Deepfakes in Onboarding, KYC, and Financial Fraud: Authenticity Standards and Liability Framework for Digital Banks and FinTech

Juan Emmanuel Delva Benavides*, Jorge Antonio Leos Navarro and Alejandro Paul García Hernández

Centro Universitario de Ciencias Económico Administrativas (CUCEA), Universidad de Guadalajara, Mexico

Abstract: Deepfakes and synthetic media have evolved from reputational threats to direct financial exploitation tools, enabling sophisticated impersonation during remote onboarding, evasion of biometric verification systems, and the creation of synthetic identity accounts at industrial scale. Yet most countermeasures remain fragmented, relying on ad hoc vendor controls or intrusive surveillance mechanisms that undermine both user trust and financial inclusion objectives. This article advances an integrated legal-technical framework for FinTech institutions and digital banks structured around four core contributions: (i) a FinTech-specific taxonomy of deepfake-enabled fraud vectors spanning onboarding, KYC refresh, and account takeover scenarios; (ii) a normative mapping of duties grounded in risk-based customer due diligence and security obligations, anchored in Mexican law but internationally interoperable; (iii) a Tiered Authenticity and Traceability Standard (TATS) that calibrates verification intensity with transaction risk while enforcing data minimization and auditability principles; and (iv) a pragmatic liability allocation model distributing responsibility among deployers (FinTech), users, and verification vendors based on control capacity, foreseeability, and evidentiary capability. By integrating digital identity assurance standards with transparency-by-design principles and secure capture/provenance mechanisms, TATS operationalizes a privacy-preserving trust infrastructure that supports safer digital finance and, ultimately, sustainable financial inclusion in emerging markets.

Keywords: Deepfake-enabled fraud, KYC, Remote onboarding, Biometric verification, Liveness detection, Liability, Financial fraud, Synthetic identity, Digital identity assurance, FinTech regulation.

1. INTRODUCTION: THE AUTHENTICITY CRISIS IN DIGITAL FINANCE

The global financial industry stands at a historical inflection point, characterized by the collision of two tectonic forces: the massive democratization of access to financial services through technology (FinTech) and the emergence of accessible, low-cost Generative Artificial Intelligence (GenAI) capabilities. This convergence has precipitated an unprecedented authenticity crisis wherein the foundational premise of digital banking—the capacity to verify the identity of a remote counterparty—has been structurally compromised. The proliferation of deepfakes (hyper-realistic synthetic media) and the industrialization of synthetic identity fraud represent not merely operational threats but an existential risk to the sustainability of the FinTech business model and, by extension, to financial inclusion in emerging economies (Bateman, 2020; Chesney & Citron, 2019).

The current context is alarming. According to data from Lucinity (2025), between May 2024 and April 2025, AI-enabled fraud schemes increased by an astonishing 456% in key sectors such as banking and insurance. This figure does not reflect a linear increase in criminality but rather a paradigm shift in the nature of attacks: the transition from credential theft (access attacks) to reality fabrication (truth attacks). Malicious actors no longer need to steal an existing user's identity; they now possess tools to construct complex synthetic identities composed of fragments of real data and artificially generated biometric attributes capable of bypassing traditional Know Your Customer (KYC) filters and cultivating legitimate credit histories before executing massive fraud schemes (TransUnion, 2025; Javelin Strategy & Research, 2024). Furthermore, the Deloitte Center for Financial Services (2024) projects that generative AI could drive U.S. fraud losses from $12.3 billion in 2023 to $40 billion by 2027—a compound annual growth rate of 32%.

This phenomenon carries profound implications for financial sustainability. Sustainability extends beyond ESG investments to encompass the long-term economic viability of institutions serving vulnerable populations. If fraud losses continue to escalate, risk costs will inevitably transfer to end users through higher interest rates and stricter entry barriers, reversing decades of progress in financial inclusion (FATF, 2020). In the Mexican legal context, the situation is particularly precarious for financial institutions. The Supreme Court of Justice of the Nation (SCJN) has consolidated a garantista jurisprudential line that, in cases of unrecognized charges or disputed electronic transfers, reverses the burden of proof onto the banking institution (Tesis 1a./J. 17/2021). The judicial premise holds that the bank, as custodian of technology and resources, is better positioned to prove the reliability of the transaction. However, in an environment where a deepfake can replicate a client's

*Address correspondence to this author at the Centro Universitario de Ciencias Económico Administrativas (CUCEA), Universidad de Guadalajara, Mexico; E-mail: emmanueldelva@gmail.com

Received on 27-12-2025 Accepted on 10-02-2026 Published on 19-02-2026

E-ISSN: 2755-8398/26

voice and face with hyperrealistic fidelity, and where video injections can evade camera sensors, the "proof of reliability" becomes a technically complex task (FS-ISAC, 2024).

The landmark case illustrating this threat occurred in January 2024, when an employee at Arup, a British multinational engineering firm, transferred $25.6 million to fraudsters after participating in a video conference where the company's CFO and several colleagues appeared—all of whom were deepfake recreations (Chen & Magramo, 2024; Fortune, 2024). This incident demonstrates that deepfake technology is no longer theoretical nor exclusive to state actors; it has become an operational tool for organized financial crime.

The central research question guiding this investigation is: What authenticity standards and liability allocation scheme (platform/user/technology provider) best reduces deepfake-enabled fraud in onboarding and KYC processes without defaulting to disproportionate surveillance? This question seeks to resolve the tension between technological innovation, legal certainty, and consumer protection while ensuring proportionality and preserving financial inclusion objectives (UNESCO, 2025).

The key contributions of this article are: (1) a practical, FinTech-specific threat taxonomy integrating deepfake-enabled fraud and synthetic identity fraud vectors across the customer lifecycle; (2) a duty map aligning AML/CFT customer due diligence with security and audit obligations, anchored in Mexican law but internationally interoperable; (3) a Tiered Authenticity and Traceability Standard (TATS) establishing controls, evidence requirements, and transaction limits calibrated by risk tier; and (4) a liability allocation model distributing responsibility to the actor with the greatest capacity to prevent harm and preserve evidence.

2. METHODOLOGY

To address the multidimensional complexity of this problem, this research employs a comparative legal-technological analysis methodology, integrating legal dogmatics with security systems engineering. The approach is qualitative, hermeneutic, and propositional, designed to generate actionable insights for both legal directors and information security architects. The methodological design unfolds in three sequential phases.

2.1. Jurisprudential and Normative Analysis Phase

A deep exegesis of primary sources of financial and civil law in Mexico was conducted, including: (a) jurisprudential analysis reviewing isolated theses and binding precedents from the SCJN's Tenth and Eleventh Epochs, focusing on criteria for transfer nullity, burden of proof in commercial matters, and interpretation of "reliability" in electronic banking; and (b) regulatory review examining the Financial Technology Institutions Regulation Law (Ley Fintech, Cámara de Diputados, 2025), General Provisions applicable to Electronic Payment Fund Institutions (IFPE), and the Single Banking Circular (CUB), specifically annexes related to non-face-to-face identification and biometric factor usage (CNBV, 2024).

2.2. Threat and Technology Evaluation Phase

The technical anatomy of emerging threats and available countermeasures was systematically analyzed based on industry reports and global technical standards. This included: (a) fraud taxonomy classifying attack vectors (synthetic identity, video deepfakes, voice cloning, data injection) using 2024-2025 data from sources including Veriff (2025), Entrust (2025), Signicat (2025), and Sumsub (2024); and (b) authentication standards providing technical analysis of ISO/IEC 30107-3:2023 on Presentation Attack Detection (PAD), NIST SP 800-63B-4 (2025) guidelines on Authentication Assurance Levels (AAL), and FIDO Alliance (2023) protocols for passwordless authentication.

2.3. Synthesis and Integration Phase (LegalTech)

The methodology culminates in the integration of both domains, correlating technical certification levels (*e.g.*, PAD Level 2) with degrees of diligence required by law, constructing a logical argument on how certified technology can satisfy courts' evidentiary requirements (iBeta, 2024). This synthesis enables the formulation of de lege ferenda proposals and concrete operational recommendations aligned with the FATF Guidance on Digital Identity (2020).

3. THREAT MODEL: WHERE DEEPFAKES IMPACT DIGITAL FINANCE

Deepfake-enabled fraud is not a single technique but rather an attack stack that exploits multiple vulnerabilities across the customer lifecycle (Carpenter, 2025; FS-ISAC, 2024). According to Signicat's research (2025), deepfake fraud attempts have increased by 2,137% over the past three years, now representing approximately 6.5% of all detected fraud cases in the financial sector.

3.1. Synthetic Identity Engineering

Synthetic identity fraud represents a sophisticated evolution of financial crime that the U.S. Department of Justice has characterized as the country's

fastest-growing financial crime (Federal Reserve Bank of Boston, 2025). Unlike total identity theft, where a criminal assumes an individual's complete personality, a synthetic identity is an original construction combining elements of real Personally Identifiable Information (PII) with fabricated data—frequently using legitimate identification numbers belonging to minors, deceased persons, or individuals with minimal credit footprint (Socure, 2022; TransUnion, 2025).

Generative AI acts as a force multiplier in this process. Diffusion models and Generative Adversarial Networks (GANs) can produce hyperrealistic documents at industrial scale (Veriff, 2025; Experian, 2025). The danger lies in their patience: criminals "cultivate" these identities over months or years, building solid credit scores before executing "bust-out" attacks. TransUnion's internal analysis showed that U.S. lender exposure to synthetic identities totaled $3.3 billion in potential losses at the end of 2024 (TransUnion, 2025), while the Federal Reserve Bank of Boston (2025) reports global losses crossed the $35 billion mark in 2023.

3.2. Attack Surface Across the Customer Lifecycle

The widespread adoption of facial biometric verification as a remote onboarding standard created a false sense of security. According to Entrust's Identity Fraud Report (2025), deepfake attempts now occur at a rate of one every five minutes. Attack techniques are classified into two categories with distinct legal implications:

Presentation attacks occur in the physical world in front of the camera sensor, ranging from high-resolution photos to 3D silicone masks or augmented reality projections (ISO/IEC 30107-1:2023; iBeta, 2024). Injection attacks represent the highest sophistication level: through virtual camera emulators or API interception software, attackers inject pre-recorded or real-time generated video streams directly into the banking application, bypassing physical camera sensors entirely (Trend Micro, 2024; Sumsub, 2024).

Table **1** presents the attack taxonomy by customer lifecycle phase, identifying specific vectors and risk indicators for remote onboarding, KYC refresh, and account takeover scenarios.

4. NORMATIVE FRAMEWORK: RISK-BASED IDENTITY ASSURANCE

A defensible authenticity standard must accomplish three objectives simultaneously, as established by the FATF Guidance on Digital Identity (2020): (i) Assurance—produce sufficient confidence that the applicant is the rightful subject of the identity evidence; (ii) Proportionality—scale friction and data collection with risk, not with institutional anxiety; and (iii) Auditability—preserve verifiable proof that controls were applied correctly. The key principle is to maximize trust per datum—the minimum data necessary to achieve assurance (NIST, 2025a).

4.1. The Mexican Legal Framework: Liability and Burden of Proof

The operating environment for FinTech in Mexico is defined by a garantista legal framework that imposes substantial evidentiary challenges on institutions facing deepfake technology. Thesis 1a./J. 17/2021 (10a.) from the SCJN's First Chamber determines that when a user claims nullity of an electronic transfer, the banking institution must prove that security procedures were followed and that the operation was reliable (SCJN, 2021).

The Court's reasoning is based on created risk theory and technological asymmetry. This reversal of the burden of proof is critical: in a commercial trial, the bank must technically demonstrate that the operation was performed by the user and not by an unauthorized third party. If an attacker uses a synthetic identity authenticated with a successful deepfake, and the real user proves they were elsewhere or never opened that account, the judge will invariably rule in favor of the user if the bank cannot demonstrate its system distinguishes between the user and their digital clone. As Carpenter (2025) notes, 25.9% of executives report their organizations have experienced deepfake

Table 1: Deepfake-Enabled Fraud Attack Taxonomy by Customer Lifecycle Phase

Lifecycle Phase	Attack Vectors	Risk Level
Remote Onboarding	AI-generated faces/video; voice cloning; synthetic documents; injection attacks[1]	High
KYC Refresh	Exploitation of update flows; social engineering + deepfake calls[2]	Medium-High
Account Takeover	Deepfake voice against phone support; synthetic video during verification[3]	Critical

Notes: [1]Enables full synthetic identity creation and immediate financial access. [2]Leverages existing trust relationship to expand fraud surface. [3]Direct access to existing funds and credit lines. Source: Author's elaboration based on Veriff (2025), Entrust (2025), FS-ISAC (2024), and Signicat (2025).

Table 2: Comparative Regulatory Framework for Deepfake Risk in Financial Services

Jurisdiction	Primary Framework	Technical Requirements	Liability Approach
Mexico	Ley Fintech + CUB Annex 72; SCJN jurisprudence[1]	Biometric validation required; PAD level unspecified	Garantista: burden on institution
European Union	AI Act Art. 50; eIDAS 2.0[2]	Synthetic content must be marked detectable	Risk-based; deployer responsibility
United States	NIST SP 800-63-4; state biometric laws[3]	AAL framework (not mandatory); FIDO encouraged	Varies by state; negligence standard
Singapore	MAS Technology Risk Guidelines[4]	Risk-based with explicit biometric security	Proportionate; governance emphasis

Notes: [1]Tesis 1a./J. 17/2021 establishes strict liability tendency. [2]Article 50 requires effective, interoperable marking solutions. [3]FTC enforcement discretion predominates. [4]Emphasis on accountability frameworks. Source: Author's comparative analysis.

incidents, yet only 29% of firms have taken protective steps.

4.2. Regulatory Framework: FinTech Law and CUB

The Financial Technology Institutions Regulation Law (Cámara de Diputados, 2025) and Single Banking Circular Annex 72 (CNBV, 2024) allow non-face-to-face identification using biometrics, requiring validation against official registries (INE, RENAPO) and conducting liveness tests. However, Mexican regulation does not specify the required robustness level for those liveness tests—by not explicitly requiring certifications against presentation attacks (PAD) or injection attacks, the regulation creates a gap between regulatory compliance and real security where civil liability proliferates.

4.3. Comparative Regulatory Analysis

A comparative examination reveals significant variation across jurisdictions. While Mexico emphasizes consumer protection through burden reversal, the European Union has advanced more prescriptive technical requirements through the AI Act (European Commission, 2024), and the United States operates primarily through NIST guidance (2025a). Table **2** presents a comparative framework analysis across four key jurisdictions.

5. TECHNICAL STANDARDS AS LEGAL DEFENSE

Facing a legal environment that demands "certainty" and a threat environment that manufactures "uncertainty," the only viable defense for FinTech institutions is the adoption of international standards that serve as pre-constituted expert evidence (iBeta, 2024).

5.1. ISO/IEC 30107-3: The Science of Liveness Detection

The ISO 30107 series standardizes the evaluation of Presentation Attack Detection (PAD). ISO/IEC 30107-3:2023 establishes principles and methods for performance assessment and reporting of testing results (ISO, 2023b). Certification must be performed by accredited laboratories such as iBeta, which is accredited by NIST NVLAP (iBeta, 2024). For Level 2 testing, laboratories create and apply each species' artefacts within 8 hours, targeting 150 attacks alternated with 50 genuine presentations per species. Table **3** presents the PAD levels with their corresponding attacker profiles and legal relevance.

5.2. NIST SP 800-63B-4: Redefining Digital Identity

NIST SP 800-63B-4 (finalized July 2025) focuses on authentication and authenticator management (NIST, 2025a). The Authentication Assurance Levels (AAL) framework establishes: AAL1 permits

Table 3: ISO/IEC 30107-3 PAD Levels and Legal Relevance

PAD Level	Attacker Profile	Attack Types	Legal Relevance
Level 1 (A)	Basic; common tools[1]	Printed photos, mobile screens, simple videos	Minimum; difficult to defend as "high security"
Level 2 (B)	Moderate; specific tools[2]	Latex masks, depth-simulated videos, basic mannequins	Acceptable standard; demonstrates due diligence
Level 3 (C)	Expert; advanced resources[3]	Silicone masks, injection attacks, real-time deepfakes	Gold Standard; robust evidence of reliability

Notes: [1]Office tools per iBeta (2024). [2]3D printers, latex materials. [3]Biometrics laboratory-level resources. Source: Adapted from ISO/IEC 30107-3:2023 with legal implications analysis by the author.

single-factor authentication (insecure for financial transactions); AAL2 requires multi-factor authentication where biometrics must be linked to a physical device; AAL3 demands cryptographic proof of key possession through hardware resistant to verifier impersonation (NIST, 2025b). For combating deepfakes, the framework suggests that relying solely on remote biometrics without cryptographic hardware binding is insufficient for high-risk operations.

5.3. FIDO and Hardware Cryptography

The FIDO Alliance promotes standards using asymmetric cryptography where the private key never leaves the user's device. Biometrics are used only to unlock that key locally (FIDO Alliance, 2023). This is crucial against deepfakes: even with a perfect deepfake video, attackers cannot authenticate without physical possession of the registered device containing the FIDO private key. The April 2024 NIST supplement on syncable authenticators (passkeys) further expands these capabilities (NIST, 2024).

6. LIABILITY MODEL: ALLOCATION WHEN DEEPFAKES DEFEAT CONTROLS

Deepfake-enabled fraud creates a recurring conflict among three actors: (i) the FinTech/digital bank (deployer); (ii) the user/customer; and (iii) the verification vendor (Bateman, 2020). A workable allocation must follow three criteria: control and prevention capacity, foreseeability and known risk, and evidentiary capability.

6.1. Control and Prevention Capacity

Liability should track the actor with the greatest ability to prevent harm at lowest cost. FinTech controls product design, thresholds, and step-up rules. Vendors control model performance and anti-spoof robustness. Users control device hygiene only in a limited sense—research confirms humans cannot consistently identify AI-generated content (UNESCO, 2025; Barrington *et al*., 2025).

6.2. Foreseeability and Known Risk

Deepfake capability is now foreseeable for remote identity flows. As FS-ISAC (2024) reports, 1 in 10 companies have encountered deepfake-enabled fraud, yet 6 in 10 executives say their firms have no protocols regarding such risks. Institutions that fail to implement adequate countermeasures cannot claim surprise when attacked.

6.3. Evidentiary Capability

The party who can generate and preserve verifiable evidence should bear the burden to show controls were properly applied. In remote onboarding, that is usually the FinTech and its vendors. Table **4** presents the duty map and evidentiary burden by actor.

7. THE TIERED AUTHENTICITY AND TRACEABILITY STANDARD (TATS)

Based on the foregoing analysis, this article proposes a new compliance and responsibility paradigm for the FinTech sector. TATS is a risk-based standard linking onboarding and KYC controls to three tiers, drawing on three technical pillars: identity proofing and enrollment, authentication assurance, and biometric presentation attack detection (NIST, 2025a; FATF, 2020).

7.1. Tier 1: Low Risk (Inclusion-Friendly)

Use case: Basic wallets, low transaction limits, informational or restricted services. **Controls:** Basic document capture OR verified identity token; strong MFA and device binding; transaction caps and velocity controls. **Evidence:** Onboarding timestamp, device binding proof, limit policy, authentication logs. **Liability implication:** Zone of Shared Responsibility—institution demonstrates proportionate controls; user bears greater responsibility for exceeding intended use.

Table 4: Duty Map and Evidentiary Burden by Actor

Actor	Core Duty	Minimum Evidence Required
FinTech (Deployer)	Risk classification; control design; vendor governance; audit trail maintenance[1]	Risk score rationale; step-up documentation; transaction logs; vendor selection criteria
Verification Vendor	Performance standards; anti-spoof robustness; transparency on limitations[2]	Third-party certifications (ISO, FIDO, iBeta); PAD metrics; update logs
User/Customer	Reasonable care with devices; timely reporting; non-collusion[3]	Confirmation of device control; reporting timestamps; non-participation evidence

Notes: [1]Includes step-up trigger configuration. [2]Includes timely security updates and incident disclosures. [3]Limited control capacity recognized. Source: Author's elaboration based on FATF (2020) and Mexican jurisprudential criteria.

7.2. Tier 2: Medium Risk (Default for Most Products)

Use case: Standard digital accounts, moderate limits, common payments. **Controls:** Document authenticity checks + selfie with liveness (PAD Level 2 minimum per ISO/IEC 30107-3); capture integrity detection; step-up on risk changes; challenge variability. **Evidence:** PAD/liveness score records, capture integrity attestation, audit trail with step-up triggers. **Liability implication:** Institution can present certificates and logs as discharge evidence, potentially reversing burden of proof.

7.3. Tier 3: High Risk (Enhanced Due Diligence)

Use case: Higher limits, rapid cash-out, cross-border transfers, higher-risk profiles. **Controls:** Multi-modal verification; stronger provenance chain with secure capture and cryptographic signing (C2PA, 2024); FIDO-bound authentication; manual review on specific triggers; periodic red-team testing. **Evidence:** Enhanced due diligence rationale, multi-modal results, reviewer decision log, vendor assurance artifacts, FIDO device attestation. **Liability implication:** Safe Harbor established with reinforced iuris tantum presumption of transaction validity.

Table **5** presents the TATS operational checklist summarizing requirements across all three tiers and eight key dimensions.

8. PRIVACY-BY-DESIGN AND PROPORTIONALITY: ANTI-SURVEILLANCE GUARDRAILS

TATS is explicitly not a blank check for biometric over-collection. The framework imposes guardrails preserving user rights while enabling effective fraud prevention, consistent with FATF's emphasis on supporting financial inclusion while managing ML/TF risks (FATF, 2020). These guardrails include: (i) Data minimization—collect only what the risk tier justifies; (ii) Purpose limitation—verification artifacts are for security and AML compliance, not profiling; (iii) Retention discipline—keep evidence for audit and dispute resolution, then delete safely; and (iv) Explainability—communicate why step-up verification is triggered in plain language.

This is where TATS becomes compatible with sustainable finance: inclusion requires trust, and trust requires safeguards that do not punish legitimate users for being "hard to verify" (UNESCO, 2025). The framework explicitly requires that Tier 1 remains viable for low-risk products to prevent de-banking of vulnerable populations, consistent with the World Bank's emphasis on digital ID pathways to financial access (FATF, 2020).

9. POLICY RECOMMENDATIONS

Based on the foregoing analysis, the following recommendations are directed toward Mexican regulators (CNBV, Banco de México) while remaining applicable to other emerging market jurisdictions:

(1) Adopt formal tiering in secondary regulation: Link onboarding assurance requirements explicitly to risk levels and transaction limits, following FATF's guidance on tiered CDD (FATF, 2020). **(2) Mandate auditability and proof-of-check evidence:** Require demonstrable artifacts proving controls were applied, with minimum retention periods aligned with NIST recommendations (NIST, 2025a). **(3) Clarify liability default rules:** Deployers should retain primary responsibility toward users, with contractual recourse against vendors, consistent with EU AI Act deployer obligations (European Commission, 2024). **(4) Promote provenance and integrity standards:**

Table 5: TATS Operational Checklist

Dimension	Tier 1 (Low)	Tier 2 (Medium)	Tier 3 (High)
Transaction Limits	Low (capped)	Medium (standard)	High (with EDD)
Biometrics	Optional/minimal	Liveness + PAD Level 2	Multi-modal + PAD Level 3
Capture Integrity	Basic	Required[1]	Required + cryptographic[2]
Step-up Authentication	Standard triggers	Risk-triggered	Mandatory on key triggers
Hardware Binding	Device binding only	Strong device binding	FIDO-certified hardware
Audit Trail	Basic logs	Full transaction trail	Full + enhanced review notes
Privacy Posture	Data minimization	Minimization + purpose limitation	Strict retention + access controls
Liability Zone	Shared Responsibility	Shared with burden reversal	Safe Harbor

Notes: [1]Emulator detection required. [2]Cryptographic attestation per C2PA standards. Source: Author's elaboration integrating ISO/IEC 30107-3 (2023), NIST SP 800-63-4 (2025), FIDO Alliance (2023), and FATF (2020).

Encourage adoption of capture integrity standards referencing ISO/IEC 30107 series, FIDO certification, and C2PA initiatives (2024). **(5) Implement inclusion testing:** Require that Tier 1 controls remain viable for low-risk products with explicit regulatory impact assessments on financial inclusion. **(6) Establish technical certification safe harbors:** Create regulatory acknowledgment that institutions implementing certified PAD Level 3 and FIDO-bound authentication benefit from favorable evidentiary presumptions.

CONCLUSION

The era of digital innocence has ended. The proliferation of deepfakes and synthetic identity fraud has eroded the foundations of trust upon which remote banking was built (Chesney & Citron, 2019; UNESCO, 2025). For FinTech institutions, the challenge is twofold: defending against attackers operating at machine speed and navigating a legal environment that severely penalizes security breaches.

The analysis confirms that Mexico's current legal framework, as interpreted by the SCJN, places financial institutions in an extremely vulnerable position if they cannot prove system reliability. The mere password or static selfie is no longer sufficient to meet evidentiary standards in commercial litigation. However, institutions that embrace rigorous technical standards can construct solid legal defenses where technical certification becomes legal evidence.

The Tiered Authenticity and Traceability Standard (TATS) proposed in this article offers a pathway to restore equilibrium by integrating ISO/IEC 30107-3 standards with risk-based authentication architectures aligned with NIST SP 800-63-4. The liability allocation model grounds responsibility in control capacity, foreseeability, and evidentiary capability—principles aligning with both civil law theory and practical fraud prevention.

KEY POLICY, REGULATORY, AND INDUSTRY IMPLICATIONS

Policy implications: Regulators must transition from prescriptive rules to outcome-based standards that recognize certified technical controls as presumptive compliance evidence. Financial inclusion objectives require maintaining accessible Tier 1 pathways that do not exclude vulnerable populations through excessive verification requirements.

Regulatory implications: The gap between formal regulatory compliance and actual security effectiveness must be addressed through explicit PAD and injection attack resistance requirements. Cross-border harmonization of digital identity assurance standards would reduce compliance fragmentation for multinational FinTech operators.

Industry implications: Financial institutions must invest in certified verification infrastructure not merely as operational expense but as litigation insurance. Vendor selection criteria should prioritize independently certified PAD performance over marketing claims. Collaborative threat intelligence sharing through organizations like FS-ISAC provides collective defense benefits that individual institutions cannot achieve alone.

Directions for Future Research: Several avenues merit further investigation. First, empirical testing of PAD effectiveness across diverse demographic populations would validate whether current certification standards adequately address bias and inclusion concerns. Second, longitudinal analysis of cross-jurisdictional litigation outcomes in deepfake-enabled fraud cases would provide evidence-based guidance for liability allocation frameworks. Third, the development of standardized metrics for measuring the cost-effectiveness of different TATS tier implementations would support regulatory impact assessments. Fourth, research into adversarial AI evolution—particularly the arms race between deepfake generation and detection technologies—would inform forward-looking regulatory frameworks. Finally, comparative studies of user acceptance and behavioral responses to different verification friction levels would optimize the balance between security and user experience.

Ultimately, the fight against AI-enabled fraud is not merely about protecting financial balances; it is about preserving the viability of an inclusive and democratic financial system. The correct response is neither surveillance maximalism nor "best-effort" vendor patchwork. TATS provides a practical, risk-based standard that hardens onboarding and KYC while preserving proportionality, minimizing data collection, and strengthening auditability. The adoption of verifiable authenticity standards is both the ethical and operational imperative for the next decade of Financial Technology.

REFERENCES

AI Act Service Desk. (2024). Article 50: Transparency obligations for providers and deployers of certain AI systems. European Commission. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-50 (Accessed: January 15, 2025).

Barrington, L., *et al*. (2025). Research confirms humans cannot consistently identify AI-generated voices. Journal of Experimental Psychology: Applied. [Cited in UNESCO, 2025].

Bateman, J. (2020). Deepfakes and synthetic media in the financial system: Assessing threat scenarios. Carnegie Endowment for International Peace. https://carnegieendowment.org/research/2020/07/deepfakes-and-synthetic-media-in-the-financial-system-assessing-threat-scenarios (Accessed: December 10, 2024).

C2PA. (2024). Content Credentials: Technical specification for content provenance and authenticity. Coalition for Content Provenance and Authenticity. https://c2pa.org/specifications/ (Accessed: January 20, 2025).

Cámara de Diputados (México). (2025). Ley para Regular las Instituciones de Tecnología Financiera (última reforma DOF 14-11-2025). https://www.diputados.gob.mx/LeyesBiblio/pdf/LRITF.pdf (Accessed: January 25, 2025).

Carpenter, P. (2025). AI, deepfakes, and the future of financial deception. Testimony before the U.S. Securities and Exchange Commission. KnowBe4. https://www.sec.gov/files/carpenter-sec-statements-march2025.pdf (Accessed: January 28, 2025).

Chen, H., & Magramo, K. (2024, February 4). Finance worker pays out $25 million after video call with deepfake 'chief financial officer'. CNN. https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk (Accessed: December 15, 2024).

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107, 1753-1819. https://doi.org/10.2139/ssrn.3213954

Comisión Nacional Bancaria y de Valores (CNBV). (2024). Circular Única de Bancos (Anexo 72: Disposiciones en materia de identificación). CNBV.

Deloitte Center for Financial Services. (2024). Deepfake banking fraud risk on the rise. Deloitte Insights. https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html (Accessed: December 20, 2024).

Entrust. (2025). 2025 Identity fraud report: The deepfake threat landscape. Entrust Corporation.

European Commission. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act). Official Journal of the European Union.

Experian. (2025, March 18). 'Synthetic fraud' reaches record levels [Press release]. Experian. https://www.experianplc.com/newsroom/press-releases/2025/-synthetic-fraud--reaches-record-levels (Accessed: January 22, 2025).

FATF. (2020). Guidance on digital identity. Financial Action Task Force (FATF/OECD). https://www.fatf-gafi.org/en/publications/Financialinclusionandnpoissues/Digital-identity-guidance.html (Accessed: December 5, 2024).

Federal Reserve Bank of Boston. (2025, April 17). Gen AI is ramping up the threat of synthetic identity fraud. Boston Fed News. https://www.bostonfed.org/news-and-events/news/2025/04/synthetic-identity-fraud-financial-fraud-expanding-because-of-generative-artificial-intelligence.aspx (Accessed: January 25, 2025).

FIDO Alliance. (2023). Biometric component certification requirements (v3.0). FIDO Alliance. https://fidoalliance.org/certification/ (Accessed: January 10, 2025).

Fortune. (2024, May 17). A deepfake 'CFO' tricked the British design firm behind the Sydney Opera House in $25 million fraud. Fortune. https://fortune.com/europe/2024/05/17/arup-deepfake-fraud-scam-victim-hong-kong-25-million-cfo/ (Accessed: December 18, 2024).

FS-ISAC. (2024). Deepfakes in the financial sector: Understanding the threats, managing the risks. Financial Services Information Sharing and Analysis Center. https://www.fsisac.com/hubfs/Knowledge/AI/DeepfakesInTheFinancialSector-UnderstandingTheThreatsManagingTheRisks.pdf (Accessed: January 5, 2025).

iBeta Quality Assurance. (2024). ISO 30107-3 presentation attack detection test methodology and confirmation letters. https://www.ibeta.com/iso-30107-3-presentation-attack-detection-confirmation-letters/ (Accessed: January 12, 2025).

ISO. (2023a). ISO/IEC 30107-1:2023 Information technology — Biometric presentation attack detection — Part 1: Framework. International Organization for Standardization. https://www.iso.org/standard/83828.html

ISO. (2023b). ISO/IEC 30107-3:2023 Information technology — Biometric presentation attack detection — Part 3: Testing and reporting. International Organization for Standardization. https://www.iso.org/standard/79520.html

Javelin Strategy & Research. (2024). 2024 Identity fraud study: Resolving the shattered identity crisis. https://javelinstrategy.com/research/2024-identity-fraud-study-resolving-shattered-identity-crisis (Accessed: December 22, 2024).

Lucinity. (2025). AI-enabled fraud trends 2024-2025: Annual threat assessment. Lucinity Research.

Monetary Authority of Singapore (MAS). (2024). Technology risk management guidelines. MAS.

NIST. (2023). NIST AI 100-1: Artificial intelligence risk management framework (AI RMF 1.0). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1.jpn

NIST. (2024). NIST SP 800-63Bsup1: Incorporating syncable authenticators into NIST SP 800-63B. National Institute of Standards and Technology.

NIST. (2025a). NIST SP 800-63-4: Digital identity guidelines. National Institute of Standards and Technology.

NIST. (2025b). NIST SP 800-63B-4: Digital identity guidelines — Authentication and authenticator management. National Institute of Standards and Technology. https://csrc.nist.gov/pubs/sp/800/63/b/4/final (Accessed: January 30, 2025).

Signicat. (2025, March 28). Fraud attempts with deepfakes have increased by 2137% over the last three years [Press release]. Signicat. https://www.signicat.com/press-releases/fraud-attempts-with-deepfakes-have-increased-by-2137-over-the-last-three-year (Accessed: January 28, 2025).

Socure. (2022). The state of synthetic fraud: Evolution, trends, and how we will eradicate it by 2026. https://www.socure.com/news-and-press/socure-estimates-financial-losses-from-synthetic-fraud-to-reach-nearly-5-billion-by-2024 (Accessed: January 8, 2025).

Sumsub. (2024). Global deepfake incidents surge tenfold from 2022 to 2023. https://sumsub.com/newsroom/sumsub-research-global-deepfake-incidents-surge-tenfold-from-2022-to-2023/ (Accessed: December 28, 2024).

Suprema Corte de Justicia de la Nación (SCJN). (2021). Tesis 1a./J. 17/2021 (10a.). Transferencias electrónicas. Carga de la prueba en casos de reclamación por operaciones no reconocidas. Semanario Judicial de la Federación, Décima Época.

TransUnion. (2025). Money 20/20: What's behind the rise in synthetic identity fraud. https://www.transunion.com/blog/money-2020-whats-behind-rise-synthetic-identity-fraud (Accessed: January 18, 2025).

Trend Micro. (2024, February 7). Deepfake CFO video calls result in $25MM in damages. Trend Micro Research. https://www.trendmicro.com/en_us/research/24/b/deepfake-video-calls.html (Accessed: December 12, 2024).

UNESCO. (2025, October 27). Deepfakes and the crisis of knowing. UNESCO.

https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing (Accessed: January 30, 2025).

Veriff. (2025). Identity fraud report 2025: Global trends in synthetic identity and deepfake attacks. Veriff.

https://doi.org/10.31875/2755-8398.2026.02.01

© 2026 Benavides *et al.*
This is an open-access article licensed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the work is properly cited.