Guidelines on Mitigating Deepfake Risks in Digital Authentication of Customer Identity

SEP 2026

Fintech Association for Consumer Empowerment (FACE)
RBI-recognised Self-regulatory Organisation in the FinTech Sector (SRO-FT)

Contents

Context

1. FinTech-enabled digital financial services conduct digital authentication of a customers' identities through a combination of measures, such as:
   a. Aadhaar OTP-based identity verification
   b. PAN validation and CKYC verification
   c. Liveness checks and anti-spoofing measures
   d. Geo-tagging and audit trails
   e. Device and IP monitoring
   f. V-CIP-based live customer interaction
   g. Enhanced due diligence for remote onboarding
   h. Transaction monitoring and AML controls

2. However, Artificial Intelligence (AI) enabled synthetic media tools and deepfakes, as below, can bypass and penetrate the above safeguards, altering the threat landscape by fabricating identities, biometric signals, and trust.
   a. AI-generated facial movements and biometric simulations
   b. Synthetic "live" onboarding videos
   c. Deepfake-assisted selfie verification
   d. Voice cloning tools
   e. Emulator-driven onboarding environments
   f. AI-Powered social engineering
   g. Manipulated identity artefacts and synthetic identity creation using fragmented information
   h. Organised mule-account creation networks.

3. These create systemic risk for digital authentication by:
   a. Compromising OTP-based authentication through SIM swap or mobile number manipulation (that may be linked to Aadhaar).
   b. Defeating basic selfie or static liveness using AI-generated facial simulations
   c. Industrialising synthetic onboarding at scale using automated AI tooling
   d. Proliferation across companies

4. Clearly, existing authentication measures no longer assure against AI-enabled impersonation attacks. Hence, there is a pressing need to step up defences. In this context, we guide members to adopt the following measures.

5. We have released these guidelines as per our Standards Setting Policy[1] with input from members and the FACE Standards Committee, and with Board approval on 1 Sep 2026.

[1] https://faceofindia.org/wp-content/uploads/2025/06/FACE-Standards-Policy.pdf

6. The guidelines complement existing RBI directions such as KYC, Digital Lending, Cyber Security and government laws and guidance rather than creating parallel obligations and regulations, and laws always supersede in case of conflict.

Guidelines

7. **Differentiated controls** for 3 common attack vectors as under:

Attack Vectors	**Real-Time Face Swap**	**Camera Injection (Virtual Driver)**	**Voice Cloning and Vishing**
Technical Mechanism	Bypass traditional liveness checks during digital onboarding by replacing facial movements and expressions with deepfake-generated content in real time, making them a critical threat to KYC processes.	Use software to feed synthetic video streams directly into a device's camera input, enabling fraudsters to present entirely AI-generated identities while appearing legitimate to verification systems.	Leverage AI-generated voice replication to impersonate customers, employees, or executives, increasing the effectiveness of social engineering and account takeover attempts, with documented incidents reported in India.
Threat Levels	CRITICAL: operationally active against digital onboarding globally	CRITICAL: available as a commoditised service for minimal cost	HIGH: documented incidents in India
Common Controls	▪ Challenge-response based liveness verification (random facial movements, head turns, spoken phrases). ▪ AI-based deepfake and facial artefact detection. ▪ Lip-sync and facial movement consistency analysis. ▪ Gaze tracking and eye movement verification. ▪ Background consistency monitoring during onboarding sessions. ▪ Risk-based escalation to manual review where synthetic artefacts are suspected.	▪ Detection and blocking of virtual cameras and camera driver manipulation. ▪ Device integrity and attestation checks. ▪ Emulator and virtual-device detection. ▪ Screen-sharing and remote desktop detection. ▪ Secure camera SDKs capable of verifying hardware camera access. ▪ Device fingerprinting and device-binding to identify repeated abuse patterns.	▪ Out-of-band verification for high-risk requests and sensitive account changes. ▪ Multi-factor authentication beyond voice-based verification. ▪ AI-based voice biometrics and synthetic speech detection. ▪ Transaction intent confirmation through independent channels. ▪ Callback verification procedures for high-value transactions. ▪ Employee awareness and targeted training on AI-enabled social engineering attacks.

8. **Proportionate framework**
    a. Adopt a proportionate framework factoring regulatory requirements, risk[2], technical maturity, and costs to align proportionate controls and measures[3].
    b. Categorise controls into baseline, enhanced, and advanced measures to support phased and practical implementation.
    c. Align controls with regulations by mapping their complementarity.

9. **Controls layering**
    a. Do not rely on a single control to mitigate AI-enabled identity fraud and adopt a layered defence model combining identity assurance, device assurance, behavioural analytics, transaction monitoring, and continuous risk assessment.
    b. Periodically reassess the effectiveness of controls against emerging deepfake techniques and adversarial AI capabilities.

10. **Standard evaluation criteria for deepfake detection solutions** for detection accuracy, Presentation Attack Detection (PAD), India-specific testing, and performance benchmarks, to support procurement, implementation, and compliance.

11. **Identity assurance controls** to evolve from static validation to authenticity verification, i.e., shift from merely verifying a facial image toward validating genuine human presence and behavioural authenticity, by using:
    a. Minimum liveness standards to resist presentation and injection attacks
    b. Multi-modal identity verification
    c. Advanced response-based liveness verification
    d. Deepfake detection capabilities
    e. Graph-based synthetic identity detection
    f. Randomised facial and voice prompts
    g. AI-based deepfake and synthetic artefact detection
    h. Lip-sync mismatch detection
    i. Gaze tracking and behavioural biometrics
    j. Background consistency analysis during V-CIP sessions
    k. Document forensics
    l. Human-in-the-loop risk models
    m. Dynamic risk scoring
    n. Step-up authentication for suspicious onboarding patterns
    o. Stronger controls around mobile number changes/modifications
    p. Step-up verification (including re-KYC or video-based verification) for high-risk profile changes[4], using a risk-tiering matrix defined by the regulator.
    q. Cooling-off periods (minimum 24–72 hours, risk-calibrated) before critical changes such as mobile number, bank account, or Aadhaar updates (if available) become effective.
    r. Enhance monitoring and transaction restrictions during the cooling-off window

[2] For product, customer segment, onboarding channel, transaction size
[3] Certain advanced controls are emerging capabilities and may be adopted progressively based on technical feasibility, operational maturity, and cost considerations
[4] mobile numbers, email addresses, Aadhaar linkages and bank accounts

12. **Device and telecom assurance controls** to extend identity verification beyond documents and biometrics using:
    a. Device integrity attestation such as device fingerprinting, SIM binding/age/swap recency/location, mobile number verification, call forwarding, roaming status
    b. Trusted-device validation
    c. Real-time customer alerts for onboarding attempts
    d. Cross-platform identity triangulation
    e. Real-time emulator and virtual-device detection
    f. Remote desktop and screen-sharing detection
    g. Geolocation consistency monitoring
    h. Abnormal device behaviour analytics

13. **Behavioural and transaction assurance controls** for identity assurance after onboarding. Many AI-enabled frauds become visible only through behavioural and transactional patterns.
    a. Behavioural biometrics with AI anomaly detection
    b. Contextual multi-factor authentication
    c. Rapid fund-flow detection
    d. Device-network linkage analysis
    e. Beneficiary concentration monitoring
    f. Dormant-to-active monitoring
    g. Out-of-band verification for high-risk activities
    h. Transaction intent verification mechanisms
    i. Mandatory secondary verification for high-risk onboarding events

14. **Periodic Testing and Resilience** using established testing methodologies, benchmark datasets, performance metrics, and minimum effectiveness thresholds for deepfake detection solutions.
    a. Cryptographic provenance
    b. Deepfake red teaming workflows
    c. Deepfake injection audits
    d. Bias and degradation testing
    e. Adversarial perturbation testing
    f. Multimodal fusion
    g. SDK Safeguards
    h. API Security and cryptographic proof
    i. Human-in-the-Loop (HITL)
    j. Zero-trust architectures
    k. Intercepting & deconstructing API Traffic
    l. Payload analysis
    m. BCP, in case a vulnerability is detected during live deployment.
    n. Periodic deepfake penetration testing (at minimum annually or following material changes to KYC systems) with scope covering at least presentation attacks, digital injection attacks, voice synthesis, and document forgery.
    o. Independent audit of deepfake controls as part of existing ICT/cyber audit frameworks

15. **Governance** to assess resilience against emerging threats by:
    a. Devising and institutionalising responsibility and ownership across the company
    b. Ensuring lawful and privacy-compliant access to data by evaluating permissible data sources, consent requirements, privacy obligations, and regulatory considerations.
    a. AI-related fraud risk tiering/classification based on materiality, complexity, impact & reversibility, and data sensitivity (e.g. High, Moderate, Low, Unacceptable)
    b. Strong AI service provider evaluation criteria based on technical performance, accuracy & reliability, explainability, latency & throughput, customisation, scalability, data privacy, model governance/training/validation, infrastructure compatibility, financial stability, pricing transparency, bias mitigation, etc.
    c. Strengthen vendor governance through due diligence, model validation, security assessments, performance monitoring, contractual safeguards, and audit rights.
    d. Independent third-party assessments/audits.
    e. Workforce training and awareness.

16. **Intelligence collaboration** to leverage industry and regulatory fraud intelligence-sharing mechanisms, risk-signalling/fraud repositories, deepfake testing frameworks, and shared learning on emerging fraud typologies.

17. **Dedicated reporting for AI-enabled identity fraud**[5] to classify and report such incidents to the relevant regulators and authorities. Define minimum required data fields for reporting such as attack vector, deepfake modality used (video, audio, document), detection outcome, and financial impact.

18. **Customer protection** through awareness initiatives on deepfake and identity fraud risks, privacy-compliant access to verified telecom, device, and identity intelligence signals, multiple authentication and cooling-off for sensitive identity updates/high-risk events, real-time and effective support and dispute resolution for victims of synthetic identity fraud and balancing customer experience and inclusion to avoid excessive onboarding friction or false declines for genuine customers.

References

19. I4C-MHA Advisory on AI-driven authentication bypass | targeting financial infrastructure[6]
20. CERT-in Advisory: Deepfakes, threats and countermeasures[7]
21. Government Strengthens Regulatory Framework to Address AI-Generated Deepfakes[8]
22. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021[9]

[5] Such standardised reporting would enhance visibility into emerging fraud typologies, facilitate regulatory intelligence gathering, support evidence-based policy interventions, and enable the sharing of aggregated industry insights to strengthen ecosystem-wide resilience against evolving AI-driven threats

[6] https://i4c.mha.gov.in/theme/resources/advisories/ADVISORY%20TAU-ADV-016-3.pdf

[7] https://www.cert-in.org.in/s2cMainServlet?pageid=PUBVLNOTES02&VLCODE=CIAD-2024-0060

[8] https://www.pib.gov.in/PressReleasePage.aspx?PRID=2295500&reg=3&lang=1

[9] https://www.meity.gov.in/static/uploads/2026/02/550681ab908f8afb135b0ad42816a1c9.pdf

Fintech Association for Consumer Empowerment (FACE) is an RBI-recognised Self-Regulatory Organisation in the FinTech sector (SRO-FT). FinTech companies of all kinds come together at FACE to build an industry that enables borrower-centric financial services that are safe, suitable, and transparent, delivering positive impacts on society and the economy.