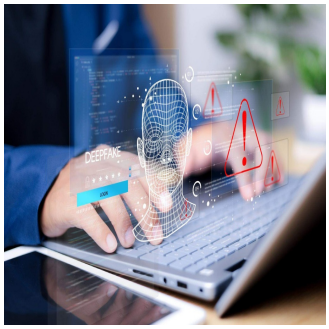
HealthManagement
Promoting Management and Leadership

Deepfakes Expose Gaps in Healthcare Security



Artificial intelligence is enabling convincing audio, video and image fabrications that are moving beyond social platforms into clinical systems, telemedicine and patient communications. These manipulations do not rely on malicious code that can be scanned or quarantined, they exploit human trust at the point of care. The operational and governance implications are wide-ranging, from the integrity of diagnostic artefacts to the authenticity of clinician identities and public health messaging. With widely available tools and low barriers to entry, healthcare is already a target. Security controls now must verify what people see and hear in the flow of care, not only protect networks and devices behind the scenes.

Clinical Artefacts, Trusted Voices and a Vulnerable Surface

The impact of synthetic media spans diagnosis, access and public confidence. A falsified medical image can misguide clinical decisions. A cloned clinician voice can be used to unlock access to sensitive systems or to elicit disclosure of credentials. A fabricated video of a public health official can circulate rapidly and seed misinformation at scale. Training on its own has limits in the face of well-crafted deception. Many employees completed cybersecurity awareness training, yet many still failed phishing simulations, underscoring how real-world tactics can overwhelm vigilance when delivered at scale. As healthcare digitises, deception will arrive through realistic synthetic voices, manipulated scans and fabricated video consultations that look and sound familiar.

Healthcare's exposure is amplified by the nature of its work. A falsified CT scan could prompt unnecessary intervention. A cloned doctor's voice captured from a webinar or press briefing can be sufficient to generate a convincing audio command capable of issuing fraudulent orders in a clinical setting. Recent research has shown that generative adversarial networks can insert or remove disease indicators in CT images, deceiving both radiologists and machine-learning diagnostic tools. Analyses also note that while synthetic data can enrich AI training sets, it simultaneously opens routes for fraud and ethical misuse. Trust, long treated as the bedrock of care, becomes fragile when core artefacts and authoritative voices can be forged from minimal source material.

Must Read:** **AI Honeypots Strengthen Healthcare Cyber Defences

Limits of Today's Defences and Detection Progress

Conventional cyber controls remain necessary but are not sufficient against synthetic media. Identity Threat Detection and Response (ITDR), endpoint protection and multi-factor authentication (MFA) address malware, credential abuse and device compromise. They are not designed to detect a fabricated face on a telemedicine call or an altered MRI file stored in an imaging system. These tools operate at system and network layers, while deepfakes exploit perception, context and the authority of familiar voices and images.

Detection research is moving quickly, though the challenge is significant. Approaches that use visual prompt tuning with pre-trained models aim to adapt as manipulation techniques evolve and to improve robustness across varied datasets. Ensembles that combine multiple detection algorithms seek higher accuracy by fusing complementary signals. Beyond algorithmic screening, cryptographic provenance, digital watermarking and blockchain-based signing of medical records and images are gaining traction to ensure that clinical artefacts have not been tampered with. Across these avenues, a consistent requirement emerges: verification needs to happen in real time and within clinical workflows rather than after an incident is reported.

Leadership, Governance and Workflow Integration

The response demanded of healthcare leaders spans architecture, process and oversight. Security designs should extend beyond perimeter and

identity controls to embed deepfake detection within electronic health record (EHR) systems, imaging platforms and telemedicine tools. Incident response playbooks need to cover scenarios where a scan, a recorded order or a live consultation is fraudulent, with clear steps for containment, secondary verification and communication. Staff education should evolve from generic awareness to structured verification protocols for unexpected requests, including those that appear to come from trusted colleagues, so that awareness is paired with practical procedures that reduce blind trust.

Regulatory frameworks have not yet caught up with synthetic media risks. Existing regimes focus on privacy and device integrity rather than media authenticity. Organisations that move early by implementing provenance checks and real-time validation can reduce risk and help shape emerging standards. Collaboration across hospitals, insurers and regulators is also necessary, as no single entity can address the breadth of the threat surface alone. With attackers already deploying voice clones and synthetic videos to defraud other sectors, healthcare's combination of sensitive data, dispersed endpoints and high-trust interactions makes preparedness an immediate priority.

Verifying authenticity is becoming as critical as blocking malicious code. Deepfakes threaten to distort diagnostics, impersonate clinicians and undermine public confidence, revealing gaps that traditional controls were not built to close. Embedding media validation into EHRs, imaging systems and telemedicine, expanding incident response to cover deception scenarios and strengthening verification protocols offer a pragmatic path forward. With regulation still evolving, early, coordinated action positions organisations to protect patients, safeguard data and preserve the trust on which care depends.

Source: **Healthcare IT Today**

Image Credit: **iStock**

Published on : Tue, 18 Nov 2025