International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(21s) 2026 pp. 1602-1617
https://doi.org/10.70917/ijcisim-2026-5898
Article

The Synthetic Threat: AI-Generated Identity Fraud, Strategic Risk Management, and Deep Learning in U.S. Banking.

Shariful Haque[1], Sandipon Chowdhury[2], Kamron Naher[3], Morium Akter Munny[4], Roksana Akter[5], Azamat Mambetaliev[6]

[1] Department of Business Administration, International American University, Los Angeles, California, United States.
Email: research@sharifulhaque.org ORCID: 0009-0003-0832-5539
[2] Technology and Leadership, Indiana Wesleyan University, Woodside, New York, United States.
Email: sandipon.chowdhury@myemail.indwes.edu
ORCID: 0009-0004-2588-6247
[3] MBA in Project Management, Central Michigan University, Mt Pleasant, Michigan, United States.
Email: naher1k@cmich.edu ORCID: 0009-0009-5086-8856
[4] Master of Science in Computer Science, California State University San Marcos, San Marcos, California, USA.
Email: moriumakter5453@gmail.com ORCID: 0009-0000-5920-9608
[5] Department of Business Administration (Cybersecurity), California state university San Bernardino, San Bernardino, California, United States. Email: 008794273@coyote.csusb.edu ORCID: 0009-0009-8609-3570
[6] College of Technology & Engineering, Westcliff University Irvine, California, United States.
Email: a.mambetaliev.2674@westcliff.edu ORCID: 0009-0007-8524-6761
Correspondence Email: n.mohammad.254@westcliff.edu

Abstract: Generative artificial intelligence has reduced the cost of producing realistic identity documents, facial imagery, voice samples, and supporting application narratives. These capabilities intensify a long-standing U.S. banking problem: synthetic identities that combine genuine and fabricated attributes and mature over time before a credit bust out or account takeover. This paper develops a privacy-preserving deep learning architecture for detecting AI-generated identity fraud during remote onboarding and early account activity. The design transfers three source-derived elements: Hybrid Masked Encryption (HME) for protected data aggregation, an Intel SGX-style Trusted Execution Environment (TEE) for attested enclave processing, and a five-layer cybersecurity model connecting data resources, preprocessing, intelligent analytics, decision orchestration, and governance. A multimodal detector is then proposed to fuse document-forensics, selfie-video, liveness, application, credit-file, device, network, and behavioral signals. The source HME study reports 25 ms encryption and 20 ms decryptions for a 1 MB dataset, while the TEE study offers proof-of-concept architecture using remote attestation, local attestation, and sealed storage. These findings support technical feasibility but do not establish fraud-detection accuracy in banking. Accordingly, this paper defines a temporal and generator-holdout validation protocol centered on precision-recall performance, recalling operational false-positive constraints, calibration, demographic error analysis, latency, and privacy leakage. The principal contribution is an auditable blueprint that joins fraud intelligence with data protection, while preserving a clear boundary between source evidence and results that still require empirical testing.
Keywords: Synthetic Identity Fraud, Deepfake Detection, Deep Learning, U.S. Banking, Trusted Execution Environment, Hybrid Masked Encryption, Cybersecurity Governance, Privacy-Preserving Analytics

1. Introduction

Synthetic identity fraud is neither ordinary identity theft nor a single forged document. It is an assembled persona. A fraudster may combine a valid Social Security number with an invented name, date of birth, address, email

Copyright: © 2026 by the authors

history, facial image, and employment story. The account can appear low risk at first because the identity has little adverse history. Over months or years, the perpetrator may cultivate creditworthiness and then exploit the accumulated limits. The Federal Reserve describes this build-and-bust pattern as especially difficult to resolve because the nominal customer never existed as a real person (Federal Reserve Board, 2019).

Generative AI changes the scale and texture of the problem. High-quality face swaps, synthetic portraits, document images, voice clones, and automatically generated correspondence can be created rapidly and varied across applications. These tools help an attacker maintain consistency within a single channel while hiding contradictions that only become visible when a bank connects biometric, documentary, device, credit, and behavioral evidence. The issue therefore cannot be reduced to deepfake classification. It is an identity-resolution problem, a presentation-attack problem, and a longitudinal fraud detection problem.

The scale of identity exploitation provides a practical reason for building stronger controls. FinCEN reported that approximately 1.6 million identity-related Bank Secrecy Act reports filed for 2021 represented $212 billion in suspicious activity, although the reported amount includes attempted or continuing activity and should not be interpreted as confirmed loss (Financial Crimes Enforcement Network [FinCEN], 2024a). More recently, FinCEN observed increased suspicious activity reporting involving deepfake media, including fraudulent identity documents intended to bypass verification and authentication (FinCEN, 2024b).

Detection alone is not sufficient. Remote onboarding systems process highly sensitive evidence: government identifiers, face and voice biometrics, device telemetry, addresses, financial history, and the bank's own fraud labels. Centralizing these materials for model training or inference creates a second risk: the defense system itself can become a high-value surveillance or breach target. The present paper responds by treating privacy protection as part of the model architecture rather than as a policy appendix.

1.1 Research Purpose and Questions

The purpose of this study is to design a multimodal deep learning framework that can identify AI-assisted synthetic identity fraud while limiting the exposure of raw identity evidence. The study asks whether document, biometric, transactional, and device signals can be fused into a calibrated risk score; how HME and TEE mechanisms can protect the data path; and which evaluation design can measure generalization to unseen generators, institutions, and time periods without overstating performance.

1.2 Contributions

The paper makes four contributions. First, it separates synthetic identity fraud into documentary, biometric, relational, device, and behavioral subproblems. Second, it adapts HME, enclave-based trusted processing, and layered cybersecurity governance from the attached studies to a banking use case. Third, it specifies a multimodal detector and a deployment workflow that produces calibrated scores and supports human review. Fourth, it defines an empirical protocol with temporal, cross-bank, and generator-holdout testing, including fairness, privacy, and latency criteria. The framework is intentionally described as proposed until those experiments are completed.

2. Background and Related Work

2.1 Synthetic Identities and Remote Banking

Remote account opening distributes identity proofing across mobile devices, third-party document services, biometric vendors, credit bureaus, and internal decision engines. Each component can be individually plausible while the combined identity remains false. Thin credit files recently issued contact points, reused devices, conflicting geolocation, improbable household relationships, and repeated document templates may become informative only when compared across applications. The operational challenge is therefore one of entity resolution under adversarial conditions.

U.S. interagency guidance recommends layered authentication and reliable verification that does not depend solely on knowledge-based questions. It also recognizes synthetic identities and impersonation as risks that financial institutions should address through ongoing monitoring and risk assessment (Federal Financial Institutions Examination Council, 2021). A model deployed in this setting must support, rather than replace, customer identification, due diligence, sanctions, suspicious activity, privacy, and information security processes.

2.2 Deepfake, Presentation-Attack, and Adversarial Detection

The deepfake literature offers a mature map of manipulation methods, yet its most useful lesson for banking is caution. Tolosana et al. (2020) organize face manipulation and detection research across the rapidly expanding range of synthetic techniques. Verdoliva (2020) places deepfakes within the wider field of media forensics, where traces may arise from content, acquisition, or post-processing. Mirsky and Lee (2021) likewise describe creation and detection as a moving contest rather than a problem solved by one artifact classifier. That framing matters at account opening because a fraudster can change the generator, compress the video, replay it through a virtual camera, or combine genuine and synthetic evidence within the same session.

Large surveys also reveal why benchmark scores do not transfer automatically to financial identity proof. Nguyen et al. (2022) review both generation and detection pipelines and emphasize the breadth of the attack space. FaceForensics++ supplies a reproducible benchmark across several manipulation and compression conditions (Rössler et al., 2019). Celeb-DF raises the visual quality of synthetic videos and thereby makes detection more demanding (Li, Yang, et al., 2020). Even so, neither benchmark recreates a mobile banking journey with identity documents, liveness prompts, device telemetry, credit-file history, and repeated applications linked across time.

Presentation-attack detection adds a second line of defense. Yu et al. (2023) survey deep face anti-spoofing methods and the persistent difficulty of generalizing across attack instruments and capturing domains. A detector that recognizes a familiar face swap may still miss a high-quality replay, a display attack, or a camera-injection pipeline. The practical implication is to test media origin, liveness, sensor integrity, and cross-modal consistency together instead of asking one visual network to carry the entire decision.

Identity-aware learning is especially relevant because synthetic fraud concerns a claimed person, not merely a manipulated frame. Huang et al. (2023) learns discrepancies between the identity represented in a source image and the identity expressed by a suspected face swap. Dong et al. (2023) show the opposite hazard: a deepfake detector can learn who appears in a dataset instead of learning the manipulation. Their warning supports identity-disjoint partitions and unseen-generator tests in the proposed banking protocol.

Adversarial robustness deserves the same attention as ordinary classification error. Akhtar and Mian (2018) describe how small, crafted input changes can mislead computer-vision models. Biggio and Roli (2018) place those attacks within a broader decade of adversarial machine learning, including evasion and poisoning threats. Zhou et al. (2023) connects modern attacks and defenses directly to cybersecurity. For a bank, the relevant test set must therefore include adaptive perturbations, poisoned updates, capture-channel abuse, and novel generators, not only clean examples drawn from the same distribution used for training.

2.3 Financial Fraud Detection Under Operational Constraints

Financial fraud research provides an equally important counterweight to the media-forensics literature. Ngai et al. (2011) classify the data-mining techniques used in financial fraud detection and show that the problem spans several fraud types and analytic traditions. Abdallah et al. (2016) describe fraud detection as a system problem shaped by data, algorithms, and the changing behavior of offenders. West and Bhattacharya (2016) review intelligent financial-fraud methods and highlight the difficulty of comparing models across inconsistent datasets and evaluation designs. Taken together, these reviews argue against presenting a synthetic-identity detector as an isolated image classifier.

Payment-card work also exposes the gap between academic accuracy and operational usefulness. Ryman-Tubb et al. (2018) compare research practice with industry requirements and underline the importance of realistic validation.

Dal Pozzolo et al. (2018) address the rarity and evolving nature of credit-card fraud through a realistic modeling and learning strategy. Jurgovsky et al. (2018) treat card activity as a sequence rather than a collection of independent transactions. Synthetic identity fraud is not identical to card fraud, but the same principles carry over: preserve the real class imbalance, respect time, model linked behavior, and report performance at the threshold the institution can operate.

2.4 Current Identity-Proofing Expectations

NIST's current identity-proofing guidance directly addresses injection attacks and forged media. It calls for media-origin controls, analysis of submitted images and video for manipulation, authenticated protected channels, sensor or device attestation, documented false-positive and false-negative performance, human review, and demographic testing. For remote biometric collection, the guidance also specifies presentation attack detection with an impostor attack presentation accept rate below 0.07 under the referenced testing standard (National Institute of Standards and Technology [NIST], 2025). The proposed system maps these requirements to measurable technical controls rather than treating compliance as a binary label.

2.5 Recent Privacy, Security, and Financial-AI Literature

Recent research strengthens the connection between adversarial biometric security and financial identity verification. Raihan et al. (2026) specifically examine adversarial robustness for biometric verification across mobile financial applications. That setting is the closest sector-specific match among the additional sources because synthetic onboarding attacks can manipulate the same biometric decision boundary that mobile financial services use to authenticate customers.

Privacy-preserving collaboration is addressed from several complementary perspectives. Tuhin et al. (2024) frame federated machine learning as a means of sharing cyber threat intelligence across cloud-integrated MIS platforms without centralizing all institutional data. Das et al. (2025) extend this direction through homomorphic federated models designed to protect information during collaborative analytics. Roy et al. (2026) connect federated learning with blockchain and zero-trust principles in privacy-preserving MIS, providing a governance-oriented complement to the protected aggregation and enclave controls used in this paper.

The remaining sources reinforce the operational security layer. Himel et al. (2026) focus on privacy-aware data-protection models for cloud environments, a direct concern when onboarding evidence and fraud models are processed through external infrastructure. Tuhin et al. (2026) examine AI-augmented security analytics for real-time threat identification, supporting the need for rapid monitoring around model inference and account opening. Ansar et al. (2026) address resilient MIS design through cryptographic and cybersecurity techniques, which aligns with the paper's emphasis on encryption, attestation, access control, auditability, and failure containment.

2.6 Distributed Learning, Cryptography, and Explainability

Cross-institution learning is attractive because a single bank sees only a narrow slice of a coordinated fraud campaign. Yang et al. (2019) explain the main federated-learning settings and the opportunity to train across data owners without pooling every raw record. Li, Sahu, et al. (2020) describe the statistical, systems, and privacy difficulties that accompany that architecture. Mothukuri et al. (2021) focus specifically on federated learning's security and privacy risks. In other words, keeping data local changes the attack surface; it does not make collaborative learning private or trustworthy by default.

Cryptography can strengthen that collaboration when its role is stated precisely. Acar et al. (2018) survey homomorphic-encryption schemes that permit computation over protected values, while also documenting differences in capability and implementation cost. In the proposed framework, encrypted aggregation is therefore a controlled mechanism for portfolio statistics or distributed updates, not a blanket claim that every stage of model development is confidential.

Interpretability becomes consequential once a score influences access to a financial service. Rudin (2019) argues that high-stakes settings should prefer interpretable models when an equally capable transparent alternative is available. Rudin (2022) restates the concern more briefly: a post-hoc explanation of a black box is not the same thing as understanding the decision process. The present design therefore uses modality-level evidence, uncertainty, calibrated scores, and contestable human review; it does not treat a polished explanation as proof that the underlying model is fair or correct.

3. Evidence Base from the Attached Articles

3.1 Hybrid Masked Encryption

The first attached article proposes Hybrid Masked Encryption for privacy-preserving cloud aggregation. Its workflow masks sensitive attributes, applies a hybrid of symmetric and homomorphic encryption, aggregates protected values in the cloud, and releases results only to authorized consumers with managed keys. For a 1 MB dataset, the study reports HME encryption in 25 ms, decryption in 20 ms, and total computational overhead of 45 ms. The same evaluation reports encryption times of 180 ms for AES, 250 ms for RSA, 100 ms for differential privacy, and 85 ms for fully homomorphic encryption (Kesavan et al., 2026).

These measurements support a feasibility hypothesis: a protected feature path may be fast enough for near-real-time onboarding. They do not establish performance in a U.S. bank, at production concurrency, with hardware key management, failures, network delay, or regulatory logging. The source also compares nominal 'security strength' in bits across primitives that do not share a directly comparable key-size interpretation. This paper therefore carries forward the workflow and reported runtime values, but not the claim that the bit values constitute a valid universal ranking of security.

3.2 Trusted Execution Environment-Based Aggregation

The second attached article proposes a fog/cloud data aggregation design using Intel SGX. A data manager creates trusted aggregators inside enclaves. Remote attestation establishes trusted channels between devices or gateways and the enclave controller; local attestation protects communication among enclaves; sealing protects persistent data; and enclave pooling or sharing is suggested to reduce initialization cost. The article presents preliminary proof-of-concept findings and identifies communication, enclave creation, memory, and response-time as matters for future measurement (Kesavan & Srinivasan, 2024).

For banking, the transferable contribution is the trust boundary. Raw images, biometric templates, application attributes, and model parameters can be decrypted only inside a verified enclave. Keys should be released after successful attestation, and outputs should be limited to necessary scores, reason codes, and audit events. This reduces exposure to a compromised host or cloud administrator, although it does not eliminate side-channel, supply-chain, model-poisoning, or endpoint-capture risks.

3.3 Layered Cybersecurity and MIS Governance

The third attached article uses a bounded systematic-review method to synthesize fifteen studies published in 2023 and 2024, supplemented by seven foundational sources. It groups the literature into AI- and ML-driven detection, cloud infrastructure and sustainability, MIS governance and decision support, human and organizational factors, and privacy-preserving or federated approaches. From those themes, the authors derive a five-layer conceptual framework comprising data and resources, collection and preprocessing, intelligent analytics, decision and orchestration, and response and governance (Mudaliyar & Kumar, 2024).

The review finds comparatively strong attention to detection capability and MIS governance but thinner evidence concerning implementation-level explainability, human behavior in AI-assisted operations, energy-aware security, and the practical limits of federated learning. Its framework is conceptual rather than experimentally validated. The present paper therefore uses the layered structure to organize the banking system and its governance

loop, while treating fraud accuracy, analyst effectiveness, and protected deployment performance as questions for empirical testing.

Table 1. Crosswalk From the Attached Studies to the Proposed Banking Framework

Evidence component	Transferred mechanism	Banking adaptation	Evidence limitation
Hybrid Masked Encryption	Masking, hybrid encryption, encrypted aggregation, key management	Protect feature transfer and portfolio-level analytics	Quantitative runtime results exist, but banking workload and security claims require independent replication
SGX/TEE aggregation	Remote and local attestation, enclave controller, sealed storage, trusted aggregators	Confine model inference and sensitive feature fusion to an attested boundary	Architecture and preliminary evidence only; no complete performance or adversarial validation
Layered cybersecurity review	Data, preprocessing, analytics, orchestration, and governance layers with continuous feedback	Organize onboarding detection, human review, monitoring, and policy ownership	Fixed-corpus review and conceptual framework; no banking deployment or performance validation
Proposed deep learning layer	Multimodal encoders, attention fusion, calibrated risk and uncertainty	Detect synthetic identities across documents, biometrics, applications, and devices	New contribution of this manuscript; no detection metrics are claimed

4. Proposed Methodology

4.1 Study Design and Evidence Boundary

The study follows a design-science structure. The artifact is a privacy-preserving fraud-detection system; the evaluation is a staged set of experiments rather than a claim that deployment has already occurred. Source-derived findings are used to choose security components and formulate latency hypotheses. Fraud accuracy, cross-bank portability, generator robustness, privacy leakage, and demographic performance remain empirical questions. This boundary should be retained in any journal submission until reproducible experiments are completed.

4.2 Data Population and Unit of Analysis

The target population comprises U.S. consumer deposit and credit applications completed through remote channels, joined to early account and payment activity where legally permitted. The primary unit of analysis is the identity-application episode, not a single image. A secondary longitudinal unit links episodes through privacy-preserving tokens derived from devices, addresses, contact points, documents, and verified identity attributes. Labels should be based on adjudicated fraud investigations, confirmed victim reports, document-forensics outcomes, account closures for identity fraud, and well-governed negative sampling. Suspicious activity reports must not be treated as ground-truth guilt labels.

The preferred design is a multi-institution consortium with bank-local raw data and centrally comparable evaluation specifications. If only one institution is available, the temporal test set should be separated by at least one model-development cycle and preserve the natural class imbalance. Public deepfake and identity-document datasets may support pretraining, but the final model must be tested on bank-like capture conditions, document types, devices, and attack scripts.

4.3 Input Modalities and Feature Governance

Table 2. Proposed Modalities, Model Branches, and Governance Controls

Modality	Example signals	Model branch	Privacy and fairness control
Identity document	Portrait region, typography, layout, security-feature cues, barcode and visible-field consistency	Document-forensics vision encoder	Minimize retention; separate document template from person-specific content
Selfie/liveness video	Face match, temporal texture, motion, depth or challenge response, replay and virtual-camera cues	Spatial-temporal encoder with PAD head	Store derived templates only when permitted; test by sex, age, and skin tone
Application and credit	Attribute consistency, thin-file evolution, contact-point age, velocity, cross-application conflicts	Tabular transformer or residual MLP	Exclude protected attributes from decision features unless legally justified for fairness auditing
Device and network	Attestation, emulator status, IP and device reputation, sensor provenance, session timing	Categorical embeddings and sequence encoder	Use coarse location where possible; apply retention limits
Relationship graph	Shared devices, addresses, phones, employers, funding accounts, and document patterns	Graph encoder or engineered neighborhood statistics	Pseudonymize identifiers; prevent unrestricted identity graph exploration

4.4 Multimodal Deep Learning Architecture

The architecture operationalizes the five-layer source framework for a banking identity-proofing context. Bank and cloud resources form the data layer; minimization, masking, encryption, and quality control form the preprocessing layer; the multimodal network forms the intelligent-analytics layer; calibrated scoring and case routing form the orchestration layer; and human review, appeals, audit, drift monitoring, and policy updates form the response-and-governance layer. This mapping keeps model inference connected to the organizational controls that consume and supervise its output.

Each modality is encoded separately before fusion. The document, video, tabular, device, and graph encoders produce distinct feature representations. A gated cross-modal attention layer fuses those representations with a mask that records missing or unreliable modalities. The missingness mask prevents the model from treating an absent signal as a genuine zero and permits graceful operation when a capture fails.

The network produces four outputs: a synthetic-identity score, a deepfake or injection score, an attribute-consistency score, and an uncertainty estimate. A final calibrated risk score combines those outputs with policy constraints. The training objective combines weighted focal, contrastive, cross-modal consistency, and calibration terms. Focal loss reduces the influence of abundant easy negatives in an imbalanced dataset (Lin et al., 2017). The contrastive term separates genuine from synthetic identity representations, the consistency term penalizes disagreement among verified attributes and biometric evidence, and the calibration term aligns predicted risk with observed outcomes.

Training should use identity-disjoint partitions so that the same real person, device cluster, or document instance cannot appear in both development and test data. Synthetic attacks should be created with multiple generator families,

document editors, replay methods, virtual cameras, compression levels, and capture devices. At least one generator family and one attack pipeline must be excluded from training and used only for the final holdout. This design tests whether the network learns stable forensic and relational evidence rather than memorizing a generator signature.

4.5 Protected Processing with HME and TEE

The protected data path begins on the capture device or within a controlled bank service. Direct identifiers are separated from model features, unnecessary fields are removed, and sensitive feature blocks are masked. The protected payload is encrypted for an attested enclave. After the bank verifies the enclave measurement and model version, a key-management service releases a short-lived decryption key. Fusion and scoring occur inside the enclave, and raw payloads are cleared from enclave memory after the decision window. Only the minimum risk outputs, reason codes, and audit hashes leave the boundary.

Portfolio monitoring can use encrypted or differentially private aggregates rather than exporting application-level biometrics. The HME concept is most suitable for secure aggregation and transport; it should not be presented as a substitute for model governance, access control, or privacy risk assessment. Cryptography production use reviewed algorithms, documented parameters, hardware-backed keys, rotation, separation of duties, and independent penetration testing.

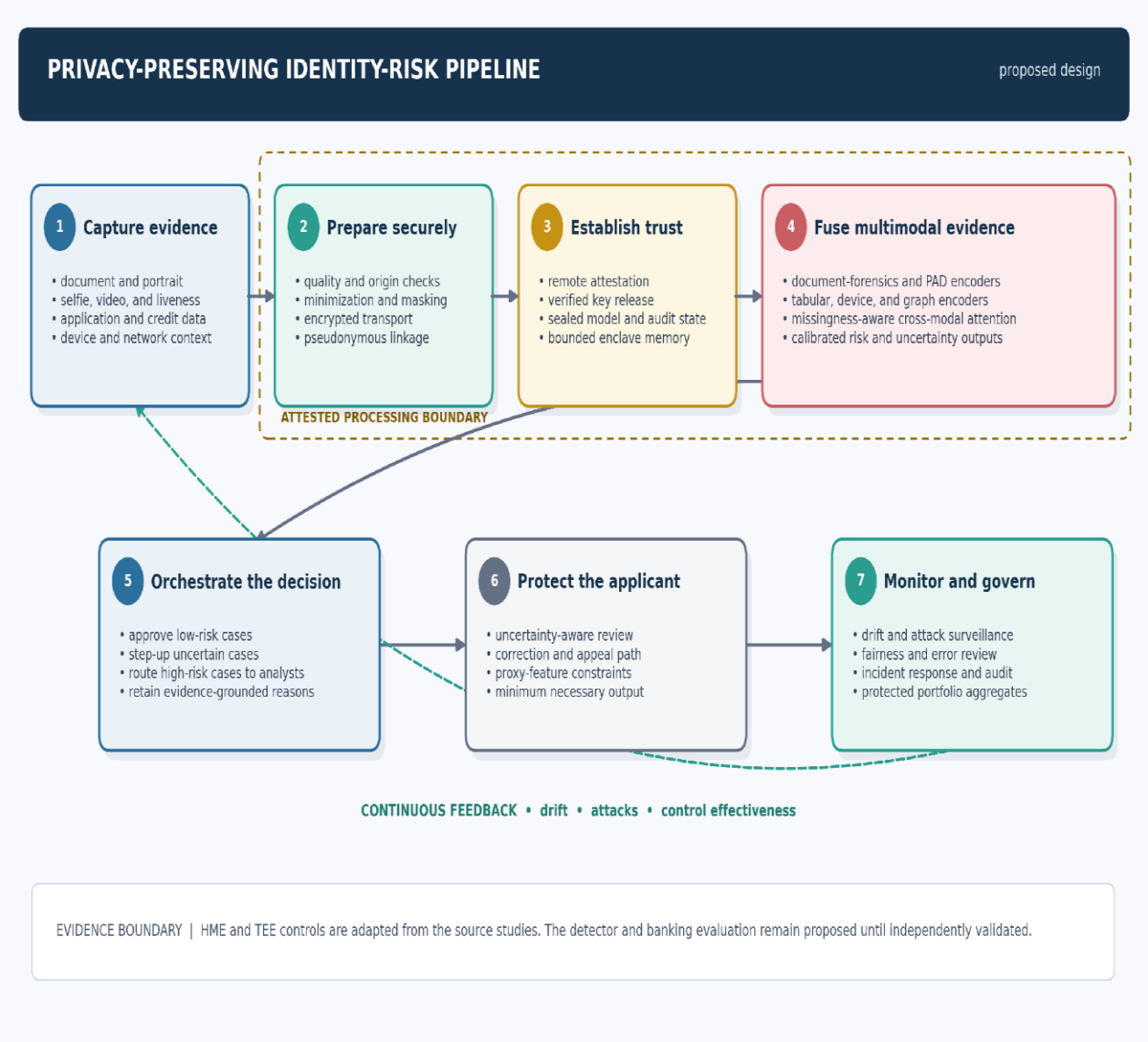


Figure 1. Proposed privacy-preserving architecture for AI-generated synthetic-identity fraud detection.

4.6 Decision Policy and Human Review

The risk score should drive a tiered response rather than an irreversible automated denial. Low-risk applications can proceed through ordinary controls. Medium-risk or uncertain cases should receive step-up proofing, live document recapture, out-of-band verification, or a trained analyst review. High-risk cases can be paused and investigated under institution policy. Human reviewers should see evidence-grounded reason codes, capture quality, model uncertainty, and conflicts among modalities, not a persuasive natural-language narrative that may encourage automation bias.

The model must also provide an appeal and correction path for legitimate applicants. A synthetic identity detector can disproportionately affect people with thin credit files, recent immigration records, nontraditional addresses, limited device history, or accessibility-related capture difficulty. These characteristics may correlate with fraud patterns without constituting fraud. Policy rules should therefore limit how weak proxy features influence a decision and require independent evidence before adverse action.

4.7 Evaluation Protocol

Table 3. Empirical Evaluation Plan

Dimension	Primary measure	Required design safeguard
Detection	PR-AUC; precision; recall; F1; recall at fixed false-positive rates	Preserve natural prevalence and report confidence intervals
Calibration	Brier score; expected calibration error; reliability curves	Fit calibration only on a validation set; never on the final test set
Generalization	Time holdout; unseen generator; unseen device; cross-bank transfer	Freeze model and thresholds before each holdout evaluation
Biometric attack resistance	IAPAR and attack-presentation classification error	Use standards-aligned PAD testing and disclose attack instruments
Fairness	False-positive and false-negative rates by sex, age, skin tone, and relevant access groups	Use the same operational threshold; investigate material disparities
Privacy and security	Membership inference, feature reconstruction, enclave threat tests, key-release failure tests	Use red-team scenarios and minimum-output review
Operations	P50/P95/P99 latency, throughput, enclave start cost, failure and fallback rates	Measure end-to-end under realistic concurrency

Because identity fraud is rare, accuracy and receiver operating characteristic area alone can create a misleading picture. Precision-recall curves should be primary, accompanied by case counts at actual operating thresholds. Thresholds must be selected using the bank's review capacity and the relative costs of a missed synthetic identity, a delayed legitimate customer, and a false adverse action. Probability calibration should be evaluated because downstream policy depends on interpretable risk estimates; temperature scaling is one established post-hoc option for neural classifiers (Guo et al., 2017).

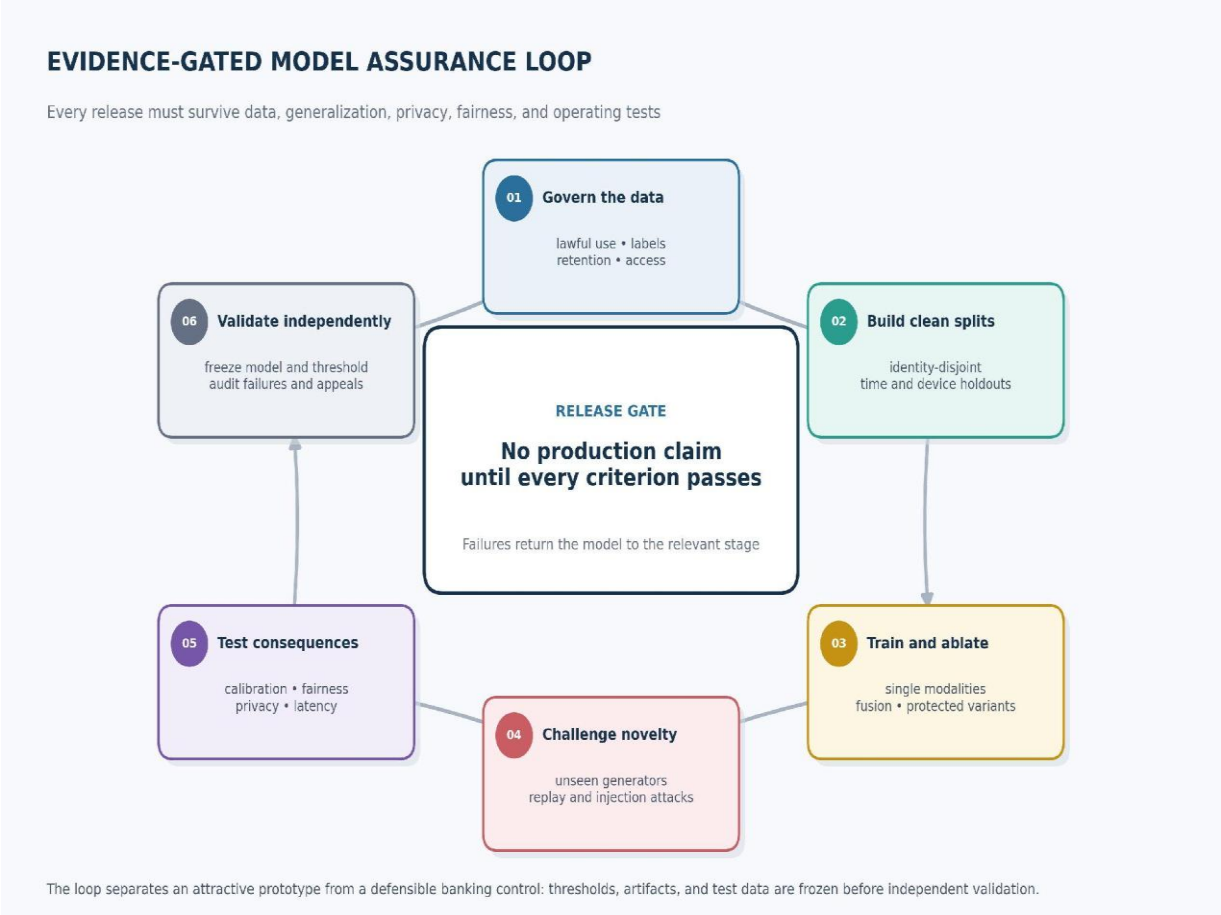


Figure 3. Evidence-gated assurance loop for development, independent validation, and monitored release.

5. Source-Derived Findings and Testable Propositions

5.1 Reported Security-Processing Findings

The quantitative evidence available from the attached materials concerns cryptographic processing, not fraud classification. In the reported 1 MB experiment, HME had the lowest encryption and decryption time among the compared methods. Figure 2 reproduces the values reported in the source. This result is relevant because onboarding systems often require responses within seconds, but it remains a small, context-specific benchmark. A bank implementation must independently reproduce the comparison using defined hardware, software, cryptographic parameters, payloads, concurrency, and network conditions.

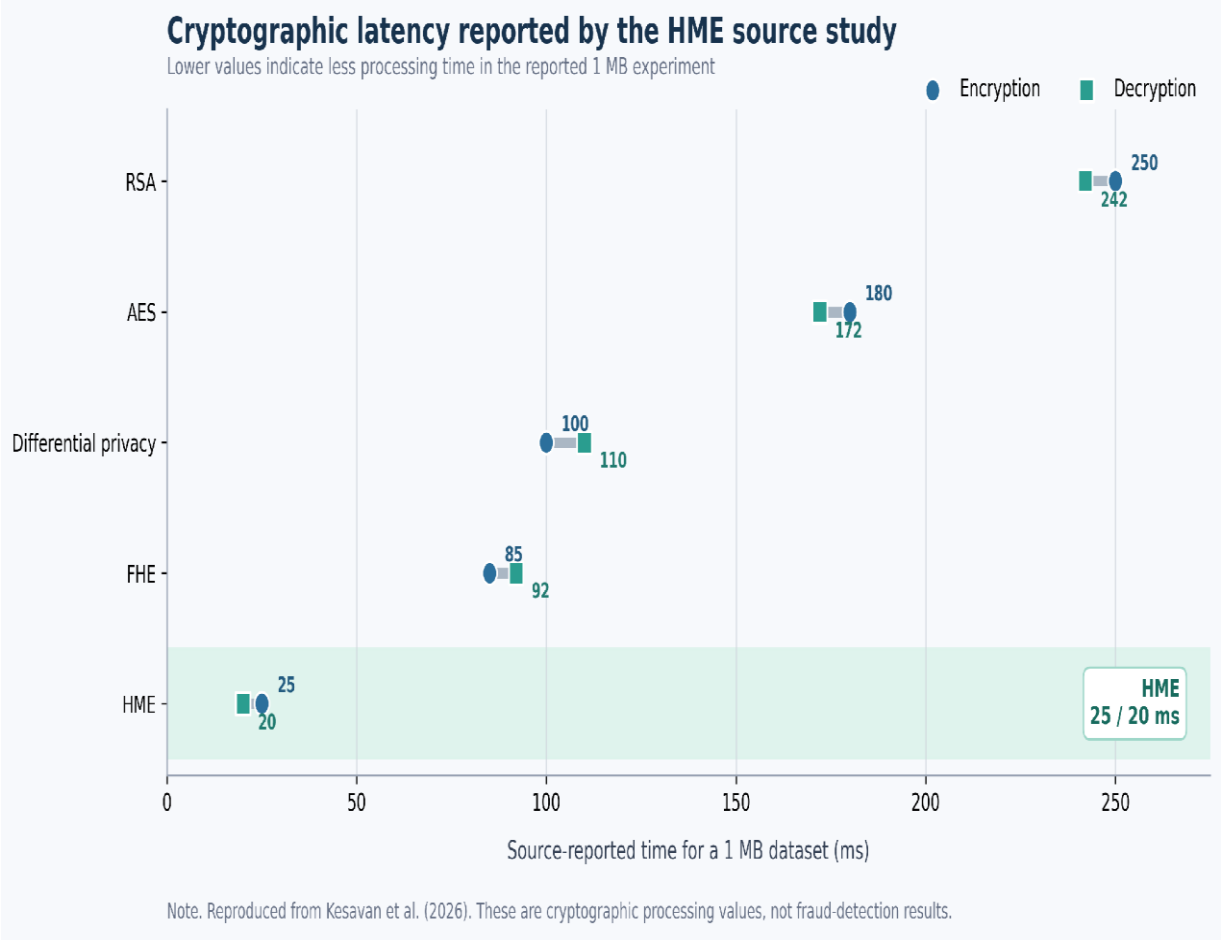


Figure 2. Source-reported encryption and decryption times for 1 MB dataset. Values are not fraud-detection metrics.

5.2 Trusted-Processing Findings

The TEE article supports the architectural proposition that heterogeneous sensitive data can be processed within attested enclaves and stored through sealing while remaining hidden from an untrusted host. Its practical contribution is a control pattern: remote attestation for external participants, local attestation among enclave components, trusted storage, and an enclave controller that routes workloads. The article does not provide final measurements for response time, encrypted communication, or enclave initialization. These factors therefore become explicit variables in the proposed banking experiment rather than assumed benefits.

5.3 Layered Cybersecurity and Governance Findings

The third source supports a cross-layer interpretation of cybersecurity: useful detection depends on governed data collection, interpretable orchestration, human response, and feedback loop, not solely on a classifier. It also identifies material evidence gaps in explainability, analyst behavior, energy-aware operations, and adversarial evaluation of federated learning. Because the source framework is a literature-derived synthesis, these findings justify the architecture and evaluation dimensions used here but do not establish that the proposed banking implementation will be accurate, efficient, or operationally effective.

5.4 Testable Propositions

The combined evidence leads to four testable propositions. First, multimodal fusion will improve recall at a fixed false-positive rate over any document-only or face-only detector because synthetic identity fraud produces

contradictions across channels. Second, generator-holdout and identity-disjoint training will reduce in-sample performance but improve the credibility of cross-attack estimates. Third, an HME-plus-TEE deployment can protect raw features with operationally acceptable latency, provided enclave initialization, key release, and encrypted communication are engineered and measured end to end. Fourth, a governed orchestration and human-review layer will improve alert actionability and audit completeness relative to presenting reviewers with raw model scores alone. None of these propositions is reported as confirmed by the present paper.

6. Discussion

6.1 Fit Between the Evidence Base and Research Scope

The evidence base aligns with a privacy-preserving synthetic-identity detector because the attached studies address secure aggregation, protected cloud processing, layered analytics, governance, and near-real-time feasibility. Those mechanisms can be integrated with supervised multimodal fraud analysis without mischaracterizing the source studies. The evidence base does not establish self-supervised representation learning, zero-day anomaly objectives, or payment-system intrusion performance. Keeping the scope centered on synthetic identity detection therefore preserves a direct chain from the source methods to the proposed banking architecture and its validation plan.

6.2 Operational Interpretation

The proposed system should be viewed as a decision-support layer within a broader control environment. A deepfake score cannot determine whether a person is legally entitled to an account, and a thin-file pattern cannot establish malicious intent. Strong cases arise when several independently governed signals converge: a suspicious document image, virtual-camera evidence, a biometric inconsistency, reused devices across unrelated identities, and a network of newly established contact points. Even then, the institution needs documented review, notice, correction, and escalation procedures.

Privacy controls also affect model quality. Minimizing raw data may reduce the richness of features, while aggressive masking can damage forensic artifacts. Enclave memory limits may constrain video models, and encrypted aggregation can complicate monitoring. These are not reasons to abandon privacy. They are design constraints that should shape model size, sampling, feature extraction, retention, and human workflow from the beginning.

6.3 Security and Governance Risks

A trusted enclave narrows but does not erase the attack surface. Side channels, rollback, vulnerable enclave code, compromised capture devices, poisoned training data, stolen reviewer credentials, and manipulated model updates remain plausible. Remote attestation verifies a measured software and hardware state; it does not prove that the model is fair, accurate, or appropriate. Likewise, encryption protects data confidentiality but cannot prevent an authorized decision maker from misusing a score. Technical controls must therefore be paired with model inventory, version approval, access review, change management, red-team testing, incident response, and independent validation.

7. Limitations

This paper has several limitations. No proprietary U.S. banking dataset was supplied, and no model was trained. The attached HME benchmark does not specify every implementation detail needed for replication, and its cross-primitive bit-strength comparison is not adopted here. The attached SGX study reports a developing proof of concept rather than completed adversarial and performance results. The layered cybersecurity study synthesizes a fixed fifteen-source corpus and proposes an unvalidated conceptual framework rather than a deployed system. Public deepfake datasets are not representative of banking capture conditions, and confirmed synthetic-identity labels are difficult to obtain. Fraud investigations can also introduce selection bias because only detected cases become positive labels. Finally, demographic fairness cannot be inferred from architecture; it must be measured with sufficiently representative data and governance-approved attributes.

These limitations determine the next research step. A credible empirical paper should preregister the data window, label rules, identity-disjoint split, excluded generator families, threshold-selection procedure, and primary metric. It should publish an ablation comparing document-only, biometric-only, tabular-only, multimodal, multimodal-plus-HME, and multimodal-plus-HME-and-TEE variants. It should also report end-to-end latency and failure behavior, not merely model inference time.

8. Conclusion

AI-generated identity fraud threatens the trust assumptions behind remote banking. The most defensible response is not a single deepfake classifier but a multimodal identity-risk system that tests whether documentary, biometric, device, credit, relational, and behavioral evidence agree. The attached studies contribute HME for protected aggregation, TEE-based attestation and sealed processing, and a five-layer model linking analytics with decision orchestration and governance. Their findings support architectural feasibility, but they do not provide banking fraud accuracy. This paper has therefore translated them into a transparent design and validation protocol rather than manufacturing results.

The proposed framework is ready for the next stage: implementation with institutionally governed data, identity-disjoint and time-based splits, unseen-generator attacks, calibrated risk scoring, standards-aligned presentation-attack testing, demographic error analysis, and end-to-end privacy and latency measurement. Only after those tests should the title's central claim detecting AI-generated identity fraud in U.S. banking - be presented as an empirically demonstrated result.

REFERENCES

1. Acar, A., Aksu, H., Uluagac, A. S., & Conti, M. (2018). A survey on homomorphic encryption schemes: Theory and implementation. *ACM Computing Surveys*, *51*(4), Article 79, 1-35. https://doi.org/10.1145/3214303
2. Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, *68*, 90-113. https://doi.org/10.1016/j.jnca.2016.04.007
3. Akhtar, N., & Mian, A. (2018). Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access*, *6*, 14410-14430. https://doi.org/10.1109/ACCESS.2018.2807385
4. Ansar, M. T. B., Sikder, M. S., Tuhin, M. K., Das, N., Chakraborty, P., & Rahman, H. (2026). Resilient MIS design using cryptographic and cyber-security techniques. In A. E. Hassanien, S. Anand, A. Jaiswal, & P. Kumar (Eds.), *Innovative computing and communications* (Lecture Notes in Networks and Systems, Vol. 2037). Springer. https://doi.org/10.1007/978-3-032-30008-9_24
5. Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, *84*, 317-331. https://doi.org/10.1016/j.patcog.2018.07.023
6. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(8), 3784-3797. https://doi.org/10.1109/TNNLS.2017.2736643
7. Das, N., Sultana, S., Sikder, M. S., Himel, H. U., Saha, U. S., & Rashed, R. A. M. (2025). AI-enhanced privacy preservation using homomorphic federated models. In *2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT)* (pp. 1-8). IEEE. https://doi.org/10.1109/ICAFT66710.2025.11453096
8. Dong, S., Wang, J., Ji, R., Liang, J., Fan, H., & Ge, Z. (2023). Implicit identity leakage: The stumbling block to improving deepfake detection generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3728-3738). https://openaccess.thecvf.com/content/CVPR2023/html/Dong_Implicit_Identity_Leakage_The_Stumbling_Block_to_Improving_Deepfake_Detection_CVPR_2023_paper.html
9. Federal Financial Institutions Examination Council. (2021). Authentication and access to financial institution services and systems: Interagency guidance. https://www.federalreserve.gov/frrs/guidance/authentication-and-access-to-financial-institution-services-and-systems-interagency-guidance.htm
10. Federal Reserve Board. (2019, July 9). Federal Reserve System white paper examines the effects of synthetic identity payments fraud. https://www.federalreserve.gov/newsevents/pressreleases/other20190709a.htm
11. Financial Crimes Enforcement Network. (2024a). Identity-related suspicious activity: 2021 threats and trends. U.S. Department of the Treasury. https://www.fincen.gov/system/files/shared/FTA_Identity_Final508.pdf
12. Financial Crimes Enforcement Network. (2024b, November 13). FinCEN issues alert on fraud schemes involving deepfake media targeting financial institutions. U.S. Department of the Treasury. https://www.fincen.gov/news/news-releases/fincen-issues-alert-fraud-schemes-involving-deepfake-media-targeting-financial

13. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In Proceedings of the 34th International Conference on Machine Learning (pp. 1321-1330). PMLR. https://proceedings.mlr.press/v70/guo17a.html
14. Himel, H. U., Mustafizur, M., Akter, T., Anzum, R., Tanni, Z. A., & Hossain, A. (2026). Privacy-aware data protection models for cloud environments. In *2026 6th International Conference on Intelligent Technologies (CONIT)* (pp. 1-10). IEEE. https://doi.org/10.1109/CONIT69683.2026.11621371
15. Huang, B., Wang, Z., Yang, J., Ai, J., Zou, Q., Wang, Q., & Ye, D. (2023). Implicit identity driven deepfake face swapping detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4490-4499). https://doi.org/10.1109/CVPR52729.2023.00436
16. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, *100*, 234-245. https://doi.org/10.1016/j.eswa.2018.01.037
17. Kesavan, S., & Srinivasan, S. (2024). A trusted execution environment-based data aggregation system that preserves privacy for fog/cloud-enhanced IoT applications. *Journal of Electrical Systems*, *20*(3), 6473-6479. https://journal.esrgroups.org/jes/article/view/6945
18. Kesavan, S., Srinivasan, S., & Vijayaraj, R. (2026). Ensuring security and privacy in data aggregation using hybrid masked encryption in cloud computing. *International Journal of Drug Delivery Technology*, *16*(18s), 348-356. https://doi.org/10.25258/ijddt.16.18s.37
19. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, *37*(3), 50-60. https://doi.org/10.1109/MSP.2020.2975749
20. Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A large-scale challenging dataset for DeepFake forensics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3207-3216). https://openaccess.thecvf.com/content_CVPR_2020/html/Li_Celeb-DF_A_Large-Scale_Challenging_Dataset_for_DeepFake_Forensics_CVPR_2020_paper.html
21. Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2980-2988). https://doi.org/10.1109/ICCV.2017.324
22. Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys*, *54*(1), Article 7, 1-41. https://doi.org/10.1145/3425780
23. Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, *115*, 619-640. https://doi.org/10.1016/j.future.2020.10.007
24. Mudaliyar, K. S., & Kumar, S. S. (2024). Intelligent cybersecurity frameworks for data protection in cloud-integrated management information systems. *The Eastasouth Journal of Information System and Computer Science*, *2*(2), 305-322. https://doi.org/10.58812/esiscs.v2i02.1165
25. National Institute of Standards and Technology. (2025). Digital identity guidelines: Enrollment and identity proofing requirements (NIST Special Publication 800-63A-4). U.S. Department of Commerce. https://doi.org/10.6028/NIST.SP.800-63A-4
26. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, *50*(3), 559-569. https://doi.org/10.1016/j.dss.2010.08.006
27. Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q.-V., & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, *223*, Article 103525. https://doi.org/10.1016/j.cviu.2022.103525
28. Raihan, M., Adnan, M., Hossain, M. J., Siddiqa, K. B., Karim, F., & Mohonta, S. C. (2026). Adversarial robustness mechanism for safeguarding biometric verification across mobile financial applications. In *2025 International Conference on Electrical Engineering and Informatics (ICEEI)* (pp. 1-7). IEEE. https://doi.org/10.1109/ICEEI68459.2025.11330502
29. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1-9). https://doi.org/10.1109/ICCV.2019.00009
30. Roy, A., Hossain, E., Himel, H. U., Imam, H., Rashed, R. A. M., & Kaur, H. (2026). Privacy-preserving MIS using blockchain, federated learning, and zero-trust models. In A. E. Hassanien, S. Anand, A. Jaiswal, & P. Kumar (Eds.), *Innovative computing and communications* (Lecture Notes in Networks and Systems, Vol. 2037). Springer. https://doi.org/10.1007/978-3-032-30008-9_16
31. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*, 206-215. https://doi.org/10.1038/s42256-019-0048-x
32. Rudin, C. (2022). Why black box machine learning should be avoided for high-stakes decisions, in brief. *Nature Reviews Methods Primers*, *2*, Article 81. https://doi.org/10.1038/s43586-022-00172-0
33. Ryman-Tubb, N. F., Krause, P., & d'Avila Garcez, A. S. (2018). How artificial intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. *Engineering Applications of Artificial Intelligence*, *76*, 130-157. https://doi.org/10.1016/j.engappai.2018.07.008

34. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, *64*, 131-148. https://doi.org/10.1016/j.inffus.2020.06.014
35. Tuhin, M. K., Bin Ansar, M. T., Sikder, M. S., Himel, H. U., Kaur, H., Imam, H., & Shan-A-Alahi, A. (2024). Federated machine learning for privacy-preserving cyber threat intelligence in global cloud-integrated MIS platforms. *The Eastasouth Journal of Information System and Computer Science*, *2*(2), 284-293. https://doi.org/10.58812/esiscs.v2i02.1123
36. Tuhin, M. K., Kaur, H., Imam, H., Hossain, A., Shan-A-Alahi, A., & Chakraborty, P. (2026). AI-augmented security analytics for real-time threat identification. In *2026 6th International Conference on Intelligent Technologies (CONIT)* (pp. 1-9). IEEE. https://doi.org/10.1109/CONIT69683.2026.11621644
37. Verdoliva, L. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*, *14*(5), 910-932. https://doi.org/10.1109/JSTSP.2020.3002101
38. West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security*, *57*, 47-66. https://doi.org/10.1016/j.cose.2015.09.005
39. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, *10*(2), Article 12, 1-19. https://doi.org/10.1145/3298981
40. Yu, Z., Qin, Y., Li, X., Zhao, C., Lei, Z., & Zhao, G. (2023). Deep learning for face anti-spoofing: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(5), 5609-5631. https://doi.org/10.1109/TPAMI.2022.3215850
41. Zhou, S., Liu, C., Ye, D., Zhu, T., Zhou, W., & Yu, P. S. (2023). Adversarial attacks and defenses in deep learning: From a perspective of cybersecurity. *ACM Computing Surveys*, *55*(8), Article 163, 1-39. https://doi.org/10.1145/3547330