Pindrop

# The 2026 AI Threat Landscape in Healthcare

How deepfakes and synthetic actors are reshaping risk across patient, provider, and workforce interactions

INTRODUCTION

# AI has a new role in healthcare

Over the past year, AI capabilities have advanced faster than most healthcare security and identity controls. In early 2026, leading AI labs released models capable of operating autonomously across workflows like patient access, billing support, IT helpdesk operations, and administrative systems—executing complex, domain-specific tasks with minimal instruction.

Generative AI fraud is projected to reach $40B in the United States by 2027, up from $12B in 2023,[1] with AI agents expected to reduce the time it takes for threat actors to exploit exposed accounts by 50%.[2] In healthcare, this acceleration directly impacts patient access, benefits, billing, and IT support workflows—where identity verification is often the gatekeeper to PHI, financial data, and clinical systems.

Healthcare organizations are recalibrating to a new reality: AI can now replicate specialized human workflows at scale—from patient calls and provider interactions to internal support requests. This matters for security because the same autonomy driving operational efficiency is also driving new, sophisticated, and hard-to-detect fraud across high-trust healthcare interactions.

**The key question is:**

***Can we still verify who—or what—is operating inside our systems?***

**1,210%**

YoY surge in AI-enabled fraud in 2025.[3]

**$40B**

Projected U.S. generative AI fraud by 2027 (from $12B in 2023).[1]

**~50%**

Human detection rate of AI-generated content — no better than a coin flip.[4]

[1] Deloitte, "Generative AI is expected to magnify the risk of deepfakes and other fraud in banking," May 2024. https://www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html

[2] Gartner: https://www.gartner.com/en/newsroom/press-releases/2025-03-18-gartner-predicts-ai-agents-will-reduce-the-time-it-takes-to-exploit-account-exposures-by-50-percent-by-2027#:~:text=AI%20Agents%20Will%20Increasingly%20Exploit,Khan%2C%20VP%20Analyst%20at%20Gartner

[3] Pindrop analysis of AI fraud data from January-December 2025

[4] Cooke, D., Abigail Edwards, Sophia Barkoff, and Kathryn Kelly. "As Good As A Coin Toss: Human detection of AI-generated images, videos, audio, and audiovisual stimuli." April 2025. https://doi.org/10.48550/arXiv.2403.16760

PART 1: THE NEW AI THREAT LANDSCAPE

# AI Automation is outpacing enterprise controls.

Attackers have shifted from one-off scams to AI-backed operations designed for scale. For example, synthetic voice, video manipulation, automated social engineering, and autonomous agents are reducing friction for fraudsters while increasing oversight for defenders.

Over the past year, AI capabilities have advanced faster than most enterprise controls. This matters for security, because the same autonomy driving productivity is now driving fraud.

In 2026, the shift is no longer on attackers "using AI," but instead using AI-native operations. These campaigns assume automation at scale, continuous model refinement, synthetic personas, and real-time impersonation. Once trained, these models operate without fatigue, iterating on scripts and adapting to resistance at an exponential rate.

> "For thousands of years, trust was based on perception, recognizing someone's voice or seeing someone face-to-face. AI has changed what we think of as proof."
>
> **Head of Fraud, Financial Crimes, and Trust Systems, HealthEquity**
> Ajit Gaddam
>
> HealthEquity

## 4 structural advantages attackers have in healthcare today

### 1. Automation at scale

AI agents do not fatigue. Once trained, they operate continuously across patient access lines, benefits IVRs, pharmacy workflows, and IT helpdesks, 24 hours a day.

### 2. Adaptive social engineering

Models refine scripts based on responses. They identify weak authentication flows in patient, provider, and member interactions and exploit them at scale, adjusting in real time to resistance.

### 3. Media realism

Synthetic voice and video no longer appear crude. Humans detect AI-generated content correctly only about 50% of the time.[4] When urgency, authority, or familiarity are introduced, detection accuracy drops further. *Human judgment is no longer a reliable control.*

### 4. Amplification

Small authentication weaknesses, when exploited across thousands of patient, provider, or member interactions, become systemic vulnerabilities. What was once a niche attack surface is now infrastructure-level risk across access centers, billing, and support operations.

**Security leaders can no longer treat deepfakes as novelty incidents. They are now infrastructure-level threats.**

# Three AI-enabled attack scenarios to watch

## Hiring fraud

Bad actors use AI to spin up fake identities en masse—then show up to interviews with polished resumes and manipulated video or audio. When fake employees gain system access, they become insider threats across environments that manage patient data, financial accounts, and clinical operations.

**REAL-LIFE STORY**

An investigation of Pindrop's hiring pipeline revealed startling patterns of fraud, including 1 in 6 applicants showing signs of fraud and 1 in 343 linked to DPRK-affiliated infrastructure.[5]

## Reconnaissance and Account Takeover

Attackers use AI-driven bots and synthetic callers to target healthcare contact centers, probing IVR systems and agent workflows to gather sensitive information and exploit authentication gaps. By harvesting or validating data like Social Security numbers, dates of birth, and account details, they gain access to patient accounts—leading to PHI exposure, benefits fraud, and financial account takeover at scale.

**REAL-LIFE STORY**

A major U.S. healthcare provider uncovered that bot attacks accounted for more than half of all fraud in their systems, with over 15,000 unique bot fraud calls observed since summer 2025. These bots systematically probed IVRs to extract data and support account takeover and financial fraud schemes.[6]

## Helpdesk Impersonation and System Access

Fraudsters impersonate providers, clinicians, or internal staff using synthetic voice and stolen credentials to manipulate IT helpdesks into resetting passwords, bypassing MFA, or granting access. Because these interactions are trusted and often urgent, attackers can gain entry into internal systems —creating pathways to data breaches, ransomware, and broader enterprise compromise.

**REAL-LIFE STORY**

A group of attackers known as Scattered Spider used a convincing voice-phishing call to trick MGM Resorts International's IT helpdesk into resetting Okta credentials, handing them access that led straight to data theft.[7]

[5] Pindrop, "From Interview to Intel Drop: The Moment We Exposed a Coordinated Hiring Scheme," July 2025, https://www.pindrop.com/article/deepfake-detection-revealed-coordinated-hiring-scheme/

[6] Anonymous Pindrop healthcare data collected in 2025

[7] Pindrop, "How Pindrop Technology Could've Prevented the MGM Breach," September 2023. https://www.pindrop.com/article/pindrops-technology-could-have-prevented-mgm-breach/

PART 2: AI ATTACKS IN HEALTHCARE-HOW THEY ACTUALLY WORK

# AI Automation is outpacing enterprise controls.

Healthcare is facing a perfect storm. The controls that once kept attacks manageable are failing at the exact moment scams are becoming faster, cheaper, and harder to detect. Legacy security checks are no longer a meaningful barrier when stolen personal data is widely available. Nearly 60% of organizations[8] using compromised personally identifiable information (PII) to bypass knowledge-based authentication.

At the same time, generative AI has changed the threat landscape. Deepfake attacks increased by 880% in 2024[9]—no longer a theoretical risk, but one showing up at scale in real accounts with real losses.

Regulatory pressure is also increasing. The largest healthcare fraud takedown in U.S. history charged 324 defendants tied to $14.6 billion in intended losses,[10] signaling a shift toward more aggressive enforcement.

**Nearly 60%**

of organizations using compromised personally identifiable information (PII) to bypass knowledge-based authentication[8]

**880% increase**

in deepfake attacks in 2024[9]

## How attacks actually work in healthcare

These bots systematically probe IVR systems for reconnaissance, gathering and validating data such as Social Security numbers, dates of birth, balances, and transaction histories. This information is then used to carry out social engineering with live agents, take over accounts, and access HSA, FSA, and other employer-funded savings accounts.

In one case, over 15,000 unique bot fraud calls were observed in a matter of months, indicating attackers are turning this into a repeatable, scalable operation—often without ever speaking to a live agent.

These interactions are increasingly difficult to detect. Researchers observed "programming-style" commands and near-human interaction speeds, with signals suggesting coordinated, call center-style fraud operations executing attacks at scale.

[8] *"TransUnion 2025 State of Omnichannel Fraud Report Insights," May 2025, https://www.hypr.com/blog/transunion-2025-state-of-omnichannel-fraud-report-insights*

[9] *Pindrop, "Voice Intelligence and Security Report," June 2025, https://www.pindrop.com/research/report/voice-intelligence-security-report/*

[10] *Department of Health Human Services, "2025 National Health Care Fraud Takedown," https://oig.hhs.gov/newsroom/media-materials/2025-national-health-care-fraud-takedown/*

## Business impact: where this shows up

AI-driven scams create immediate and measurable business impact.

Financial loss is the most direct effect. Compromised accounts—especially high-balance accounts such as HSAs and FSAs—can lead to significant exposure, with one healthcare provider facing more than $40 million in account risk tied to AI-driven bot activity.[3]

Trust is also impacted. Healthcare organizations are trusted with highly sensitive personal and financial data. When accounts are compromised, confidence in the organization can drop significantly. Even a small number of public failures can damage brand reputation, and rebuilding trust requires additional time and investment.

Operational strain is another consequence. AI-powered schemes increase the volume of suspicious interactions, which can waste agent time and increase the workload for fraud and investigation teams. The result is longer handle times, overextended teams, and slower service for legitimate patients and members.

At the same time, organizations are proving that these risks can be reduced. HealthEquity, a leading HSA administrator, reported over a 90% reduction in voice-channel fraud after strengthening identity verification—while maintaining a strong member experience.

AI-driven attacks in healthcare do not stop at the point of interaction. The same techniques used to probe contact centers, bypass authentication, and take over accounts are also being used to gain persistent access inside the organization.

One of the fastest-growing and least understood attack surfaces is hiring. As organizations expand remote work and digital hiring processes, attackers are using synthetic identities and AI-generated personas to bypass screening, secure employment, and operate inside the organization with legitimate access.

WEBINAR

### When Identity Can't Be Trusted: Defending Healthcare Against AI Attacks, Impersonation and Deepfakes

Hear from Jim Routh, Chairman of the Board for Health-ISAC and Pindrop Customer Ajit Gaddam Head of Fraud, Financial Crimes, and Trust Systems at HealthEquity.

PART 3: HIRING AS AN ATTACK VECTOR

# Pindrop research: Inside our own hiring pipeline

*This section draws on internal data and one recruiter's firsthand experience uncovering fraud in Pindrop's hiring pipeline over the past year.*

While remote work expanded career opportunities, it also expanded the attack surface. To understand how deeply AI-native fraud has penetrated into hiring, we analyzed confirmed fraudulent applicants for Pindrop engineering roles, and cross-referenced our findings with GitLab's recent Threat Intelligence Report on North Korean tradecraft.

GitLab's Report documents how DPRK-aligned actors use "Contagious Interview" and fake IT worker schemes to spread malware and generate revenue.

We did a similar analysis, and what we found in our own hiring pipeline was not anecdotal.

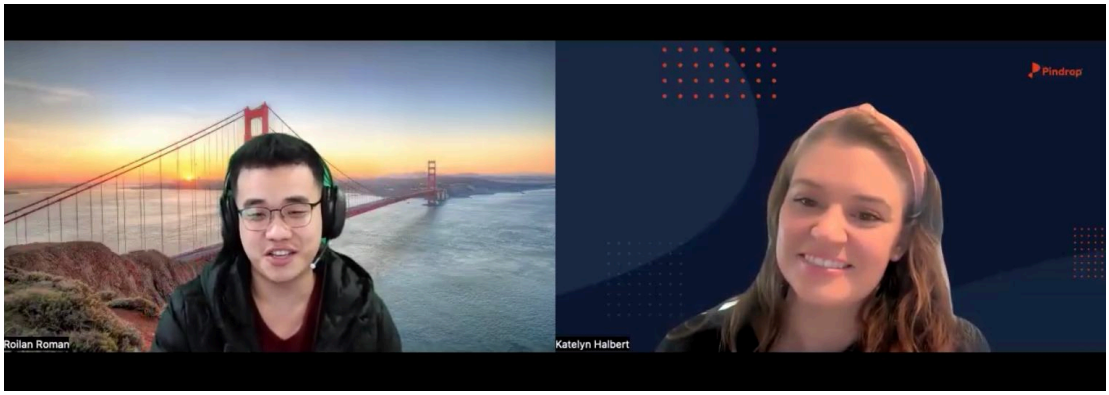
We conducted a focused review of the applicant pipeline for a single role. As part of this review, we analyzed recurring fraud indicators observed over the past 12 months, including email aliases containing "dev," "tech," or "work," newly created LinkedIn profiles, missing profile photos, low connection counts, and VoIP phone numbers.

## Key findings from Pindrop's hiring pipeline

**1 in 6**
applicants in pipeline show signs of fraud

**1 in 343**
applications are linked to DPRK-affiliated activity

**1 in 4**
North Korean linked applicants used a deepfake during a live interview



THE DEEPFAKE FILES

# Meet our candidates

## Meet “Shamar”

When we first met “Shamar,” he had applied for a backend software engineering role. On the surface, he was a strong candidate.

The quality of the deepfake was high. His lip movements were synchronized, facial expressions natural. His English was fluent. The pacing and length of his answers felt appropriate and conversational. There were no obvious visual distortions or awkward delays. Nothing about the interaction felt artificial.

We decided to move him forward for a second interview with our product manager. He used the same deepfake setup. He spoke confidently about his technical background and handled follow-up questions with reasonable depth. Whether supported by AI tools or careful preparation, he knew how to make his responses sound authentic.

It was only after digging further into his online identity that inconsistencies began to surface. His LinkedIn profile was verified through CLEAR using a Jamaican ID, which added a layer of legitimacy. However, IP analysis told a different story.

The activity was tied to infrastructure inconsistent with his claimed location, including connections associated with Astrill VPN and endpoints linked to Eastern Europe. The geographic signals did not align.

Without an alert inside our Pulse for Meetings tool, this candidate likely would have advanced through the interview process unchecked. There were no clear visual artifacts or behavioral inconsistencies that would have raised suspicion.

> What made this case notable was that it lacked any glaring flaw; that’s how convincing it was.

## Meet “Connor”

When we first interviewed Connor, nothing immediately raised concern. He was wearing glasses and a headset, details that years ago were often cited as ways to confirm it wasn’t a deepfake. That assumption no longer holds.

He communicated clearly and confidently. He claimed to be based in Kentucky, and spoke comfortably about his environment. The interaction felt natural.

Then, Pulse for Meetings generated an alert. The alert immediately shifted our observations. What stood out was not a visible flaw, but how polished the presentation was.

There were no distortions, no lag, no behavioral irregularities. We advanced Connor to the next stage and required a recorded video explanation as part of a technical assessment.

When we reviewed that submission, the continuity became clear. The same glasses and headset. The same facial structure and voice. But subtle rendering differences across sessions confirmed what Pulse for Meetings had already identified.

> **This was not a one-time manipulation. It was a coordinated synthetic identity engineered to persist through multiple stages of evaluation.**

## Inside the attack infrastructure

Across confirmed fraudulent applicants, patterns were consistent and repeatable.

Device telemetry revealed that environments were highly standardized. In fraud cases, systems consistently reported the same operating system (Windows NT 10.0), Chrome-only browser usage, and nonstandard screen resolutions—patterns that differ significantly from the diverse device profiles seen in legitimate applicants. No confirmed fraud cases originated from current-generation operating systems.

These signals suggest the use of centralized, controlled environments—such as virtual machines or containers—rather than organic user devices.

Network and geographic data showed similar patterns. A significant portion of fraudulent sessions originated from U.S.-based IP addresses, but session behavior indicated proxy usage, VPN masking, and possible handoffs between operators. In one sample, 76.6% of sessions showed VPN usage, and applicants frequently used different IP addresses across interview stages.

These patterns are consistent with coordinated infrastructure rather than independent individuals.At the identity level, attackers are constructing synthetic personas end-to-end. This includes generating profile images, creating supporting documentation, building professional profiles, and establishing communication channels. In some cases, fabricated identities have successfully passed document verification processes.

Individually, these signals may not be conclusive. Together, they form a clear pattern: centralized, repeatable systems designed to simulate legitimate candidates at scale.

In healthcare, this means unauthorized individuals can gain access to systems that support patient care, financial operations, and clinical workflows—turning hiring into a direct pathway to enterprise compromise.

In healthcare, this means unauthorized individuals can gain access to systems that support patient care, financial operations, and clinical workflows—turning hiring into a direct pathway to enterprise compromise.

PART 4: DEEPFAKE DETECTION

# Why deepfake detection is only the foundation

When Pindrop® Pulse was first developed, the focus was on detecting deepfake audio. Since then, impersonation has evolved beyond audio to include video, synthetic identities, and fully AI-driven interactions.

Deepfake detection alone is no longer sufficient.

In an AI-driven threat landscape, enterprises must move beyond single-signal detection to a multi-layered identity verification model—one that correlates media integrity, behavioral signals, device telemetry, and network intelligence.

Legacy approaches—knowledge-based authentication, one-time passcodes, static credentials, and human judgment—were designed for human actors. They do not hold in an environment where adversaries can generate, adapt, and operate at machine scale.

**Identity verification in the AI era must answer 3 questions**

### Is this a real human?

Continuous liveness detection across audio and video—critical in environments like contact centers and telehealth.

### Is this the right human?

Analysis confirming the participant is not just human, but the correct patient, provider, or employee.

### Are they in the right location?

Device and network intelligence cross-referenced with IP or VPN mismatch—helping identify anomalies in both external interactions and internal access.

PART 4: WHAT YOU CAN DO NOW TO DEFEND YOUR ENTERPRISE FROM AI FRAUD

# Three assumptions to revisit

AI fraud is not a 'future-state' risk. It is an operating condition. The organizations that respond effectively will not wait for a catastrophic event; they will adjust controls now, at the identity layer.

## 1. AI-driven attacks are low-scale and opportunistic

This assumption is increasingly challenged by observed activity patterns.

AI-driven automation enables high-volume, repeatable interactions that extend far beyond human-paced attack models. Fraud attempts are now occurring across a meaningful share of interactions—often without immediate indicators of loss.

In healthcare, this shows up as sustained activity across patient access lines, IVRs, and digital workflows—not isolated events. The same scale is now emerging in hiring pipelines, where synthetic candidates can be deployed across multiple roles, and in coordinated attempts to access internal systems through helpdesks.

Controls designed for rare or manual abuse may not provide sufficient coverage when interaction volume and repetition increase materially. What was once episodic is now continuous, and increasingly automated.

## 2. We will detect fraud when monetization begins

Observed activity increasingly emphasizes interaction with systems and workflows prior to downstream misuse, rather than immediate financial impact.

In healthcare environments, this includes IVR probing, account access attempts, and repeated interaction with identity workflows before any visible fraud occurs. Attackers are testing authentication controls, gathering information, and establishing trust early in the interaction lifecycle.

This pattern extends beyond patient interactions. Similar behavior is seen in helpdesk workflows—where attackers request credential resets—and in hiring, where synthetic identities move through multiple stages before gaining system access.

When detection relies on transaction-loss signals, identity trust may already have been granted—shifting detection from preventive to reactive.

## 3. Authentication is an operational control, not a governance concern

Identity decisions made during live interactions are increasingly being evaluated through governance, audit, and risk management lenses.

Knowledge-based authentication was not designed for environments where breached personal data and AI-assisted impersonation are prevalent. Static authentication methods can be undermined when personal data has been compromised.

In healthcare, authentication decisions now determine access across multiple high-risk points: patient accounts, benefits and billing systems, clinical workflows, and internal infrastructure via helpdesks or new hires entering the organization.

A governance-oriented approach to authentication considers not only efficiency and experience, but the defensibility of identity decisions at the moment trust is granted.

**It is no longer only:**

- Were controls in place?
- Did teams respond effectively after an incident?

**The question is now:**

Were real-time identity decisions reasonable and defensible when access was granted—whether to a patient account, a system, or an employee role?

CONCLUSION

# Strategic outlook

The world has changed. AI broke trust, and it did so faster than most enterprises were prepared for. Attacks are now operational, profitable, and increasingly indistinguishable from legitimate interactions.

In 2026 and beyond, the defining question is not whether AI-driven impersonation reaches your organization: it's whether you can recognize it-and if your systems can act at the speed, scale, and sophistication of these scams.

Trust can no longer rely on instinct. It must be engineered. Verification must evolve from static authentication to continuous, multi-layer validation across media, device, network, and behavior.

Organizations that wait for a breach to drive transformation will find the cost far exceeds the investment they initially tried to avoid. The window to build proactive, AI-aware identity infrastructure is now, before the next wave of attacks reset what "prepared" actually means in today's AI-first world.

## That is what the Pindrop® Real Human + Right Human™ Platform was built to do.

Pindrop is the Real Human + Right Human® Identity Trust Platform for the AI era. As AI-driven fraud and deepfakes erode trust in digital communication, Pindrop delivers continuous identity verification and deepfake detection across voice, video, and digital interactions in real time.

Enterprises rely on Pindrop to secure billions of high-risk customer interactions each year, including 7 of the top 10 U.S. banks and leading insurers and healthcare providers. Powered by models trained on more than 1.5 billion real-world interactions annually and protected by 300+ patents, Pindrop restores trust while reducing fraud, lowering operational costs, and improving customer experience.

Recognized by TIME as one of 2025's Best Inventions and by Inc. for Best in Business for Innovation.

www.pindrop.com/industry/healthcare/

Pindrop®