**International Journal of Financial Systems**

Volume 3, Number 2 (2025)
July-December 2025
Page : 129-158
P-ISSN : 3025-8480
E-ISSN : 3025-8537
DOI : https://doi.org/10.61459/ijfs.v3i2.89

Article history
Received : September 12, 2025
Revised : November 12, 2025
Accepted : December 9, 2025
Available online : December 31, 2025

# Artificial Intelligence and the Fragile Fortress: Addressing the Security and Data Privacy Gaps in Indonesia's Financial Sector Resilience

Yoseph Hendrik Maturbongs[1]

[1]Universitas Tarakanita, Indonesia
Email: yoseph.hendrik@utarki.ac.id

## ABSTRACT

Indonesia's financial sector embodies a paradox of a "fragile fortress": while Artificial Intelligence (AI) accelerates innovation and strengthens external facades, it simultaneously embeds systemic vulnerabilities within its digital architecture. To address this dichotomy, this article introduces the Algorithmic Digital Integrity Fault Lines Syndrome (ADIFL Syndrome) as a novel theoretical framework. ADIFL conceptualises six interlinked algorithmic risks, cyber-privacy vulnerability, data abuse by design, erosion of social legitimacy, infrastructure fragility, regulatory asymmetry, and AI service unreliability that collectively erode digital integrity. Using a qualitative-constructive methodology grounded in systemic theories and strategic case studies, the study demonstrates that these risks are not isolated incidents but manifestations of deeper structural fault lines. Findings reveal how compounding vulnerabilities undermine resilience and legitimacy, offering actionable policy recommendations to strengthen adaptive governance. Furthermore, the article proposes the Digital Integrity Risk Index (DIRI) as a future evaluative tool, complete with measurable indicators for each ADIFL subset. By framing AI risks as emergent and interconnected phenomena, this research provides a conceptual foundation for building anticipatory digital resilience, relevant not only to finance but also to critical domains such as public services, healthcare, and education.

**Keywords:** Artificial Intelligence, Financial Sector Resilience, ADIFL Syndrome, Digital Vulnerability, Policy Design

**JEL Classification:** G28, O33

## 1 Introduction

Indonesia's financial services sector now operates within an increasingly complex global landscape, marked by rising geopolitical tensions and the rapid acceleration of artificial intelligence driven digital transformation (Batool et al., 2025). In this context, financial sector resilience has become a vital element for national economic stability and growth (Cahyani & Maria, 2023). AI has emerged as a key catalyst for innovation, ranging from fraud detection and robo-advisory (Dellaert et al., 2024) to e-KYC automation (Hidayanti et al., 2025). The development and implementation of AI systems in Indonesia's banking sector are widely recognised as having the potential to transform the industry, drive innovation, and empower smarter decision-making (OJK, 2025). However, behind its efficiency potential lies a fundamental paradox: the very technology designed to strengthen resilience may inadvertently create new systemic vulnerabilities.

Despite the transformative promise of AI, Indonesia's financial services sector faces serious challenges in the form of significant vulnerabilities in cybersecurity and data privacy (Luthfah, 2024). The Financial Services Authority (OJK), in its report titled "Governance of Artificial Intelligence in Indonesian Banking," has identified several key vulnerabilities aligned with findings from the Financial Stability Board (FSB), including third-party dependency, service provider concentration, market correlation, cyber risks, model risks, data quality, and governance issues (OJK, 2025). These tensions are exacerbated by the suboptimal readiness of policy and regulatory frameworks, as highlighted by (Purba et al., 2025). Within the national regulatory context and global geopolitical dynamics, where reliance on foreign technology infrastructure and cross-border cyber threats are intensifying, these risks threaten the integrity of the financial system as a "digital fortress" that appears robust but is fragile within: a fragile fortress. These phenomena, which emerge systemically and inherently from the nature of AI algorithms, demand a new analytical framework capable of systematically explaining how these digital fault lines form and interact.

Behind the transformative potential of Artificial Intelligence (AI) in Indonesia's financial services sector (Bahoo et al., 2024), lies an inherent paradox where progress simultaneously generates significant vulnerabilities in cybersecurity and data privacy (Malatji & Tolah, 2025). The conflict between efficiency and vulnerability threatens the resilience of the national financial system, especially when the gap between technological penetration and policy readiness remains unresolved. In the context of Indonesia's regulatory landscape and increasingly complex global geopolitics (Iqbar et al., 2024), a comprehensive, adaptive, and responsive risk mitigation approach to technological challenges has yet to be optimally formulated. Therefore, the Governance of Artificial Intelligence in Indonesian Banking was developed as a complement to the digital transformation acceleration policy previously issued by OJK, with the hope of serving as a minimum reference for the banking sector in responsibly developing and implementing AI systems.

This condition underscores the need for a critical review of existing literature to understand why optimal risk mitigation approaches have not yet been formulated. Previous studies have identified AI threats in a fragmented manner, such as algorithmic bias (Panarese et al., 2025), automation service failures (Singh et al., 2025), and regulatory lag (Walter, 2024). While some studies have developed conceptual frameworks (e.g., AICD by Malatji & Tolah (2025), or Zhang et al.'s (2022) framework in cybersecurity contexts), no theoretical approach has explicitly captured and constructed the interrelation of risks as a systemic vulnerability syndrome, 'fault lines' that erode digital integrity structurally.

This article makes a significant contribution by developing and elaborating the Algorithmic Digital Integrity Fault Lines Syndrome (ADIFL Syndrome) as a new analytical framework and theory that addresses this gap. ADIFL Syndrome identifies and categorises inherent algorithmic digital integrity fault lines manifested in six critical dimensions: AI-Induced Cyber-Privacy Vulnerability (AICPV), Data-Abuse by Design (DABD), Algorithmic Social Legitimacy Breakdown (ASLB), Trust-Through-Infrastructure Failure (TTIF), Regulatory Asymmetry in Emerging Governance (RAEG), and AI Service Unreliability and System Failure (AISUF).

The primary objective of this article is to analyse the manifestations of ADIFL Syndrome within the Indonesian context, identify the root causes of these "digital fault lines," and formulate adaptive, responsive, and governance-based policy recommendations. To address the complexity of digital vulnerabilities outlined above, a theoretical foundation is needed that not only explains risks individually but also maps the systemic relationships among them. Therefore, the next section will elaborate on the key theories and literature analysis that underpin the construction of ADIFL Syndrome as a new analytical framework. The article is structured as follows: Section II presents the literature review and theoretical foundation; Section III outlines the research methodology; Section IV discusses findings and conceptual analysis; Section V concludes with theoretical contributions; and Section VI offers actionable policy recommendations.

## 2 Literature Review

This section presents the theoretical foundation and critical review of relevant literature to construct and justify the ADIFL Syndrome framework as the article's primary scholarly contribution.

To construct the ADIFL Syndrome framework, this research rests on two primary benchmarks. First, theoretically, this study does not rely on a single theory but instead builds a synthesis of five relevant systemic theories that serve as a multi-dimensional foundation. Second, in terms of prior research, the framework is positioned directly against the most current preceding studies to clearly identify the existing conceptual gap and ensure the novelty of the contribution being offered. Both benchmarks will be elaborated upon in detail in the following section.

### 2.1 Theoretical Foundations of ADIFL Syndrome

To understand and elaborate the Algorithmic Digital Integrity Fault Lines Syndrome (ADIFL Syndrome), this article draws upon several systemic theories:

#### 2.1.1 Socio-Technical Systems Theory (STS)

Originally developed by Eric Trist and the Tavistock Institute, STS emphasises the harmonious design of technological and social systems. Technology is viewed as an integral part of the social system, not an external entity (Trist, 1981) . Rafael (2013) further argues that technology functions as a societal subsystem, observing tools, techniques, and applications through the lens of "state-of-the-art obsolete." STS provides a lens to analyse how ADIFL fault lines emerge from disharmony between technical components (AI algorithms) and social components (users, regulators, governance systems). In the context of ADIFL, STS explains how vulnerabilities such as DABD and ASLB arise from misalignment between algorithmic design and socio-regulatory structures.

### 2.1.2 Complex Adaptive Systems (CAS)

CAS theory posits that AI-based digital systems are non-linear and adaptive, allowing vulnerabilities to emerge unpredictably. Holland demonstrated that complex system behaviour cannot be deterministically predicted (Holland, 1992) . Tristan Lim notes that ESG, AI, and finance research landscapes exhibit evolving techniques, domain differentiation, and unique interactions between digital systems and sustainability contexts (Lim, 2024) . CAS supports the analysis of emergent systemic failures such as AISUF and TTIF, which arise from interactions among ADIFL sub-sets.

### 2.1.3 Information Asymmetry Theory (IAT)

Introduced by Akerlof and expanded by Stiglitz & Weiss, IAT explains how information imbalances between regulators and industry actors can lead to systemic risks. In the ADIFL context, this theory underpins the sub-set Regulatory Asymmetry in Emerging Governance (RAEG) (Akerlof, 1978; Stiglitz & Weiss, 1981) . In Indonesia, AI regulation in finance lags behind innovation, creating information gaps that allow algorithmic risks to grow unchecked (Pradana et al., 2025).

### 2.1.4 Trust Theory

Developed by Luhmann (1979) and Lewis & Weigert (1985), this theory posits that trust reduces social complexity more effectively than prediction. In ADIFL, sub-sets ASLB and TTIF illustrate how algorithmic and infrastructure failures erode social and institutional trust in financial systems. Trust Theory enables analysis of how ADIFL manifestations, such as service failures or privacy breaches, undermine trust at both social and institutional levels.

### 2.1.5 Technological Mediation Theory (TMT)

Developed by Peter-Paul Verbeek and other postphenomenologists, TMT asserts that technology is not neutral (Verbeek, 2023), but mediates human-world relations, shaping perceptions, values, and social actions. Verbeek emphasises that technology shapes social perception and action through mediation structures involving amplification and reduction of interpretation. TMT is particularly relevant to DABD and ASLB, explaining how algorithmic design inherently facilitates data exploitation and how AI mediates perceptions of fairness and trust through opaque or biased processes.

These five theories were selected strategically based on a review of seminal and contemporary literature demonstrating their widespread use in studies of AI risk, digital governance, and systemic vulnerability in finance. STS explains the interaction between algorithmic design and social structures; CAS outlines the emergent nature of digital system failures; IAT highlights information gaps between regulators and developers; Trust Theory examines the impact of technological failure on social legitimacy; and TMT offers a philosophical dimension on how technology mediates values and perceptions. Together, they form a multidimensional foundation for ADIFL Syndrome, positioning it not as an isolated concept but as a synthesis of established theoretical insights.

## 2.2 Previous Studies and Research Gaps

Existing literature has identified various individual risks associated with AI adoption in finance. Studies have extensively explored AI-enhanced cybersecurity threats (Singh et al., 2025), data privacy violations (Malatji & Tolah, 2025; Zhang et al., 2022), algorithmic bias (Panarese et al., 2025), and regulatory lag (Walter, 2024).

Some studies have developed robust conceptual frameworks. For example, Zhang et

al. (2022) proposed an AI application framework in cybersecurity encompassing access authentication, anomaly monitoring, and situational awareness. Malatji & Tolah (2025) introduced the AICD Framework, integrating attack types, mitigation strategies, actor motivations, and social impacts in AI systems. These studies indicate a shift toward more integrative approaches in academic literature.

However, a conceptual gap remains: the absence of a theoretical framework that explicitly constructs these risks as a systemic vulnerability syndrome, where risks interact and reinforce one another. Studies like Walter (2024) and Singh et al. (2025) underscore the importance of understanding system dysfunction and regulatory imbalance, but often treat them as separate entities rather than manifestations of deeper algorithmic structural fractures.

Global institutions like the Financial Stability Board (FSB) have identified key vulnerabilities threatening financial stability due to AI (OJK, 2025). Yet, ADIFL Syndrome goes further by offering a systemic analytical lens to explain how and why digital fault lines form, interact, and intensify. Rather than viewing vulnerabilities as isolated threats, ADIFL conceptualises them as architectural fragilities, an inherent syndrome that structurally erodes digital integrity and renders financial systems a "fragile fortress."

The core contribution of ADIFL Syndrome is its synthesis of these risk dimensions into a unified framework that explains why and how financial systems can appear digitally robust yet architecturally fragile. ADIFL not only responds to empirical challenges but also expands the conceptual horizon of systemic risk in AI-driven financial governance by identifying diverse risks as manifestations of a shared algorithmic architectural fragility.

### 2.3 Official Definition of ADIFL Syndrome and Its Subsets

As the article's principal theoretical contribution, this section formally defines the Algorithmic Digital Integrity Fault Lines Syndrome (ADIFL Syndrome) along with its six interrelated subsets.

ADIFL Syndrome is defined as a theoretical framework that explains patterns of digital integrity vulnerabilities and failures that are systemic and inherent, not merely incidental. These patterns arise from the nature and complexity of artificial intelligence algorithms embedded within the digital systems of the financial sector. ADIFL not only identifies individual digital fault lines but also elucidates the dynamic interconnections among these vulnerabilities as a unified systemic syndrome that undermines sectoral resilience.

The six subsets of ADIFL Syndrome, each representing a manifestation of these "fault lines," are defined as follows:

#### 2.3.1 AI-Induced Cyber-Privacy Vulnerability (AICPV)

This subset refers to vulnerabilities in data protection and cybersecurity that emerge or are exacerbated by the deployment of AI, including automated attacks, unauthorised processing of sensitive data, and deviations from default privacy principles. (Zhang et al., 2022) observed that AI contributes to heightened security threats in authentication and anomaly detection systems, particularly within e-KYC contexts. Contemporary scholarship reinforces the critical nature of AICPV. Recent studies indicate that financial AI models are increasingly susceptible to adversarial attacks, where malicious inputs manipulate algorithmic outcomes without triggering traditional firewalls (Fursov et al., 2021; James et al., 2024). Furthermore, the integration of extensive datasets in Open Banking ecosystems has introduced new vectors for model inversion attacks, allowing attackers to reconstruct sensitive financial data from model outputs (Fredrikson et al., 2015; Parisot et

al., 2021; Veale et al., 2018). Consequently, scholars argue that conventional cybersecurity frameworks are insufficient, necessitating a shift towards "AI-native security" paradigms that address these specific algorithmic vulnerabilities (Sankalp et al., 2025).

In the Indonesian context, these vulnerabilities manifest acutely where biometric e-KYC systems deliver speed and compliance gains but simultaneously widen the attack surface for personal data. Legal-procedural reviews indicate that while digital onboarding accelerates inclusion, its safety hinges precariously on how banks manage biometric templates and access controls under the country's evolving data-protection regime; weak operational practices can quickly translate into privacy risks at scale (Fitriyanti et al., 2024). Moreover, a secondary fault line is the surge in synthetic identity fraud. Industry evidence points to rising losses, suggesting that without robust privacy architectures, such as on-device matching and strict role-based access scaling biometrics can amplify data-breach impacts and enable "function creep," thereby eroding trust in digital finance.

The rapid infusion of Generative AI adds another layer of complexity. While GenAI accelerates threat detection, surveys from central banks suggest it raises social-engineering risks and demands significant investment in human capital to keep models aligned with legal boundaries. Collectively, these sources position AICPV not merely as a technical glitch, but as a multi-layered "fault line" spanning technical design, organisational operations, and ecosystem dependencies. Without an integrated mitigation program, pairing privacy engineering with strict governance, the efficiency gains of AI in Indonesia's financial sector risk becoming systemic privacy vulnerabilities.

### 2.3.2 Data-Abuse by Design (DABD)

This fault line arises when the technical architecture of AI systems inherently enables or even encourages, the unethical and non-transparent collection, processing, or misuse of data. Zuboff (2023) argues that the design logic of modern digital systems is often built to maximise data extraction, making exploitation a systemic feature rather than a technical anomaly.

This systemic exploitation is empirically supported by recent investigations into dark patterns in FinTech interfaces, which manipulate user consent to maximise data extraction (Rakovic, 2022; Rakovic & Inal, 2023). Moreover, the concept of "privacy engineering failure" highlights how rapid development cycles in startups often deprioritise privacy safeguards in favor of aggressive user profiling (Gupta et al., 2023; Nguyen et al., 2021). Critics warn that without strict "ethical-by-design" mandates, the financial sector risks normalising a surveillance-based business model that fundamentally erodes consumer sovereignty (Li, 2025; Power, 2022).

### 2.3.3 Algorithmic Social Legitimacy Breakdown (ASLB)

This subset describes the erosion of public trust due to algorithmic bias, unaccountable automated decisions, and the absence of social accountability in AI systems. Eubanks (2018) illustrates how biased algorithms in welfare distribution systems undermine public legitimacy and reinforce structural discrimination.

The breakdown of legitimacy is further evidenced by studies on algorithmic fairness in credit scoring, which found that "neutral" algorithms frequently penalise unbanked populations due to historical data biases (Trinh & Zhang, 2024). The lack of contestability and the ability for users to challenge automated decisions, has been identified as a primary driver of public distrust (Alfrink et al., 2022; Bayamlıoğlu, 2022). Sociological analyses suggest that when financial institutions deploy "black box" models without transparency,

they violate the implicit social contract, leading to a profound deficit in institutional trust (Lu, 2022; Thalpage, 2023; Eschenbach, 2021).

### 2.3.4 Trust-Through-Infrastructure Failure (TTIF)

This fault line refers to breakdowns or excessive dependence on critical digital infrastructure such as AI cloud services, cross-border network systems, or public APIs. When disrupted, these systems directly diminish institutional and public trust in financial systems. Insuretech Asia (2023) reported that failures in cloud-based underwriting systems led to widespread service disruptions and consumer trust crises.

This fragility is exacerbated by cloud concentration risk, where the financial sector's over-reliance on a few dominant AI service providers creates single points of failure (Ryan et al., 2024). Furthermore, the complexity of interconnected API ecosystems means that a failure in one AI component can trigger cascading disruptions across the entire banking network (Basak & Tiwari, 2025; Cate, 2025). Resilience scholars emphasise that current disaster recovery protocols are often ill-equipped to handle such synchronised algorithmic failures, threatening the stability of the entire financial infrastructure (Iyer, 2022; Karakasilioti, 2024; Neumannová et al., 2023).

### 2.3.5 Regulatory Asymmetry in Emerging Governance (RAEG)

This subset highlights the chronic gap between the rapid evolution of AI and the capacity of existing regulatory frameworks, creating policy grey zones and widening accountability gaps. Veale & Edwards (2018) critique how AI development often outpaces legislative responsiveness, resulting in digital power imbalances, a concern echoed by Trisnawati (2024).

This asymmetry is theoretically framed as the "pacing problem," where technological innovation exponentially outpaces legislative adaptation (Lachmann, 2025). In the financial context, this results in "regulatory arbitrage," allowing FinTech firms to exploit grey zones in AI governance before regulators can intervene (Grennan, 2022; Kandpal et al., 2025; Ravshan, 2025). Comparative studies suggest that traditional static regulations are failing, prompting a global shift toward adaptive governance models like regulatory sandboxes to bridge this widening gap (Novelli et al., 2025; Yaksan, 2024).

### 2.3.6 AI Service Unreliability and System Failure (AISUF)

This fault line encompasses operational instability in AI systems, such as erroneous outputs, automation failures, or service disruptions that compromise the integrity and continuity of financial systems. Singh et al. (2025) note that AI-based credit scoring services in Southeast Asia exhibit high false rejection rates with opaque logic, indicating systemic failure potential.

Technical literature attributes these failures to model drift, where AI performance degrades as real-world data diverges from training data, a common occurrence in volatile financial markets (Sun et al., 2024). Additionally, the phenomenon of "automation surprise", where operators fail to intervene during AI errors due to over-trust has been cited as a key factor in service outages (Agudo et al., 2024; Romeo & Conti, 2025). These findings underscore that reliability is not just a coding issue but a dynamic challenge requiring continuous algorithmic auditing and robust "fail-safe" mechanisms (Minkkinen et al., 2022; Zhang et al., 2025).

By defining ADIFL Syndrome both conceptually and operationally through these six interconnected subsets, the article establishes a robust theoretical foundation for

understanding algorithmic vulnerabilities as a systemic syndrome that weakens the resilience of digital financial systems. This definition is not merely descriptive but is designed to be analytically operationalised and used as an evaluative tool for assessing real-world dynamics in Indonesia.

To test the utility and relevance of the ADIFL framework within the national financial services context, the next section outlines the methodological approach employed in this study. Section III will detail the research design, literature selection strategy, and conceptual analysis framework underpinning the findings and policy recommendations.

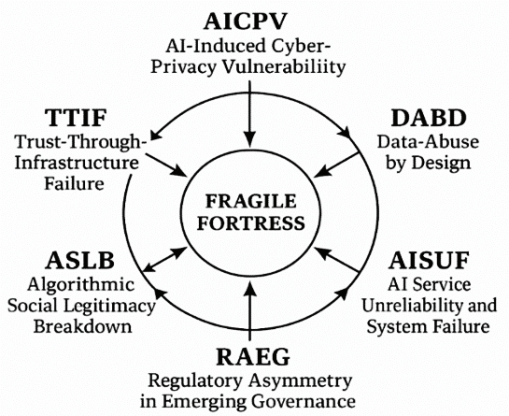


**Figure 1.** Illustrating the Algorithmic Digital Integrity Fault Lines Syndrome (ADIFL Syndrome) framework

## 3 Methodology

To construct the ADIFL Syndrome theoretical framework in a systematic and structured manner, this section outlines the methodological approach used in the study. The article adopts a qualitative-constructive approach, emphasising theory development based on systematic literature review, thematic analysis, and deductive-abstract reasoning. This process is supported by comparative policy studies to contextualise the empirical relevance of ADIFL in both Indonesian and global digital financial governance.

The central methodological aim of this research is to construct a new conceptual framework by synthesising insights from five established systemic theories. The following sections will detail the systematic process undertaken to achieve this synthesis.

### 3.1 Research Design Justification and Problem Framing

This study begins from problematisation: AI adoption in financial services is accelerating, yet authorities warn of monitoring gaps and policy sufficiency issues, with vulnerabilities concentrated in third-party dependencies, service-provider concentration, market correlations, cyber risk, and model governance (FSB, 2024). In a banking context like Indonesia and ASEAN, these risks intersect with nascent sector-specific governance frameworks (MAS, 2018; OJK, 2025).

Given the heterogeneity of sources (policy reports, sector toolkits, academic articles) and the mixed maturity of measurement, this study does not adopt meta-analysis (which presupposes comparable effect sizes) nor purely grounded theory (which presupposes minimal prior structure). Instead, following MacInnis's (2011) guidance on conceptual

contribution types and Jaakkola's (2020) design templates for non-empirical work, we employ a conceptual "Typology/Model" design. This design is purpose-built to synthesise fragmented literatures, bridge disciplines, and construct testable propositions under clear boundary conditions, making explicit the role of theory and policy frameworks in a non-empirical article. For local relevance, the model is anchored in the OJK Artificial Intelligence Governance for Banking (focusing on lifecycle governance, reliability, accountability) and the MAS FEAT/Veritas Toolkit, treating these as boundary conditions for claims regarding risks and mitigations in the ASEAN region.

Justification of Method: A conceptual typology model enables the integration of FSB's vulnerability blocks with sector toolkits (OJK/MAS) and academic literature, mapping where risks originate, how they propagate, and which governance instruments mitigate them. This approach explicitly addresses the methodological necessity of explaining the why and how of systemic reasoning, ensuring the research design fits the emergent nature of the evidence and aligns practically with sector governance.

### 3.2 Phases of ADIFL Syndrome Theory Development

To transition from disparate empirical observations to a unified theoretical construct, this study employed a systematic developmental protocol. This process was designed to ensure methodological rigor and analytical depth, moving beyond simple description to achieve conceptual abstraction. The development trajectory followed a logic of recursive synthesis: starting with the deconstruction of anomalies in current risk literature, progressing through inductive pattern recognition, and culminating in deductive theoretical grounding. This structured approach ensures that the ADIFL Syndrome is established not merely as a heuristic device, but as a robust theoretical framework capable of explaining the systemic nature of digital fragility. The construction process is delineated into four distinct phases:

#### 3.2.1 Phase 1: Deconstruction of Anomalies (Literature Identification)

Consistent with the "fragmented literature" problem identified in the research design, this phase focused on capturing risk phenomena that defied standard classification. The process began with the selection and critical review of scholarly literature related to AI adoption risks in the financial sector. We conducted a structured search of academic databases (Scopus Q1-Q2, ScienceDirect) and grey literature (FSB, OJK reports) from 2019–2025. The analytical filter was set to isolate "architectural anomalies", specifically, incidents where compliant systems failed (e.g., privacy breaches occurring despite robust encryption). Included literature was critically assessed not merely to list risks, but to identify structural flaws that standard frameworks failed to explain, providing the raw material for the typology construction.

#### 3.2.2 Phase 2: Conceptualisation and Pattern Recognition

Moving from raw data to theoretical constructs, this phase employed inductive thematic analysis. The primary objective was to synthesise the isolated anomalies identified in Phase 1 into a coherent conceptual structure. We necessitated an inductive approach because existing frameworks typically categorise risks based on "symptoms" (e.g., cyber risk vs. legal risk), which often obscures the underlying causal mechanisms connecting them.

To address this, we coded the identified anomalies to detect recurring structural patterns. This process involved a level of abstraction: for instance, disparate incidents such as "dark patterns" in user interfaces and "excessive data permissions" in fintech apps were clustered not merely as compliance issues, but as a unified design-logic flaw termed "Data-

Abuse by Design." Following MacInnis's (2011) and Jaakkola's (2020) framework for conceptual contribution, this synthesis transcended descriptive categorisation to identify "systemic architectural fragility." This rigorous abstraction process crystallised six distinct "risk clusters," forming the proto-typology of the ADIFL framework and establishing the logical premise for a "syndrome" rather than a simple list of threats.

#### 3.2.3 Phase 3: Derivation of ADIFL Sub-sets

In this phase, the six conceptual clusters were formally derived as the constituent sub-sets of the ADIFL framework (AICPV, DABD, ASLB, TTIF, RAEG, AISUF). Crucially, to ensure the local relevance argued in the research design, these definitions were calibrated against boundary conditions specific to the Indonesian and ASEAN banking context. We employed source triangulation, cross-referencing theoretical definitions with regulatory artifacts, such as OJK's specific requirements for data sovereignty and MAS's FEAT principles. This rigorous calibration ensured that the resulting typology was operationally valid for developing economies, distinguishing structural "fault lines" that require systemic intervention from transient operational errors.

#### 3.2.4 Phase 4: Systemic Grounding and Theoretical Integration

The final phase involved the deductive validation of the framework to ensure its theoretical robustness. Having inductively derived the sub-sets in the previous phases, we then mapped them against five core systemic theories (STS, CAS, IAT, Trust Theory, and TMT) to test their structural logic. This step served as a theoretical "stress test": for example, the linkage between Data-Abuse by Design (DABD) and Algorithmic Social Legitimacy Breakdown (ASLB) was validated using Technological Mediation Theory, confirming that technical design choices actively mediate social trust. By anchoring the sub-sets in these established theories, this phase transformed the framework from a simple descriptive taxonomy into a unified "Syndrome," confirming the central thesis that digital fragility is a systemic, emergent property rather than a collection of isolated errors.

### 3.3 Case-Based Analysis

As contextual verification, a strategic case study analysis approach was used to examine the manifestation of ADIFL in Indonesia. The method employed is qualitative content analysis of incident reports, relevant policy responses, and reports from trusted media outlets (Kompas, Tempo, Bisnis Indonesia) and regulators (OJK, BI). Case selection criteria were based on: (1) Direct relevance to one or more ADIFL sub-sets, (2) Significant impact (e.g., customer service chatbot errors, fraud due to automated scoring systems, or data breaches involving AI-based systems), and (3) Availability of sufficient public data for analysis.

The analysis was conducted using the directed qualitative content analysis method (Zhang & Wildemuth, 2009), which enables categorisation of incidents based on the theoretically derived ADIFL dimensions. This approach was chosen for its ability to bridge theory and contextual data systemically.

### 3.4 Ethical Considerations

This research is non-empirical and based on secondary literature, yet it upholds principles of academic integrity, interpretive transparency, and objectivity. Risk interpretations were conducted critically and free from conflicts of interest with any financial institutions or entities. All citations and references are explicitly included to ensure traceability and scholarly accountability.

Additionally, the researcher consciously avoided speculative interpretations and ensured that all data were sourced from publicly verifiable archives to prevent privacy violations or contextual bias.

Through this methodological approach, the article not only formulates ADIFL as a conceptual construct but also reinforces it through a systematic process and relevant contextual studies. The next section presents the conceptual findings and theoretical discussion derived from this methodological framework.

### 3.5 Validation, Justification, and Methodological Limitations

To ensure the conceptual validity of the ADIFL Syndrome framework, triangulation was conducted across literature sources and cross-confirmation with the systemic theories previously outlined. Triangulation involved comparing vulnerability patterns found in various case studies, academic literature, and policy frameworks, then testing their consistency against five core theories (STS, CAS, IAT, Trust Theory, and TMT).

This approach enabled the researcher to verify that ADIFL sub-sets are not only theoretically relevant but also have strong empirical correspondence within Indonesia's digital financial sector. Triangulation was performed not only across sources but also across disciplines (AI studies, financial regulation, and social systems theory), and across directions (deductive from theory to case and inductive from case to concept construction) to ensure the integrity and relevance of the ADIFL framework.

Case selection was strategically conducted to allow theoretical generalisation of the ADIFL framework, despite the non-empirical nature of the research. Cases were chosen based on their direct relevance to ADIFL sub-sets, their systemic impact on the financial sector, and the availability of verifiable public data. Using directed qualitative content analysis, case studies served not only as illustrations but also as contextual verification of the developed theoretical construct.

While this conceptual approach enables the development of a systemic and multidimensional theoretical framework, several limitations must be acknowledged: Since the research is based on secondary literature and does not involve primary data collection, empirical validation of ADIFL Syndrome remains limited. Generalisation of findings depends on the representativeness of case studies and the quality of sources used. This approach does not yet capture the quantitative dimensions of systemic risk, such as failure probabilities or impact distributions.

Future research is needed to empirically test this framework across sectors using mixed methods, such as risk perception surveys or regression analysis of AI failure incidents, to quantitatively validate ADIFL. Additionally, as an early-stage conceptual theory, ADIFL Syndrome has not yet integrated probability estimates or impact distributions. Therefore, further studies are recommended to develop quantitative instruments, such as digital risk measurement scales or complex network modeling, to mathematically verify the systemic structure among ADIFL sub-sets.

## 4 Result and Discussion

As the central pillar of this theoretical framework, the Result & Discussion section seeks to articulate the tangible manifestations of the Algorithmic Digital Integrity Fault Lines Syndrome (ADIFL Syndrome), as previously conceptualised and elaborated. Employing a systematic methodology that integrates literature review, case study analysis, and theoretical interpretation, this section presents a nuanced examination of six ADIFL

sub-sets. The analysis is structured into three layers: What/How (empirical evidence and case descriptions), Why (conceptual reasoning and systemic roots), and What Else (structural implications and global relevance).

This approach not only affirms the existence of each "fault line" but also elucidates their interrelations as components of a broader systemic syndrome. In doing so, it positions ADIFL as a novel evaluative lens through which the fragility of Indonesia's financial digital resilience can be understood amidst the accelerating adoption of artificial intelligence.

### 4.1 AICPV: AI-Induced Cyber-Privacy Vulnerability

The AICPV sub-set highlights vulnerabilities in data privacy and cybersecurity that are exacerbated by the integration of AI technologies within the financial sector, particularly in e-KYC systems and digital infrastructures (Cypriva, 2025). Between 2019 and 2024, multiple data breach incidents were reported, stemming from weak encryption protocols and the absence of algorithmic audits. Notable examples include the leakage of two million customer records from BRI Life (Detiknet, 2021) and unauthorised access manipulation on the Jenius platform (Shahnaz, 2021).

Reports from the OJK Institute further underscore deficiencies in facial recognition systems and the storage of biometric data without adequate safeguards. Zhang et al. (2022) argue that AI-driven cybersecurity systems may inadvertently create new blind spots, particularly when deployed for authentication or sensitive data management, as algorithmic decisions often lack real-time explainability.

Within the ADIFL framework, AICPV is not merely a technical anomaly but a symptom of a deeper structural disjunction between algorithmic design and the principles of privacy governance. Addressing this issue requires systemic mitigation strategies, including AI privacy audits, sandbox testing environments, and cross-vendor security validation to uphold the integrity and trustworthiness of digital financial systems.

The fragility observed in AICPV can be interpreted through the lens of Complex Adaptive Systems (CAS), wherein AI technologies form intricate networks of interdependence, producing emergent effects that defy linear prediction. In this context, cyberattacks or privacy breaches do not originate from isolated vulnerabilities but emerge from dynamic interactions among code, data, and infrastructure.

Additionally, Information Asymmetry Theory sheds light on the disparity in understanding between end-users and system administrators. In e-KYC systems, customers are often unaware of how their data is processed by AI, while financial institutions themselves may lack full visibility into the operations of third-party providers.

The implications of AICPV are profound within the metaphor of the "fragile fortress." A financial system that appears digitally robust can suffer a rapid erosion of public trust following a single breach. Unlike technocratic approaches that focus narrowly on enhancing firewalls or encryption, ADIFL interprets AICPV as a structural fault, specifically, the disconnect between algorithmic design and privacy governance.

Globally, the European Union's AI Act has designated biometric recognition systems as "high-risk," while the Monetary Authority of Singapore (MAS) mandates data minimisation principles in financial AI systems (MAS, 2024). These developments reflect a growing international recognition of AICPV as a systemic risk, though such awareness remains limited in the Indonesian regulatory landscape.

### 4.2 DABD: Data-Abuse by Design

DABD encapsulates vulnerabilities that arise from AI system architectures which, by

design, facilitate the exploitation of user data. In Indonesia, this phenomenon is evident in fintech applications that overreach in accessing personal data, such as contacts, location, and media files, and in startups that monetise behavioral data without transparency (aisensum, 2019). This exploitation is not a mere technical oversight but a feature embedded within the system's design logic.

Research by Panarese et al. (2025) reveals that credit scoring algorithms do more than assess financial risk, they actively shape consumer behaviour, positioning AI as a tool for constructing social realities. Consequently, DABD calls for mitigation strategies grounded in ethical design, transparent data practices, and comprehensive algorithmic audits to ensure fairness and accountability in financial AI systems.

The Technological Mediation Theory (TMT) offers a critical perspective: technologies, including AI algorithms, are not neutral instruments but active mediators that influence human actions, perceptions, and social relationships. When AI systems are designed without ethical foresight, they may inadvertently mediate exploitative patterns through interface design, computational logic, and incentive structures geared toward maximising data extraction (Boivin, 2025).

Moreover, Socio-Technical Systems Theory (STS) emphasises that such exploitation is not merely a technical flaw but the outcome of complex interactions among system design, permissive regulatory frameworks, and organisational cultures driven by data monetisation (Hughes et al., 2017). In essence, DABD reflects "systemic incentives" embedded across multiple layers, from algorithmic architecture to corporate policy.

DABD significantly undermines digital system resilience by creating a persistent tension between business efficiency and user rights. Over time, this not only erodes public trust but also heightens legal and reputational risks. Internationally, the EU General Data Protection Regulation (GDPR) enshrines privacy by design as a legal mandate (EU, 2016), while MAS advocates for algorithmic audits of user data processing systems (MAS, 2024). Unfortunately, Indonesia has yet to adopt a systemic framework for regulating design ethics at the engineering stage, allowing DABD to persist as a concealed yet strategically critical fault line.

### 4.3 ASLB: Algorithmic Social Legitimacy Breakdown

The ASLB sub-set reflects the erosion of social legitimacy in algorithmic systems, driven by bias, discrimination, and a lack of accountability. In Indonesia's financial sector, this is evident in the unequal access to financing for vulnerable groups such as informal MSMEs and female-headed households, who are often marginalised by AI-based credit scoring systems (Cristine et al., 2025).

Although there is no explicit evidence of systemic exclusion, various reports highlight structural barriers, such as low digital literacy and the absence of formal documentation, that diminish these groups' chances in digital financing ecosystems. Imbalanced training data and non-inclusive algorithmic design further amplify the risk of undetected discrimination (Respati & Sukmana, 2023).

This phenomenon is mirrored globally, with examples including racial bias in facial recognition systems (Buolamwini & Gebru, 2018), algorithmic censorship of activists (Alrasheed & Lim, 2021), and Amazon's AI recruitment tool that perpetuated gender disparities (Chang, 2023) . These cases underscore how algorithms, if left unchecked, can reinforce social injustice.

ASLB thus calls for the implementation of Explainable AI, fairness audits, and public appeal mechanisms as part of a governance framework that ensures algorithmic justice

and preserves the social legitimacy of digital financial systems.

The emergence of ASLB can be explained through the synergy of Trust Theory and Information Asymmetry Theory. When users are unable to understand how algorithmic decisions are made, or are denied the opportunity to challenge or correct those decisions, trust becomes fragile. What appears to be a technical decision often carries significant social consequences, yet lacks avenues for public participation or institutional safeguards.

ASLB is also closely tied to institutional failure in ensuring procedural fairness, as outlined in Socio-Technical Systems Theory (STS). AI systems do not merely automate services; they shift the burden of proof from institutions to individuals, exacerbating power imbalances between service providers and users. When society no longer perceives AI as a fair tool, the legitimacy of the system erodes collectively.

ASLB poses a long-term threat to the resilience of digital financial systems by undermining the social consensus that underpins technological adoption. In the global context, European jurisprudence, such as the Schufa ruling by the Court of Justice of the European Union, has declared that automated decisions without human oversight may violate fundamental rights. Singapore, meanwhile, has adopted Explainable AI policies and the AI Fairness Toolkit as preventive measures against legitimacy crises.

In contrast, Indonesia currently lacks adequate legal mechanisms to facilitate appeals against algorithmic decisions, reinforcing the urgency of ASLB as a fault line that must be anticipated and addressed.

### 4.4 TTIF: Trust-Through-Infrastructure Failure

The TTIF sub-set captures the breakdown of trust resulting from disruptions or excessive dependence on critical digital infrastructure in the financial sector. A notable example occurred in February 2025, when BSI's BYOND application experienced a 72-hour downtime, triggering public panic and exposing systemic fragility (Tempo, 2025).

A report by Hitachi Vantara (Javia, 2025) revealed that most global financial institutions deploy AI without sandbox environments or failover protocols, heightening the risk of systemic disruption. In Indonesia, low digital literacy and the absence of AI infrastructure audits exacerbate this vulnerability, as acknowledged by OJK (Sandy, 2025).

TTIF is not merely a technical glitch, it represents a systemic failure in maintaining reliability and transparency. A study by INSURETECH Asia (2023) warns that AI infrastructure lacking redundancy can trigger a domino effect, undermining institutional trust.

According to Socio-Technical Systems Theory (STS), such failures cannot be explained solely through technical shortcomings. AI infrastructure should be developed as an adaptive system that accounts for the reciprocal relationships between technology, users, institutions, and organisational contexts. When infrastructure architecture lacks resilience, due to minimal redundancy, dependence on foreign vendors, or weak interoperability, social trust collapses, even in the absence of malicious intent.

Trust Theory emphasises that trust is not merely a product of functional performance, but also of consistency, predictability, and responsiveness during crises. TTIF arises when the digital ecosystem fails to guarantee these qualities, especially when damaged infrastructure layers are inaccessible, opaque, or unaccountable to the public.

Dependence on global vendors (such as AI-as-a-Service providers) introduces geopolitical risks and challenges to technological sovereignty. Europe addresses this through the principle of digital strategic autonomy in its cloud policy, while Singapore mandates AI Infrastructure Resilience Assessments for large-scale financial entities.

Indonesia, however, lacks explicit provisions for AI infrastructure audits within OJK or BI regulations, highlighting the strategic importance of TTIF as a foundational concern for systemic resilience.

### 4.5 RAEG: Regulatory Asymmetry in Emerging Governance

The RAEG sub-set illustrates the growing gap between the pace of AI innovation and the readiness of regulatory frameworks in Indonesia's financial sector. Technologies such as machine learning and robo-advisory have been in use since 2021–2022, yet remain largely unregulated in terms of algorithmic transparency, decision accountability, and ethical safeguards (Susilo, 2023). AI system audits continue to face technical challenges and capacity limitations, as acknowledged by OJK in its banking AI governance documentation (OJK, 2025).

The information asymmetry between regulators and technology developers has created a policy grey zone, allowing algorithmic risks to proliferate without sufficient oversight (Puannandini et al., 2025). Regulatory delays in understanding AI's technical architecture and operational logic have led to slow policy responses, amplifying systemic risks and weakening the overall governance of digital innovation.

Advanced economies have begun adopting risk-based governance principles to regulate AI. The European Union, for instance, classifies AI systems into risk tiers under the EU AI Act, assigning levels of oversight based on potential harm (EU, 2016). Singapore employs a Sandbox Plus approach, enabling controlled experimentation under adaptive supervision (MAS, 2024). Indonesia could benefit from adopting similar frameworks to fill regulatory gaps and prevent normative lag that could lead to long-term trust deficits.

### 4.6 AISUF, AI Service Unreliability and System Failure

The AISUF sub-set captures systemic and disruptive operational failures in AI systems, despite their widespread adoption in the financial sector. In Indonesia, incidents such as digital bank chatbots providing incorrect information on interest rates and loan tenures (Teknologi.id, 2023 ) have directly impacted customer financial decisions, highlighting the absence of real-time validation and human-in-the-loop mechanisms.

The digital service outage at Bank DKI in March 2025 revealed weaknesses in responsive architecture, as automated recovery features failed, causing a three-day disruption in banking services (Setyowati, 2025). This incident underscores the lack of rigorous system testing and inadequate IT governance within AI infrastructure.

Research by Singh et al. (2025) reinforces that systemic AI failures stem not only from technical errors but also from a lack of algorithmic transparency and the absence of audit mechanisms. AISUF calls for strengthened oversight systems, reliability audits, and more robust AI design to safeguard the integrity and public trust in digital financial services.

These failures can be understood through Socio-Technical Systems Theory (STS), which emphasises the integration of technology, human oversight, and organisational procedures. AISUF reveals the consequences of excessive reliance on automated systems without sufficient human supervision.

Additionally, Trust Theory explains how such dysfunctions immediately erode public confidence, not only in service providers but in AI systems as a whole. When users encounter errors in AI interactions, institutional credibility suffers, triggering broader resistance to technological adoption.

A study by Dorochowicz et al. (2025) shows that AI system reliability in finance hinges on two key factors: technical robustness and operational clarity. While South Korea does

not explicitly mandate real-time failover or AI audit logs, its digital system design implicitly integrates these practices to enhance resilience and transparency (Bank for International Settlements, 2023). Singapore, through MAS, has formalised AI audit standards and system monitoring as part of financial sector operations (MAS, 2024).

In Indonesia, such approaches remain underutilised, particularly in the non-bank fintech sector, making AISUF the most vulnerable ADIFL sub-set in the short term.

### 4.7 Synthesis & Implications: Mapping the Fragile Fortress

The six sub-sets of the ADIFL Syndrome, AICPV, DABD, ASLB, TTIF, RAEG, and AISUF, do not exist in isolation. Rather, they reinforce one another in shaping the systemic vulnerabilities of the digital financial services sector. Each sub-set represents a distinct fault line that may appear isolated, but in reality, they interact through complex relationships among technology, policy, and social dynamics.

This interconnectedness forms what this article refers to as the "Fragile Fortress": a condition in which financial digital infrastructure appears technologically resilient, yet harbors deep fragility at its core. For instance, flaws in system design (DABD) can lead to privacy breaches (AICPV), which in turn erode social legitimacy (ASLB), compounded by weak infrastructure (TTIF) and lagging regulation (RAEG), ultimately culminating in service failures (AISUF). This pattern reveals the compound vulnerability characteristic of complex adaptive systems.

These findings contribute to both academic and practical discourse on AI-related risks in finance in several key ways. First, ADIFL offers a unified framework that consolidates previously fragmented and sectoral risk perspectives into a systemic entity. Second, by linking to systemic theories such as Complex Adaptive Systems and Technological Mediation, ADIFL provides new insights into how digital failures emerge, not merely as technical glitches but as systemic phenomena. Third, the mapping of inter-subset relationships offers an evaluative tool for policy audits, risk assessments, and institutional resilience planning.

Normatively, these insights underscore the urgency of strengthening regulation, principle-based oversight, and multidisciplinary approaches to prevent structural failure in digital financial systems. Siloed or reactive responses are no longer sufficient. ADIFL Syndrome not only explains why systems fail but also guides how they can be fortified.

Furthermore, the visualisation of the ADIFL Syndrome and Fault Line Map can be operationalised as an evaluative instrument, such as an audit tool or self-assessment framework. Regulators like OJK, BI, and BSSN could use the diagram of inter-subset interactions as a template for digital risk audits, with evaluation indicators for each fault line. Each sub-set could be linked to mitigation checklists, compliance status, and cross-dimensional risk interaction potentials.

Meanwhile, financial institutions could adopt the Fault Line Map as an internal dashboard to assess the position and intensity of algorithmic risks. Quadrant positioning (e.g., Social-Macro or Techno-Micro) could help determine the type of intervention needed, whether technical, ethical, or structural. Technically, this visualisation could be integrated into data-driven audit systems, digital oversight platforms, or systemic simulations to test policy impacts on ADIFL sub-set interactions.

In this way, the visualisation not only enhances theoretical understanding but also provides a practical and technical foundation for building adaptive and anticipatory algorithmic risk governance frameworks.

### 4.8 Conceptual Visualisation: The ADIFL Framework and Fault Line Map

To deepen systemic understanding of the interrelations among the sub-sets of the ADIFL Syndrome, and how these vulnerabilities interact to shape the architecture of the "fragile fortress", this article presents two forms of conceptual visualisation:

#### 4.8.1 ADIFL Syndrome Framework Diagram

This diagram portrays ADIFL as a systemic entity, with six dynamically interconnected sub-sets. Each sub-set is arranged in a circular topology, rather than a hierarchical structure, to emphasise that there is no fixed causal order. Instead, the relationships are complex, non-linear, and characterised by feedback loops among fault lines. This structure reflects the foundational principle of Complex Adaptive Systems Theory, which posits that systemic failure rarely originates from a single point, but rather from subtle disturbances that propagate through a fragile yet interdependent network of vulnerabilities.

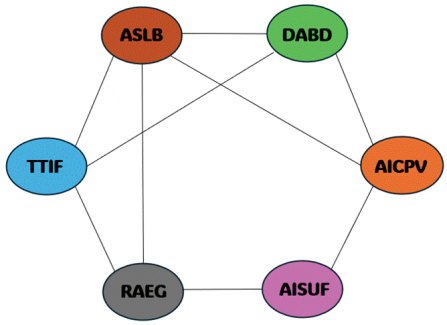


**Figure 2.** Circular Topology of ADIFL Sub-sets

#### 4.8.2 Digital Integrity Fault Line Map

This map classifies the six ADIFL sub-sets along two critical dimensions: The Techno-Social Dimension, ranging from algorithmic to socio-political aspects, and The Risk Institutionalisation Dimension, spanning from micro-level design failures to macro-level governance asymmetries.

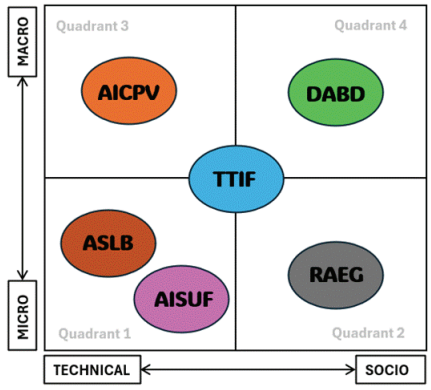


**Figure 3.** Digital Integrity Fault Line Map

Together, these visualisations serve as analytical tools for: Mapping risk mitigation priorities, based on the severity and reach of each vulnerability Illustrating the need for

multi-point interventions, showing that addressing one fault line (e.g., ASLB) without engaging others (e.g., DABD or RAEG) may prove ineffective. Through these visual instruments, the ADIFL Syndrome evolves from a conceptual theory into a diagnostic framework for assessing the structural resilience of digital financial systems.

### 4.9 Comparative Analysis with Global AI Governance Frameworks

As a theoretical construct designed to capture the systemic complexity of algorithmic risk, the ADIFL Syndrome must be tested for its relevance within the broader landscape of global AI governance. This section presents a critical comparison between ADIFL and three internationally recognised frameworks: the EU AI Act from the European Union, the FEAT Principles & Veritas Toolkit from Singapore, and the Digital Financial Inclusion Framework from the World Bank. The aim is not to replace these established models, but to position ADIFL within the global governance spectrum, highlighting its unique analytical depth and synthetic capacity as a meta-framework that complements normative, ethical, and resilience-oriented approaches.

#### 4.9.1 European Union – The EU AI Act

The EU AI Act (Floridi, 2021; European Commission, 2021 ) stands as one of the most structured regulatory frameworks globally, adopting a risk-based approach. AI systems are categorised into four tiers: unacceptable risk, high risk, limited risk, and minimal risk. Each category is governed by stringent technical and procedural requirements to ensure compliance with safety standards and the protection of fundamental rights.

The strength of this model lies in its emphasis on human rights and its capacity to provide a comprehensive compliance structure. However, from the perspective of the ADIFL framework, certain limitations emerge. The EU AI Act's compliance-centric approach does not fully capture the systemic interactions among risks, for example, how algorithmic bias (ASLB) might trigger infrastructure failures that erode public trust (TTIF). This is where ADIFL's relevance becomes evident: rather than treating risks as isolated entities, ADIFL conceptualises them as interacting fault lines that form a syndrome of systemic failure, what this article refers to as the "fragile fortress" of digital financial architecture.

#### 4.9.2 Singapore – MAS AI Governance Model (FEAT Principles & Veritas Toolkit)

Singapore adopts a more adaptive and technocratic approach to AI governance through two key initiatives: the FEAT Principles and the Veritas Toolkit, developed by the Monetary Authority of Singapore (MAS). FEAT, standing for Fairness, Ethics, Accountability, and Transparency, provides a principled foundation for ethical AI development in finance. The Veritas Toolkit offers practical instruments for financial institutions to quantitatively assess and audit their adherence to FEAT principles.

The strength of Singapore's model lies in its governance-by-design philosophy, encouraging institutions to embed ethical considerations throughout the AI lifecycle. However, when viewed through the lens of ADIFL, this approach tends to emphasise internal ethical compliance, while underplaying the systemic dynamics and inter-risk relationships that constitute a failure syndrome. For instance, the Veritas Toolkit does not explicitly address how service failures (AISUF) might be linked to algorithmic bias (ASLB) or privacy vulnerabilities (AICPV). ADIFL complements this framework by offering a systemic perspective, recognising that algorithmic failures are structurally interconnected and must be mitigated holistically, not just ethically or individually.

### 4.9.3 World Bank – Digital Financial Inclusion Framework

The World Bank's Digital Financial Inclusion Framework focuses on expanding public access to digital financial services in a safe, inclusive, and trustworthy manner. It emphasises digital identity, data privacy, cybersecurity, and regulatory strengthening as pillars of sustainable financial inclusion.

This framework's strength lies in its systemic view of digital governance, particularly in developing economies. However, within the ADIFL context, the World Bank's approach tends to treat digital risks as sectoral and normative domains, regulatory gaps, data privacy, and user trust are often addressed as separate policy areas. ADIFL's advantage lies in its ability to reframe these risks as a syndrome, where each fault line reinforces others, weakening the structural integrity of digital finance.

ADIFL also adds a conceptual layer to the World Bank's inclusion agenda by showing how technological failures (e.g., AISUF or TTIF) not only disrupt services but also erode social legitimacy (ASLB) and widen regulatory asymmetries (RAEG), ultimately undermining the very goals of digital inclusion.

As a conclusion of the comparative analysis, The ADIFL Syndrome demonstrates a distinctive synthetic strength: it is not merely a risk classification model like the EU AI Act, nor solely an ethical and technical guide like Singapore's Veritas Toolkit. Rather, it serves as a meta-framework that explains the systemic interconnections among risks and analyses how a failure in one domain (e.g., data privacy or algorithmic bias) can escalate into architectural collapse.

In the context of the World Bank's inclusion framework, ADIFL fills a critical gap by showing that digital resilience is not only about expanding access, but also about preventing internal fragility. Thus, ADIFL is not only compatible with existing global frameworks, it enhances them by offering a deeper, more integrated conceptual lens.

**Table 1.** Summary of ADIFL's Synthetic Advantages

| Aspect | EU AI Act | MAS-FEAT | World Bank | ADIFL Syndrome |
|---|---|---|---|---|
| Basis | Risk-Class Compliance | Ethical Principles | Digital Resilience | Fault-Line Systemic Theory |
| Focus | Protection of Rights | AI Accountability | Technology Resilience | Algorithmic Architectural Fragility |
| Approach | Normative-Legal | Self-Regulatory | System Capacity | Theoretical-Conceptual & Diagnostic |
| Innovation | Risk Levelling | AI Ethics | System Security | Risk Interaction as a Syndrome |

The ADIFL Syndrome offers an integrative lens that not only explains what the risks are and how they occur, but also why they interact and intensify within the systemic structure of the financial sector. As such, ADIFL serves as both a complement and a refinement to existing global approaches, especially in developing countries like Indonesia, where infrastructure and governance challenges coexist and compound.

## 5 Conclusion and Policy Recommendations

### 5.1 Conclusion

This article has proposed and formulated the Algorithmic Digital Integrity Fault Lines Syndrome (ADIFL Syndrome) as a novel theoretical framework for understanding the systemic fragility emerging from the penetration of artificial intelligence (AI) into the

digital financial sector. By integrating conceptual approaches, systematic literature review, and strategic case studies, this research constructs a theoretical synthesis that reveals how risks, often treated as isolated phenomena such as algorithmic bias, digital infrastructure failure, or regulatory asymmetry, are in fact interlinked components of a systemic risk syndrome, each triggering and amplifying the others.

This is the core contribution of the ADIFL Syndrome: it offers a meta-framework that maps algorithmic failures as fault lines forming a fragile structure within digital financial systems, a "fragile fortress" that appears robust on the surface but is vulnerable at its core.

Substantively, the framework is articulated through six primary sub-sets, AICPV, DABD, ASLB, TTIF, RAEG, and AISUF, each representing a distinct dimension of algorithmic risk with its own characteristics and interconnection dynamics. Through the integrated discussion in Section IV, this article demonstrates that these sub-sets not only manifest in the Indonesian context but also resonate with global challenges, challenges that remain only partially addressed by international frameworks such as the EU AI Act, FEAT & Veritas Toolkit, and the World Bank's digital inclusion agenda. In this regard, ADIFL offers a unique bridge between technological complexity and systemic architectural failure.

The implications of the ADIFL framework are twofold. Theoretically, it expands the literature on AI risk and digital governance by framing risk as a systemic syndrome, rather than a modular list of ethical or technical problems. Practically, it opens new pathways for regulators, policymakers, and industry actors to evaluate the readiness of digital financial systems more holistically, by accounting for the hidden interdependencies among vulnerabilities that are often obscured by the rhetoric of technological innovation. In other words, ADIFL serves as both a diagnostic and evaluative tool for assessing the architectural resilience of digital systems, beyond normative compliance checklists.

Nonetheless, this study acknowledges several limitations. First, the approach is conceptual and based on secondary literature, and thus has not yet empirically tested the ADIFL framework within specific financial institutions or populations. Second, while case studies have been used illustratively, the generalisability of findings remains limited. Third, the qualitative nature of this research has not captured the quantitative dimensions of systemic risk, such as failure probabilities or impact distributions.

Future research could focus on empirically testing the ADIFL Syndrome. This can be pursued through several avenues, such as: large-scale surveys to C-level executives and risk managers to statistically validate the causal relationships between fault lines using methods like Structural Equation Modeling (SEM); quantitative studies on historical AI failure data in the financial sector; comparative case studies between Indonesia and other ASEAN countries to test the framework's generalisability; in-depth interviews with regulators and system developers; or systemic simulations to explore inter-subset dynamics. Moreover, applying ADIFL in non-financial domains, such as e-government, digital healthcare, or AI-driven education, could broaden its relevance and cross-sectoral validity.

Beyond its analytical function, the ADIFL Syndrome holds significant potential to be developed into an evaluative instrument, such as a national digital risk index. By operationalising the six ADIFL sub-sets as indicators, regulators and financial institutions could construct a Digital Integrity Risk Index (DIRI). DIRI is envisioned not merely as a measurement tool, but as a strategic instrument for national resilience. It would function like a 'National Weather Radar' for digital risk, enabling regulators to shift from a reactive to an anticipatory stance. Furthermore, publicly available DIRI scores could create 'positive

pressure' on institutions to proactively improve, and the data collected would serve as a rich foundation for OJK's future Supervisory Technology (SupTech) capabilities. This index could be used for risk rating, policy auditing, and data-driven mitigation planning.

In conclusion, the ADIFL Syndrome represents an initial step toward a new paradigm for understanding and responding to AI risks in a systemic and multidimensional manner. By placing algorithmic architectural fragility at the center of analysis, this article invites both academic and policy communities to move beyond sectoral approaches, toward a holistic and anticipatory understanding of digital resilience that is not only adaptive, but also self-aware of its own failure anatomy.

### 5.2 Policy Recommendations

In response to the conceptual findings outlined in the Algorithmic Digital Integrity Fault Lines Syndrome (ADIFL Syndrome) framework, this section presents a set of policy recommendations that are concrete, context-sensitive, and designed for phased implementation within Indonesia's financial services sector. These recommendations focus on mitigating the six ADIFL sub-sets, each identified as a critical architectural fault line in the digital ecosystem, with the overarching goal of building comprehensive systemic resilience.

Each recommendation articulates the substantive policy objective followed by detailed implementation strategies, specifying responsible actors and operational mechanisms. Given the cross-sectoral and complex nature of algorithmic risks, all recommendations emphasise the need for coordinated action among the Financial Services Authority (OJK), Bank Indonesia (BI), the Ministry of Communication and Informatics (Kominfo), the National Cyber and Crypto Agency (BSSN), and AI-driven financial industry stakeholders.

#### 5.2.1 AICPV (AI-Induced Cyber-Privacy Vulnerability)

To address the AI-Induced Cyber-Privacy Vulnerability (AICPV), it is imperative to mandate AI-based privacy audits and cybersecurity standards specifically tailored to AI systems, particularly those utilized in high-stakes functions such as e-KYC, credit scoring, and customer service chatbots. The implementation of this policy requires the Financial Services Authority (OJK) and the National Cyber and Crypto Agency (BSSN) to collaboratively develop AI privacy audit modules based on the ISO/IEC 27701 standard, explicitly incorporating algorithmic risk indicators. At the operational level, industry actors must be strictly required to submit a Privacy Impact Assessment (PIA) for every AI system prior to its public deployment. Furthermore, to ensure pre-market safety, all high-risk AI systems must undergo mandatory sandbox testing as a prerequisite for obtaining operational approval.

To operationalize this, high-risk classification criteria shall be formally defined, covering data sensitivity, scale of impact, algorithmic opacity, and potential bias, while cybersecurity controls shall be aligned with ISO/IEC 27001/27002 and AI risk management guidance (e.g., ISO/IEC 23894, NIST AI RMF). Finally, sandbox governance, enforcement mechanisms, and reporting requirements shall be jointly stipulated by OJK and BSSN to ensure consistent compliance.

#### 5.2.2 DABD (Data-Abuse by Design)

To mitigate the risks of Data-Abuse by Design (DABD), regulators must enforce legally binding Data Ethics by Design principles intended to prevent exploitative data practices embedded within AI system architectures. These principles must explicitly encompass

fairness, accountability, explainability, and privacy by design, while being harmonized with Indonesia's Personal Data Protection Law (UU PDP) and global standards such as the OECD AI Principles and ISO/IEC 23894. This strategy requires the Ministry of Communication and Informatics (Kominfo) and the OJK to co-develop legal guidelines for ethical AI design that ensure fairness and transparency while strictly limiting exploitative data use.

To ensure that compliance goes beyond mere declaration, the framework must incorporate robust oversight mechanisms and graduated sanctions for violations. Furthermore, companies must be mandated to publish a Data Use Transparency Sheet detailing specifically how user data is collected, processed, and monetised. To guarantee utility and accessibility, this disclosure document should be formatted to be machine-readable, made publicly available on official corporate websites, and updated periodically to reflect evolving data practices. These regulations are to be formally integrated into revisions of data protection laws as well as fintech licensing frameworks.

### 5.2.3 ASLB (Algorithmic Social Legitimacy Breakdown)

To prevent an Algorithmic Social Legitimacy Breakdown (ASLB), it is critical to establish robust grievance and redress mechanisms for individuals affected by algorithmic bias within the financial services sector. This initiative necessitates that the OJK form a specialized Algorithmic Dispute Resolution Unit in collaboration with the Alternative Dispute Resolution Institution for the Financial Services Sector (LAPS SJK). To ensure operational effectiveness, this unit must function under clear Service Level Agreements (SLAs) regarding resolution timelines while explicitly guaranteeing the right to appeal for consumers.

At the institutional level, financial service providers must be mandated to implement Explainable AI (XAI) features intended to allow users to review and understand the rationale behind automated decisions using plain and accessible language. Furthermore, institutions are required to publish annual algorithmic transparency reports that include detailed metrics on bias, data representation, and model interpretability. These disclosures must be subject to mandatory independent fairness audits aligned with global standards such as ISO/IEC TR 24028 or the NIST AI Risk Management Framework. Finally, strictly defined enforcement mechanisms and graduated sanctions must be applied to ensure adherence to these fairness protocols.

### 5.2.4 TTIF (Trust-Through-Infrastructure Failure)

To mitigate Trust-Through-Infrastructure Failure (TTIF), regulators must introduce mandatory periodic resilience testing for AI systems embedded within critical digital financial infrastructure. Implementation requires Bank Indonesia (BI), the OJK, and BSSN to jointly develop AI Infrastructure Stress Test Protocols covering core banking systems, AI cloud platforms, payment gateways, and critical third-party integrations. These protocols should be aligned with international standards such as ISO 22301 on business continuity and the principles of the Digital Operational Resilience Act (DORA) to ensure global interoperability and regulatory harmonization.

Crucially, semi-annual audits must be conducted to assess reliability, redundancy, Recovery Time Objectives (RTO), and potential social impact in the event of disruption. Given the reliance on external vendors, these audits must extend to Third-Party Risk Management (TPRM) to address concentration risks in cloud computing. Non-compliance with resilience testing or audit requirements shall trigger regulatory sanctions and

mandatory remediation within defined timelines.

Finally, to balance transparency with security, test results are to be disclosed in a strictly limited-access format restricted to regulators and authorized stakeholders as part of systemic risk monitoring. Reports must include standardized metrics on reliability, redundancy, RTO, and social impact to ensure accountability and comparability across institutions.

### 5.2.5 RAEG (Regulatory Asymmetry in Emerging Governance)

To address Regulatory Asymmetry in Emerging Governance (RAEG), the government must establish a cross-sectoral and adaptive task force equipped with a horizon-scanning mandate incorporating scenario planning to formulate responsive AI policies. This National AI Financial Task Force shall comprise representatives from the OJK, Bank Indonesia (BI), Kominfo, the Ministry of Finance, and academic experts, alongside industry practitioners, to ensure comprehensive regulatory foresight. To ensure global and local relevance, policy recommendations should be informed by a Regulatory Technology Watch while being aligned with evolving legal frameworks in the European Union and ASEAN, as well as harmonized with national regulations such as the Personal Data Protection Law (UU PDP). Finally, a green paper mechanism and broad public consultations must be implemented to ensure that these regulatory developments maintain social legitimacy and public accountability.

### 5.2.6 AISUF (AI Service Unreliability and System Failure)

To address AI Service Unreliability and System Failure (AISUF), regulators must establish rigorous standards for AI system reliability audits and mandatory reporting of systemic incidents caused by AI failures. Implementation requires the Financial Services Authority (OJK) to mandate the submission of an AI Reliability Scorecard for all institutions using AI in critical operations, which includes customer service and risk decision-making. Simultaneously, Bank Indonesia (BI) must require a dedicated Business Continuity Plan (BCP) specifically addressing AI-related service failures in payment systems and digital banking to ensure the availability of manual fallback mechanisms during outages.

To ensure global interoperability and resilience, these reliability audits and BCP frameworks should be explicitly aligned with ISO/IEC TR 24028 regarding AI trustworthiness and ISO 22301 on business continuity management. Finally, to ensure a systematic and effective rollout, the implementation of these measures shall commence with the top 10 banks and fintech firms ranked by transaction volume.

These policy recommendations affirm that mitigating the ADIFL Syndrome requires a holistic, adaptive, and interconnectivity-aware approach. It is no longer sufficient for financial institutions to merely adopt ethical principles or comply with sectoral regulations. What is needed is a policy architecture that recognises and addresses algorithmic risk as a systemic syndrome with mutually reinforcing fault lines.

The ADIFL framework provides a conceptual foundation for future policy design, one that is responsive to digital complexity while grounded in Indonesia's contextual realities.

## Disclosure

### Conflict of Interest

The author declares that there are no conflicts of interest that could have influenced the results or interpretation of this research. The analysis and policy recommendations

presented are independent and do not reflect the official stance of any specific financial institution.

### Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The study was conducted as an independent academic contribution to the field of digital financial governance.

### Author Contributions

Yoseph Hendrik Maturbongs (Y.H.M.): Conceptualisation; Methodology; Formal Analysis; Investigation; Writing original draft preparation; Writing review and editing; Visualisation. The author confirms sole responsibility for the study conception and design.

### Author Acknowledgement and Use of AI Tools

The author extends gratitude to the organisers and anonymous reviewers of the Karisma OJK 2025 competition for their constructive feedback which significantly improved the manuscript.

Disclosure of Generative AI Use: The author acknowledges the use of Google Gemini (Generative AI) during the preparation of this work. The tool was used exclusively for:

1. Linguistic refinement: Improving the clarity, flow, and grammatical accuracy of the English text.
2. Structural brainstorming: Assisting in the organisation of the theoretical arguments (ADIFL Syndrome framework) and refining the presentation of methodological phases. The author assumes full responsibility for the final content, data interpretation, and conclusions drawn in this article. No AI tool was used to generate underlying data or fabricate references.

### Data Availability

Data sharing is not applicable to this article as no new datasets were generated or analysed during the current study. The analysis is based on publicly available secondary data, including scholarly articles, regulatory reports (OJK, BI, FSB), and media reports, all of which are cited in the References section.

## References

Agudo, U., Liberal, K. G., Arrese, M., & Matute, H. (2024). The impact of AI errors in a human-in-the-loop process. *Cognitive Research: Principles and Implications*, *9*(1), 1. https://doi.org/10.1186/s41235-023-00529-3

Akerlof, G. A. (1978). The market for "lemons": Quality uncertainty and the market mechanism. In *Uncertainty in economics* (pp. 235–251). Elsevier. https://www.sciencedirect.com/science/article/pii/B9780122148507500022X

Alfrink, K., Keller, I., Kortuem, G., & Doorn, N. (2022). Contestable AI by Design: Towards a framework. *Minds and Machines*, *33*(4), 613–639. https://doi.org/10.1007/s11023-022-09611-z

Alrasheed, G., & Lim, M. (2021, May 16). *Beyond a technical bug: Biased algorithms and moderation are censoring activists on social media*. The Conversation. http://theconversation.com/beyond-a-technical-bug-biased-algorithms-and-moderation-are-censoring-activists-on-social-media-160669

Bahoo, S., Cucculelli, M., Goga, X., & Mondolo, J. (2024). Artificial intelligence in Finance: A comprehensive review through bibliometric and content analysis. *SN Business & Economics*, *4*(2), 23. https://doi.org/10.1007/s43546-023-00618-x

Bank for International Settlements, (BIS). (2023). *A step toward new financial market infrastructure:*

*Infrastructure: Bank of Korea's initiative Bank of Korea's initiative* (p. 25). Bank of Korea.
Basak, A., & Tiwari, D. (2025). *API security risk and resilience in financial institutions*. https://www.theseus.fi/handle/10024/883344
Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics*, *5*(3), 3265–3279. https://doi.org/10.1007/s43681-024-00653-w
Bayamlıoğlu, E. (2022). The right to contest automated decisions under the general data protection regulation: Beyond the so-called "right to explanation." *Regulation & Governance*, *16*(4), 1058–1078. https://doi.org/10.1111/rego.12391
Boivin, J. (2025). The ecological catastrophe of algorithmic individuation: Technological mediation and its social implication in the age of the anthropocene. In D. Binns & R. Najdowski (Eds.), *Confronting the Climate Crisis: Activism, Technology and Ecoaesthetics* (pp. 293–311). https://doi.org/10.1007/978-3-031-89606-4_15
Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Conference on Fairness, Accountability and Transparency*, 77–91.
Cahyani, F. D., & Maria, N. S. B. (2023). Analisis stabilitas sistem keuangan dan faktor-faktor yang mempengaruhinya. *Diponegoro Journal of Economics*, *12*(3), 57–61.
Cate, M. (2025). *Data privacy risks and vulnerabilities in API-driven financial ecosystems*.
Chang, X. (2023). Gender bias in hiring: An analysis of the impact of amazon's recruiting algorithm. *Advances in Economics Management and Political Sciences*, *23*(1), 134–140.
Cristine, M. A., Risakota, F. M. A., & Sirait, S. R. C. (2025). Perlindungan data pribadi dalam sistem skoring kredit otomatis oleh fintech di indonesia: Analisis yuridis normatif berdasarkan undang- undang nomor 27 tahun 2022. *Jurnal Studi Hukum Modern*, *7*(3), Article 3. https://journalversa.com/s/index.php/jshm/article/view/860
Cypriva. (2025, March 8). *Kebocoran data pribadi di indonesia: ancaman, kasus, dan sol*. https://www.cypriva.id/articles/kebocoran-data-pribadi-di-indonesia-ancaman-nyata-kasus-terkini-dan-solusi-pencegahannya-
Dellaert, B. G. C., Baker, T., & Johnson, E. J. (2024). Regulating robo-advice for consumers' financial decisions: The interplay between algorithm quality & digital choice architecture. *Behavioral Science & Policy*, *10*(2), 1–7. https://doi.org/10.1177/23794607241296686
Detiknet, T. (2021, July 28). *Kominfo investigasi dugaan kebocoran data 2 juta pengguna BRI Life* [News]. detikinet. https://inet.detik.com/security/d-5659592/kominfo-investigasi-dugaan-kebocoran-data-2-juta-pengguna-bri-life
Dorochowicz, A., Jankowski, D., Ksieniewicz, P., Topolska, K., Topolski, M., & Zyblewski, P. (2025). A prototype of an ai-driven operational security system for fintechs: A faas-based approach to fraud detection, *Progress in Pattern Classification and Machine Learning* (pp. 103–112). https://doi.org/10.1007/978-3-032-01773-4_11
Eschenbach, W. J. V. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, *34*(4), 1607–1622. https://doi.org/10.1007/s13347-021-00477-0
EU. (2016, April 27). *Regulation, 2016/679, EN - gdpr, EUR-Lex* [Official Journal of the European Union]. European Union. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
Fitriyanti, F., Devty, S., Putri, S., & Thora, R. E. (2024). Securing personal data in e-kyc: vital for digital economy growth. *Diponegoro Law Review*, *9*(1), 104–120.
Fredrikson, M., Jha, S., & Ristenpart, T. (2015). Model inversion attacks that exploit confidence information and basic countermeasures. *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 1322–1333. https://doi.org/10.1145/2810103.2813677
FSB. (2024). *The Financial stability implications of artificial intelligence*.
Fursov, I., Morozov, M., Kaploukhaya, N., Kovtun, E., Rivera-Castro, R., Gusev, G., Babaev, D., Kireev, I., Zaytsev, A., & Burnaev, E. (2021). Adversarial attacks on deep models for financial transaction records. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge*

*Discovery & Data Mining*, 2868–2878. https://doi.org/10.1145/3447548.3467145
Grennan, J. (2022). FinTech regulation in the united states: Past, present, and future. *Present, and Future (August 31, 2022)*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4045057
Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access*, *11*, 80218–80245.
Hidayanti, N. F., Ariani, Z., & Syaharuddin, S. (2025). The integration of artificial intelligence in islamic financial services: A review on digital innovation for sharia financial inclusion. *Integrating Religion, Social Economy, and Law: Conference Series*, *1*(1), 8–17. https://journal.ummat.ac.id/index.php/ics/article/view/32559
Holland, J. H. (1992). Complex adaptive systems. *Daedalus*, *121*(1), 17–30.
Hughes, H. P. N., Clegg, C. W., Bolton, L. E., & Machon, L. C. (2017). Systems scenarios: A tool for facilitating the socio-technical design of work systems. *Ergonomics*, *60*(10), 1319–1335. https://doi.org/10.1080/00140139.2017.1288272
Iqbar, M. D., Darrel, M., & Akilah, U. I. (2024). *Analisis politik luar negeri indonesia dinamika dan tantangan lima tahun terakhir*. https://doi.org/10.13140/RG.2.2.34325.95203
Iyer, S. (2022). *Operational Resilience for Financial Institutions* [PhD Thesis, Trident University International].
Insurtech Asia. (2023). *InsureTech Connect Asia 2023 Summary Report*. Abeam Consulting. https://www.abeam.com/id/en/insights/insure_tech_2023/
Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, *10*(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0
James, U. U., Idika, C. N., Enyejo, L. A., Abiodun, K., & Enyejo, J. O. (2024). Adversarial attack detection using explainable AI and generative models in real-time financial fraud monitoring systems. *International Journal of Scientific Research and Modern Technology*, *3*(12), 142–157.
Javia, S. (2025, March 27). *Kesenjangan besar dalam AI: pemimpin di sektor jasa keuangan hadapi tantangan tata kelola data dan tuntutan infrastruktur*. IndonesiaSatu.co. https://indonesiasatu.co/detail/kesenjangan-besar-dalam-ai--pemimpin-di-sektor-jasa-keuangan-hadapi-tantangan-tata-kelola-data-dan-tuntutan-infrastruktur
Kandpal, V., Ozili, P. K., Jeyanthi, P. M., Ranjan, D., & Chandra, D. (2025). Regulation of the fintech industry. in *digital finance and metaverse in banking: decoding a virtual reality towards financial inclusion and sustainable development* (pp. 181–198). *Emerald Publishing Limited*. https://www.emerald.com/insight/content/doi/10.1108/978-1-83662-088-420251009/full/html
Karakasilioti, G. M. (2024). *Supporting the digital operational resilience of the financial sector: The EU's DORA Digital Operational Resilience Act* [Master's Thesis,
Lachmann, N. (2025). Chasing the elusive bird?: The technological development of the digital economy and international trade law's susceptibility to a pacing problem. *The Journal of World Investment & Trade*, *26*(3), 479–511.
Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces*, *63*(4), 967–985.
Li, Q. (2025). Research on consumer data sovereignty and algorithmic fairness. *International Journal of Management Science Research*, *8*(4), 6–10.
Lim, T. (2024). Environmental, social, and governance (ESG) and artificial intelligence in finance: State-of-the-art and research takeaways. *Artifcial Intelligence Review*, *57*(76), 1–64. https://doi.org/10.1007/s10462-024-10708-3
Lu, S. (2022). Data privacy, human rights, and algorithmic opacity. *Cal. L. Rev.*, *110*, 2087.
Luhmann, N. (1979). *Trust and Power*. Chichester: Jhon Wiley and Son Inc.
Luthfah, D. (2024). Penguatan keamanan siber pada sektor jasa keuangan Indonesia. *Jurnal Penelitian dan Karya Ilmiah Lembaga Penelitian Universitas Trisakti*, *9*(1), 259–267. https://doi.org/10.25105/pdk.v9i1.18643
MacInnis, D. J. (2011). A framework for conceptual contributions in marketing. *Journal of Marketing*, *75*(4), 136–154. https://doi.org/10.1509/jmkg.75.4.136
Malatji, M., & Tolah, A. (2025). Artificial intelligence (AI) cybersecurity dimensions: A

comprehensive framework for understanding adversarial and offensive AI. *AI and Ethics*, *5*(2), 883–910. https://doi.org/10.1007/s43681-024-00427-4

MAS. (2018). *Principles to promote fairness, ethics, accountability and transparency (feat) in the use of artificial intelligence and data analytics in singapore's financial sector* (p. 15). Monetary Authority of Singapore.

MAS. (2024). *Artificial intelligence model risk management: observations from a thematic review* (Artificial Intellegent Model Risk Management, pp. 1–50) [Information Paper]. Monetary Authority of Singapore.

Minkkinen, M., Laine, J., & Mäntymäki, M. (2022). Continuous auditing of artificial intelligence: a conceptualization and assessment of tools and frameworks. *Digital Society*, *1*(3), 21. https://doi.org/10.1007/s44206-022-00022-2

Neumannová, A., Bernroider, E. W. N., & Elshuber, C. (2023). The digital operational resilience act for financial services: a comparative gap analysis and literature review, *Information Systems* (Vol. 464, pp. 570–585). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-30694-5_40

Nguyen, V.-L., Lin, P.-C., Cheng, B.-C., Hwang, R.-H., & Lin, Y.-D. (2021). Security and privacy for 6G: A survey on prospective technologies and challenges. *IEEE Communications Surveys & Tutorials*, *23*(4), 2384–2428.

Novelli, C., Hacker, P., McDougall, S., Morley, J., Rotolo, A., & Floridi, L. (2025). Getting regulatory sandboxes right: Design and governance under the AI act. *Available at SSRN 5332161*.

OJK. (2025). *Tata kelola kecerdasan artifisial perbankan Indonesia*. Otoritas Jasa Keuangan (OJK).

Panarese, P., Grasso, M. M., & Solinas, C. (2025). Algorithmic bias, fairness, and inclusivity: A multilevel framework for justice-oriented AI. *AI & SOCIETY*. https://doi.org/10.1007/s00146-025-02451-2

Parisot, M. P. M., Pejo, B., & Spagnuelo, D. (2021). *Property inference attacks on convolutional neural networks: Influence and implications of target model's complexity* (No. arXiv:2104.13061). arXiv. https://doi.org/10.48550/arXiv.2104.13061

Power, M. (2022). Theorizing the economy of traces: from audit society to surveillance capitalism. *Organization Theory*, *3*(3), 26317877211052296. https://doi.org/10.1177/26317877211052296

Pradana, A. E., Herawati, A. R., & Dwimawanti, I. H. (2025). Tantangan kecerdasan buatan dalam implikasi kebijakan pemerintah di indonesia: Studi Literatur. *Jurnal Good Governance*, 51–66.

Puannandini, D. A., Fabian, R., Firdaus, R. A. P., Mustopa, M. Z., & Herdiyana, I. (2025). Liabilitas produk AI dalam sistem hukum Indonesia: Implikasi bagi pengembang, pengguna, dan penyedia layanan. *Iuris Studia: Jurnal Kajian Hukum*, *6*(1), 24–33.

Purba, D. S., Permatasari, P. D., Tanjung, N., Rahayu, P., Fitriani, R., & Wulandari, S. (2025). Analisis perkembangan ekonomi digital dalam meningkatkan pertumbuhan ekonomi di indonesia. *Jurnal Masharif Al-Syariah: Jurnal Ekonomi dan Perbankan Syariah*, *10*(1). https://journal.um-surabaya.ac.id/Mas/article/view/25367

Rafael, E. F. (2013). Technology as a social system: A systems theoretical conceptualization. *Philippine Sociological Review*, 319–347.

Rakovic, I. (2022). *Dark finance: Exploring deceptive design in investment apps* [Master's Thesis, NTNU]. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3005227

Rakovic, I., & Inal, Y. (2023). Dark finance: Exploring deceptive design in investment apps, *Human-Computer Interaction–Interact 2023*. 14142 (pp. 339–348). https://doi.org/10.1007/978-3-031-42280-5_20

Ravshan, K. (2025). Risks and regulatory challenges associated with fintech technologies. *International Conference on Interdisciplinary Science*, *2*(6), 282–284.

Respati, A. R., & Sukmana, Y. (2023, December 7). *Evaluasi penyaluran KUR, dari debitur tak punya NPWP sampai Biaya "siluman"* [News]. KOMPAS.com. https://money.kompas.com/read/2023/12/07/194311126/evaluasi-penyaluran-kur-dari-debitur-tak-punya-npwp-sampai-biaya-siluman

Romeo, G., & Conti, D. (2025). Exploring automation bias in human–AI collaboration: A review and implications for explainable AI. *AI & SOCIETY*. https://doi.org/10.1007/s00146-025-02422-7

Ryan, M., Withers, G., & Den Hartog, F. (2024). The cloud conundrum: Are financial institutions heading for a catastrophic disruption event? *IEEE Transactions on Technology and Society*. https://ieeexplore.ieee.org/abstract/document/10706249/

Sandy, K. F. (2025, February 9). *Begini strategi OJK hadapi serangan siber perbankan* [IDX Channel]. https://www.idxchannel.com/. https://www.idxchannel.com/banking/begini-strategi-ojk-hadapi-serangan-siber-perbankan

Sankalp, M. R., Lokapal, G., Mohan, B. A., & Basavaraj, G. N. (2025). Addressing cybersecurity challenges in 6g networks through ai-driven adaptive defense mechanisms and quantum-resilient protocols. *2025 International Conference on Computing for Sustainability and Intelligent Future (COMP-SIF)*, 1–12. https://ieeexplore.ieee.org/abstract/document/10969886/

Setyowati, D. (2025, April 10). *Kronologi sistem Bank DKI eror, data dan dana nasabah aman? - Teknologi Katadata.co.id*. https://katadata.co.id/digital/teknologi/67f74c66822b5/kronologi-sistem-bank-dki-eror-data-dan-dana-nasabah-aman

Shahnaz, K. (2021, Agustus). *BTPN soal pembobolan nasabah jenius: Itu ulah social engineering* [Financial News]. Bisnis.Com. https://finansial.bisnis.com/read/20210826/90/1434378/btpn-soal-pembobolan-nasabah-jenius-itu-ulah-social-engineering

Singh, S., Rahman, A., & Kaur Johl, S. (2025). The looming labyrinth: Risks of artificial intelligence in financial sector. In *(Eds) Green Horizons: Role of AI in Sustainable Finance* (pp. 215–235). Springer, Singapore. https://doi.org/10.1007/978-981-96-6495-5_12

Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American Economic Review*, *71*(3), 393–410.

Sun, J., Zhao, M., & Lei, C. (2024). Class-imbalanced dynamic financial distress prediction based on random forest from the perspective of concept drift. *Risk Management*, *26*(4), 19. https://doi.org/10.1057/s41283-024-00150-8

Susilo, A. (2023). regulatory technology untuk digitalisasi proses kepatuhan (Studi kasus bank swasta di Indonesia). *Infotech Journal*, *9*(1), 252–258. https://doi.org/10.31949/infotech.v9i1.5460

Tempo.com. (2025, February 12). *Sempat error berhari-hari, sistem IT pada aplikasi byond dari bsi kini diklaim sudah stabil | tempo.co*. Tempo. https://www.tempo.co/ekonomi/sempat-error-berhari-hari-sistem-it-pada-aplikasi-byond-dari-bsi-kini-diklaim-sudah-stabil-1205998

Thalpage, N. (2023). Unlocking the black box: Explainable artificial intelligence (XAI) for trust and transparency in ai systems. *J. Digit. Art Humanit*, *4*(1), 31–36.

Trinh, T. K., & Zhang, D. (2024). Algorithmic fairness in financial decision-making: Detection and mitigation of bias in credit scoring applications. *Journal of Advanced Computing Systems*, *4*(2), 36–49.

Trisnawati. (2024). Artificial intelligence governance and regulation: A roadmap to developing legal policies for artificial intelligence deployment. *Journal of Governance and Administrative Reform (JGAR)*, *5*(2), 185–194.

Trist, E. L. (1981). *The evolution of socio-technical systems* (Vol. 2). Ontario quality of working life Centre Toronto. https://www.lmmiller.com/blog/wp-content/uploads/2013/06/The-Evolution-of-Socio-Technical-Systems-Trist.pdf

Veale, M., Binns, R., & Edwards, L. (2018). Algorithms that remember: Model inversion attacks and data protection law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180083. https://doi.org/10.1098/rsta.2018.0083

Verbeek, P.-P. (2023). Postphenomenology and ethics. In *Technology Ethics* (pp. 42–51). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003189466-8/postphenomenology-ethics-peter-paul-verbeek

Walter, Y. (2024). Managing the race to the moon: Global policy and governance in Artificial Intelligence regulation, A contemporary overview and an analysis of socioeconomic consequences.

*Discover Artificial Intelligence*, *4*(1). https://doi.org/10.1007/s44163-024-00109-4
Yaksan, S. (2024). Regulating AI with a legal policy fuelled by innovation and accountability: Regulatory sandboxes as the way forward. *Available at SSRN 5068132*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5068132
Zhang, X., Wang, T., Ma, L., & Mahadevan, S. (2025). Reliability engineering, risk management, and trustworthiness assurance for AI systems. *Journal of Reliability Science and Engineering*, *1*(2), 022001.
Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. *Applications of Social Research Methods to Questions in Information and Library Science*, *308*(319), 1–12.
Zhang, Z., Ning, H., Shi, F., Farha, F., Xu, Y., Xu, J., Zhang, F., & Choo, K.-K. R. (2022). Artificial intelligence in cyber security: Research advances, challenges, and opportunities. *Artificial Intelligence Review*, *55*(2), 1029–1053. https://doi.org/10.1007/s10462-021-09976-0
Zuboff, S. (2023). The age of surveillance capitalism. In *Social theory re-wired* (pp. 203–213). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003320609-27/age-surveillance-capitalism-shoshana-zuboff

This page is intentionally left blank