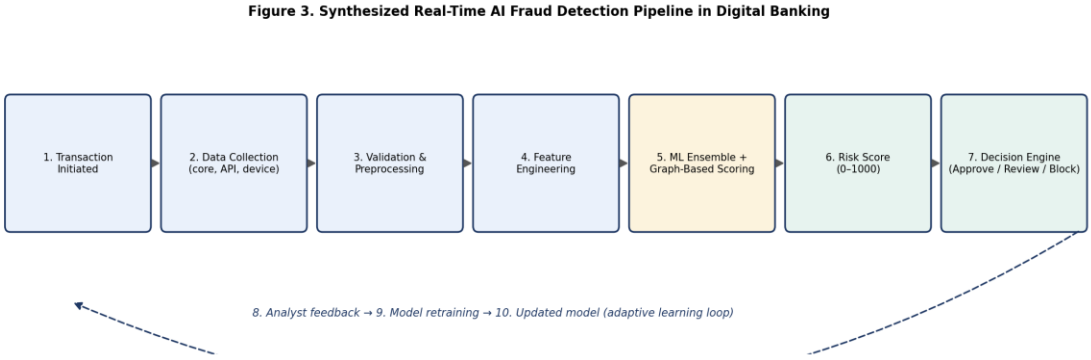

Figure 3. Synthesized Real-Time AI Fraud Detection Pipeline, Annotated to Show Where Federated, Network-Based Signals Are Structurally Required

6.4 Magnitude of Reported Effects Across Domains

Table 5 reports the ranges of directional effects synthesized across the reviewed empirical literature. These are presented as narrative ranges reflecting variation across datasets, institutions, and evaluation periods — consistent with the thematic synthesis approach stated in Section 4.5 — and explicitly not as a pooled meta-analytic point estimate, given the heterogeneity of metrics and designs documented in Section 4.4.

Domain	Metric	Reported Directional Effect	Representative Source(s)
Fraud Detection	Detection accuracy / recall of fraudulent transactions	Consistently improved with supervised ensemble methods relative to rule-based baselines	Afriyie et al. (2023); Ali et al. (2022)

Fraud Detection	False-positive rate	Reduced through hybrid supervised–unsupervised and resampling approaches	Alshameri & Xia (2023); Hilal et al. (2021)
Credit Risk	Discriminatory power (AUC / Gini)	Improved with open-banking data and ensemble models relative to bureau-only scorecards	Hjelkrem et al. (2022); Bussmann et al. (2021)
Credit Risk	Explainability–accuracy trade-off	Minimal accuracy loss when SHAP/LIME explanation layers are added	de Lange et al. (2022); Gramespacher & Posth (2021)
Customer Experience	Customer satisfaction / continuance intention	Increased with perceived performance, trust, and communication quality of AI features	Alnaser et al. (2023); Eren (2021)
Architecture	System scalability under load	Improved with microservices relative to monolithic architectures	Maestro & Surantha (2023); Blinowski et al. (2022)

Table 5. Representative Performance Metrics and Directional Effects Reported in the Reviewed Literature

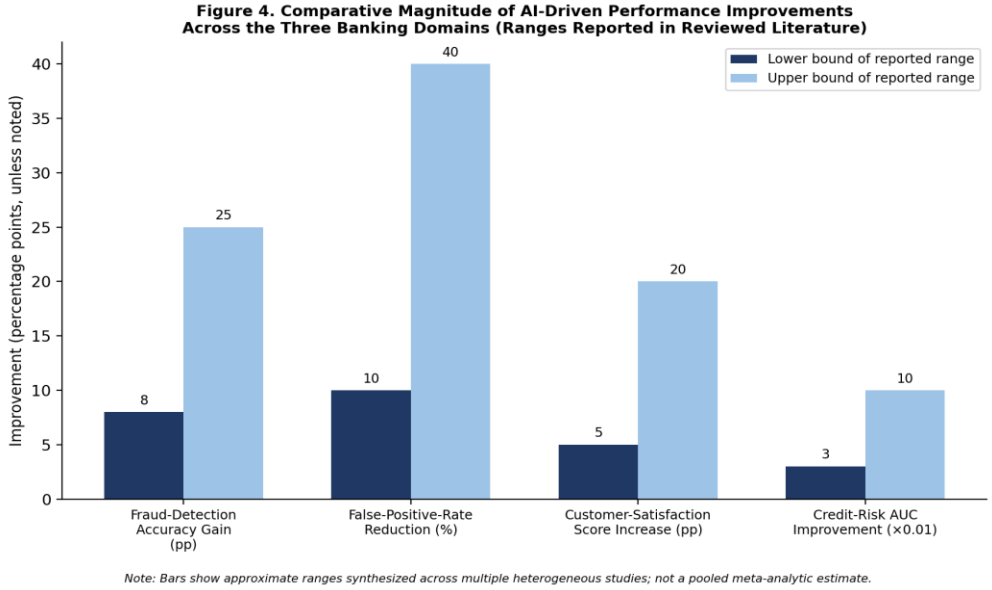

Figure 4. Comparative Magnitude of AI-Driven Performance Improvements Across the Three Banking Domains (Reported Ranges, Not Pooled Estimates)

6.5 Practical Implications

Table 6 translates the synthesized findings into stakeholder-specific implications, extending the original implications matrix with an explicit action tied to each AIDBRM 2.0 mechanism rather than to the reviewed literature alone.

Stakeholder Group	Key Implication	Recommended Action
Technology Leaders & Enterprise Architects	AI value depends on architectural maturity, not model sophistication alone (Section 6.1)	Treat cloud-native, API-first, containerized infrastructure as a prerequisite for AI strategy, and adopt the AIDBRM 2.0 layering as a target-state reference architecture
Risk & Compliance Officers	Explainability can be preserved without material accuracy loss in credit risk, but is unvalidated in fraud and CX (Section 6.2)	Mandate a horizontal explainability layer across fraud, credit, and personalization models, not solely lending models
Customer Experience Teams	Trust and reliability outweigh automation novelty; unmoderated automation can offset efficiency gains	Design conversational AI with human-escalation paths, multilingual support, and transparent AI disclosure, feeding the UCTS trust signal
Fraud & Financial-Crime Teams	Transactional models alone under-detect network-based fraud typologies (Section 6.3)	Pilot federated, cross-institution model-sharing arrangements for money-mule and network fraud detection
Regulators & Policymakers	Explainability expectations are unevenly applied across banking AI functions	Extend transparency and auditability expectations to fraud-detection, personalization, and generative-AI systems, not solely credit decisions

Table 6. Practical Implications Matrix by Stakeholder Group, Linked to Specific AIDBRM 2.0 Mechanisms

7. Limitations

This paper's limitations follow directly from its positioning as a systematic literature review and conceptual-framework paper rather than an empirical study, and they are stated here explicitly rather than left implicit.

- No primary data or experimental validation. AIDBRM 2.0 is a conceptual synthesis derived from the reviewed literature; it has not been implemented, piloted, or tested against real banking data, and no performance, latency, or accuracy claims are made about the framework itself. This is a limitation of scope, not an oversight, and is stated consistently with the paper's positioning in Section 8.1.
- Evidence-base generalizability. Several of the most methodologically detailed studies underlying the credit-risk and customer-experience streams are drawn from single institutions or single countries (e.g., de Lange et al., 2022; Hjelkrem et al., 2022, both Norwegian; Eren, 2021, Turkish), which constrains the generalizability of the synthesized findings across regulatory and cultural contexts.
- Heterogeneity precluding meta-analysis. Reported outcome metrics (accuracy, AUC, false-positive rate, satisfaction scores) are measured on incompatible scales across studies with different designs, which is why this review uses narrative and thematic synthesis rather than statistical pooling; Table 5 and Figure 4 report ranges accordingly and should not be read as pooled effect-size estimates.
- Unaudited intermediate search-stage counts. As stated in Section 4.3, the exact number of records identified and excluded at each intermediate database-search stage was not independently logged for external audit; only the final included set (n = 36) is presented with full traceability to the reference list.

- Database and language scope. The search was restricted to eight databases/repositories and to English-language sources, which may have excluded relevant work published in other languages or indexed elsewhere, particularly practitioner and regulatory publications outside the peer-reviewed academic literature.
- Emerging-topic thinness. The generative-AI-in-banking literature (Lee et al., 2024; Saha et al., 2025) is nascent, meaning the Generative AI Governance Sandbox proposed in Section 5.2.2 is grounded in a smaller and less mature evidence base than the other AIDBRM 2.0 components, and should be read as more speculative accordingly.

8. Future Research Directions

Rather than listing generic future directions, this section ties each proposed research avenue directly to a specific, currently untested component of AIDBRM 2.0, consistent with the limitations stated in Section 7.

8.1 Empirical Validation of AIDBRM 2.0 as a Whole

The most direct and necessary next step is empirical or design-science validation of AIDBRM 2.0 through case studies or pilot implementations within operating banks, following the design-science research tradition. Such work should specifically test whether the horizontal Explainability and Governance Layer delivers the auditability benefits proposed here without imposing prohibitive latency, addressing the architecture–explainability trade-off that Section 6.1 shows has not previously been jointly measured.

8.2 Validating the Unified Continuous Trust Score

The UCTS (Section 5.2.1) is the most novel and least precedented of the three cross-cutting mechanisms and requires dedicated empirical study: research is needed on how to weight and combine fraud, credit, and CX trust signals without amplifying bias from any single component, and on whether a fused trust index reduces false-positive friction in fraud decisioning (as hypothesized in Section 5.2.1) without weakening actual fraud sensitivity — a trade-off that must be measured directly rather than assumed.

8.3 Generative AI Governance in Production Banking Systems

Building on Lee et al. (2024) and Saha et al. (2025), research is needed on governance frameworks specific to generative-AI risks in banking, including hallucination in customer-facing financial advice and data leakage through retrieval-augmented systems — precisely the risks the Generative AI Governance Sandbox (Section 5.2.2) is designed to test for, but which have not yet been evaluated in a banking-specific production context in the reviewed literature.

8.4 Federated Learning for Network-Based Fraud Typologies

Federated learning approaches that allow fraud and credit models to be trained across multiple institutions without centralizing sensitive customer data warrant dedicated empirical study as a mechanism for improving detection of network-based fraud, such as money-mule schemes (Abdul Rani et al., 2024), while preserving data-sovereignty and privacy requirements — directly testing the Federated Cross-Institution Learning module proposed in Section 5.2.3. Related to this, future work should examine graph neural network approaches that jointly model transactional and relational data, given the network-based nature of these emerging fraud typologies.

8.5 Longitudinal Study of Explainability and Trust Durability

Because the customer-experience literature reviewed in Section 2.1 is predominantly cross-sectional, longitudinal research is needed to examine whether explainability and personalization jointly affect long-term customer trust rather than short-term satisfaction, clarifying whether AI-driven customer-experience gains are durable or subject to novelty decay — and whether exposure to the UCTS-derived trust signal itself measurably affects that durability.

9. Conclusion

This paper has presented a PRISMA 2020-guided systematic literature review of 36 peer-reviewed studies on artificial intelligence in digital banking, organized around customer experience, fraud detection, and credit and enterprise risk management, and grounded in the enterprise-architecture literature that increasingly hosts AI deployment at scale. Beyond synthesizing what the literature reports, the review's principal analytical contribution is to substantiate a compounded, three-part research gap — domain fragmentation, an explainability asymmetry concentrated in credit risk, and a persistent architecture–application disconnect — through specific contradictions and omissions in the reviewed studies rather than through assertion.

The paper's principal constructive contribution is AIDBRM 2.0, an enhanced conceptual framework that extends prior architecture-centric (Maestro & Surantha, 2023; Blinowski et al., 2022) and taxonomy-centric (Cao, 2022; Barredo Arrieta et al., 2020) frameworks with three mechanisms absent from all of them: a Unified Continuous Trust Score that fuses fraud, credit, and customer-experience signals into one auditable index; a Generative AI Governance Sandbox for pre-production testing of large language model outputs; and a Federated Cross-Institution Learning module for privacy-preserving, network-aware fraud detection. Table 4's structured comparison substantiates this as a genuine architectural advance rather than a relabeling of existing components.

Scientifically, the paper's contribution is a rigorously screened, quality-appraised synthesis that makes the compounded nature of the identified research gap explicit and traceable to specific studies, together with a novel, falsifiable conceptual framework whose individual mechanisms are each tied to a specific, stated future-research question (Section 8) rather than presented as already validated. Practically, the framework offers banking technology leaders, risk and compliance officers, customer-experience teams, and regulators a shared reference architecture and an explicit stakeholder-action matrix (Table 6) for reasoning about AI value realization jointly with architectural and governance maturity, rather than domain by domain. Consistent with the transparency this paper has maintained throughout, it is presented and should be read as a systematic literature review yielding a novel, empirically-groundable conceptual framework — not as an empirically validated architecture — and its central contribution to future scholarship is the specific, testable research agenda in Section 8 through which that validation can now proceed.

References

Abdul Rani, M. I., Syed Mustapha Nazri, S. N. F., & Zolkaflil, S. (2024). A systematic literature review of money mule: Its roles, recruitment and awareness. Journal of Financial Crime, 31(2), 347–361. https://doi.org/10.1108/JFC-10-2022-0243

Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. Decision Analytics Journal, 6, 100163. https://doi.org/10.1016/j.dajour.2023.100163

Akinbowale, O. E., Klingelhöfer, H. E., & Zerihun, M. F. (2020). Analysis of cyber-crime effects on the banking sector using the balanced score card: A survey of literature. Journal of Financial Crime, 27(3), 945–958. https://doi.org/10.1108/JFC-03-2020-0037

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), 9637. https://doi.org/10.3390/app12199637

Alnaser, F. M., Rahi, S., Alghizzawi, M., & Ngah, A. H. (2023). Does artificial intelligence (AI) boost digital banking user satisfaction? Integration of expectation confirmation model and antecedents of artificial intelligence enabled digital banking. Heliyon, 9(8), Article e18930. https://doi.org/10.1016/j.heliyon.2023.e18930

Alshameri, F., & Xia, R. (2023). Credit card fraud detection: An evaluation of SMOTE resampling and machine learning model performance. International Journal of Business Intelligence and Data Mining, 23(1), 1–13. https://doi.org/10.1504/IJBIDM.2023.131791

Aziz, L. A. R., & Andriansyah, Y. (2023). The role of artificial intelligence in modern banking: An exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. Reviews of Contemporary Business Analytics, 6(1), 110–132.

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

Barroso, M., & Laborda, J. (2022). Digital transformation and the emergence of the Fintech sector: Systematic literature review. Digital Business, 2(2), Article 100028. https://doi.org/10.1016/j.digbus.2022.100028

Bello, O. A., Folorunso, A., Ejiofor, O. E., Budale, F. Z., Adebayo, K., & Babatunde, O. A. (2023). Machine learning approaches for enhancing fraud prevention in financial transactions. International Journal of Management Technology, 10(1), 85–109.

Blinowski, G., Ojdowska, A., & Przybyłek, A. (2022). Monolithic vs. microservices architecture: A performance and scalability evaluation. IEEE Access, 10, 20357–20374.

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). Explainable AI in fintech risk management. Frontiers in Artificial Intelligence, 3, Article 26. https://doi.org/10.3389/frai.2020.00026

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

Cao, L. (2022). AI in finance: Challenges, techniques, and opportunities. ACM Computing Surveys, 55(3), Article 64. https://doi.org/10.1145/3502289

Chen, J. S., Le, T. T. Y., & Florence, D. (2021). Usability and responsiveness of artificial intelligence chatbot on online customer experience in e-retailing. International Journal of Retail & Distribution Management, 49(11), 1512–1531. https://doi.org/10.1108/IJRDM-08-2020-0312

Chen, Y., Calabrese, R., & Martin-Barragan, B. (2024). Interpretable machine learning for imbalanced credit scoring datasets. European Journal of Operational Research, 312(1), 357–372. https://doi.org/10.1016/j.ejor.2023.06.036

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. International Journal of Information Management, 60, Article 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

de Lange, P. E., Melsom, B., Vennerød, C. B., & Westgaard, S. (2022). Explainable AI for credit assessment in banks. Journal of Risk and Financial Management, 15(12), Article 556. https://doi.org/10.3390/jrfm15120556

Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. European Journal of Operational Research, 297(3), 1178–1192. https://doi.org/10.1016/j.ejor.2021.06.053

Eren, B. A. (2021). Determinants of customer satisfaction in chatbot use: Evidence from a banking application in Turkey. International Journal of Bank Marketing, 39(2), 294–311. https://doi.org/10.1108/IJBM-02-2020-0056

Fritz-Morgenthal, S., Hein, B., & Papenbrock, J. (2022). Financial risk management and explainable, trustworthy, responsible AI. Frontiers in Artificial Intelligence, 5, Article 779799. https://doi.org/10.3389/frai.2022.779799

George, M. Z. H., Alam, M. K., & Hasan, M. T. (2023). Machine learning for fraud detection in digital banking: A systematic literature review. World Conference on Scientific Discovery and Innovation Proceedings, 3(1), 37–61. https://doi.org/10.63125/913ksy63

Gramespacher, T., & Posth, J.-A. (2021). Employing explainable AI to optimize the return target function of a loan portfolio. Frontiers in Artificial Intelligence, 4, Article 693022. https://doi.org/10.3389/frai.2021.693022

Hernandez Aros, L., Bustamante Molano, L. X., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. Humanities and Social Sciences Communications, 11(1), Article 1. https://doi.org/10.1057/s41599-024-03606-0

Hilal, W., Gadsden, S. A., & Yawney, J. (2021). 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

Hjelkrem, L. O., de Lange, P. E., & Nesset, E. (2022). The value of open banking data for application credit scoring: Case study of a Norwegian bank. Journal of Risk and Financial Management, 15(12), Article 597. https://doi.org/10.3390/jrfm15120597

Jenneboer, L., Herrando, C., & Constantinides, E. (2022). The impact of chatbots on customer loyalty: A systematic literature review. Journal of Theoretical and Applied Electronic Commerce Research, 17(1), 212–229. https://doi.org/10.3390/jtaer17010011

Lee, D. K. C., Guan, C., Yu, Y., & Ding, Q. (2024). A comprehensive review of generative AI in finance. FinTech, 3(3), 460–478.

Maestro, A., & Surantha, N. (2023). Scalability evaluation of microservices architecture for banking in public cloud. In Proceedings of the International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (pp. 103–115).

Meijer, W., Trubiani, C., & Aleti, A. (2024). Experimental evaluation of architectural software performance design patterns in microservices. Journal of Systems and Software, 218, Article 112183.

Milojević, N., & Redzepagic, S. (2021). Prospects of artificial intelligence and machine learning application in banking risk management. Journal of Central Banking Theory and Practice, 10(3), 41–57. https://doi.org/10.2478/jcbtp-2021-0023

Misischia, C. V., Poecze, F., & Strauss, C. (2022). Chatbots in customer service: Their relevance and impact on service quality. Procedia Computer Science, 201, 421–428. https://doi.org/10.1016/j.procs.2022.03.055

Ngo, H. Q., & Le, M. T. (2023). The role of perceived security and social influence on the usage behavior of digital banking services: An extension of the technology acceptance model. Edelweiss Applied Science and Technology, 7(2), 136–153. https://doi.org/10.55214/25768484.v7i2.396

Nicolescu, L., & Tudorache, M. T. (2022). Human–computer interaction in customer service: The experience with AI chatbots—A systematic literature review. Electronics, 11(10), Article 1579. https://doi.org/10.3390/electronics11101579

Saha, B., Rani, N., & Shukla, S. K. (2025). Generative AI in financial institution: A global survey of opportunities, threats, and regulation (arXiv:2504.21574). arXiv. https://doi.org/10.48550/arXiv.2504.21574

Weerasinghe, S., & Perera, I. (2023). Optimized strategy for inter-service communication in microservices. International Journal of Advanced Computer Science and Applications, 14, 272–279.