The Year Trust Broke

Inside the 2025 AI Fraud Spike

February 4, 2026

Prepare for 2026 After the AI Attack Spike

AI-enabled attacks are surging because automation, generative tools, and synthetic video and voice let attackers scale faster, cheaper, and more effectively.

- **Who:** Fraudsters use AI to supercharge and scale their schemes across channels.
- **Used for:** Social engineering schemes, financial fraud, identity fraud, and account takeover.
- **Seen most in:** Real-time conversations.

Important details

- AI fraud surged **1210% in 2025**.[1]
- Non-AI fraud is also increasing—**up 195% at the end of 2025**[2]—but it's still outpaced by AI fraud growth, meaning more fraudsters are turning to AI-backed schemes.
- Healthcare is facing a new wave of bot attacks—with one major healthcare provider experiencing over **15,000 unique bot fraud calls in the summer of 2025**[3]—leaving these orgs exposed.
- Return fraud in retail is skyrocketing as fraudsters automate refund scams at scale. Among major retailers and Pindrop customers, there was a **330% increase in AI fraud in 2 months**.[4]

Quick takeaways

- AI dramatically increases fraud speed and volume.
- Many enterprises are unprepared for AI-driven fraud.

Bottom line: In 2025, AI became a magnifying force for fraud, growing faster than both non-AI fraud and the defenses built to stop it.

Next page: AI attacks didn't just evolve. They grew exponentially.

Citations

[1] Pindrop analysis of AI fraud data from January-December 2025
[2] Pindrop analysis of non-AI fraud data from January-December 2025
[3] Anonymous Pindrop healthcare data collected in 2025
[4] Pindrop analysis of aggregrated retail fraud data collected in 2025

CHAPTER ONE

AI attacks didn't just evolve. They grew exponentially.

Attackers scaled their operations and AI is the force behind it.

Attackers shift to AI-backed scams.

In 2025, many fraudsters weren't just running one-time scams; they were deploying scalable, AI-assisted schemes designed for maximum impact with minimal effort. We saw it first in contact centers, where synthetic voices and automated social engineering bypassed traditional security checks in seconds.[1] Now, those tactics are rippling across real-time interactions that rely on trust: from remote job interviews to financial transactions and beyond.[2] For CISOs and CTOs, this isn't just another trend. It's a fundamental shift in the threat landscape and how trust is exploited at scale.

Trust needs a redesign.

In this modern threat landscape, enterprises need to change how they define and defend trust. Legacy security methods often assume identity can be proven through knowledge-based authentication (KBAs), one-time passwords (OTPs), or human judgement. In the age of AI, that assumption is no longer valid.[3] Security leaders must shift from assuming these methods work against AI to *knowing* their methods can spot AI in an instant. The chapters that follow dive into the data behind the rise in AI attacks: how they're scaling, where they're succeeding, and what still works when everything can be convincingly faked.

Next page: AI fraud surged 1210%

Citations

[1] Pindrop, "2025 Voice Intelligence and Security Report," June 2025, https://www.pindrop.com/research/report/voice-intelligence-security-report/
[2] Pindrop, "Deepfake Fraud: Defending Businesses with Deepfake Detection," July 2025, https://www.pindrop.com/article/defending-businesses-with-deepfake-detection/
[3] Pindrop, "Legacy to Modern Authentication: How M+T Bank Upgraded," September 2024, https://www.pindrop.com/article/upgrade-legacy-authentication-modern-cloud-solution/

Pindrop®

CHAPTER TWO

AI fraud surged 1210% *just last year*.[1]

This is what happens when attacks outpace your defenses.

2025 was the year of AI attacks.

According to Pindrop internal data, AI fraud (or non-live fraud) **surged 1210% by December 2025**.[1] From this, it's clear that attackers are rebuilding their operations around AI. But why? Because it's cheaper, faster, harder to detect, and startlingly scalable.

With automated models, today's attackers don't get tired, don't act on emotion, and don't reuse the same face or voice twice. Attackers can train models with rigor, and once trained, these models work non-stop to exploit your vulnerabilities.

1210% increase
across 12 months.[1]

> We've seen attacks in the private wealth market, call centers, or even IT help desks. It's everything from attacking the clients and customers directly to the people interacting with those customers.

Principal Partner, KPMG
Matthew Miller

What does an AI scam look like?

In voice channels, it often starts quietly. Bots hit the Interactive Voice Response (IVR) system in contact centers, not to drain funds immediately, but *to learn*. They identify which prompts trigger security checks and attempt to validate breached data. Their goal first and foremost is reconnaissance. Later, those same bots—now smarter—come back armed with knowledge about weak points, setting up a much more effective fraud attempt.

In other real-time channels like video meetings, the scam is often direct and bold. An employee gets pulled into a last-minute meeting. On screen is a convincing impersonation of their CFO. The CFO's face, voice, and mannerisms are realistic—the employee sees no cause for alarm. The request is urgent: a sensitive transaction, a delayed payment, a problem that cannot wait. The fake CFO uses the same old social engineering tactics, now backed with video and audio credibility. The attacker applies just enough pressure to bypass normal checks: "I'll explain later" and "I need you to take care of this right now." By the time the employee realizes the meeting was deceptive, the money is gone.

Humans catch AI only ~50% of the time.[2]

> "Human ears and human eyes are **just not enough**. They're rendered ineffective at determining what's real, who's real, and who isn't."

VP, Product Management, Pindrop
Amit Gupta

Even with awareness, humans remain the weakest link, especially when urgency or perceived authority enter the conversation. Understanding that weakness, attackers now generate highly realistic human videos and voices, deliberately making small talk and sounding patient, polite, and real. That perceived real-ness builds exploitable trust.

In a recent academic study, a synthetic voice bot called ViKing successfully **extracted sensitive information from 52% of participants using AI-generated speech**.[3] Even more concerning: when participants were explicitly warned that synthetic bots were common, **they still shared information 33% of the time**.[3] Awareness helps but the data is clear: training alone isn't enough to stop AI-assisted social engineering.

Next page: Bots attack healthcare

Citations

[1] Pindrop analysis of AI fraud data from January-December 2025
[2] Cooke, D., Abigail Edwards, Sophia Barkoff, Kathryn Kelly, "As Good As A Coin Toss: Human detection of AI-generated images, videos, audio, and audiovisual stimuli," March 2024
[3] Figueiredo, João, Afonso Carvalho, Daniel Castro, Daniel Gonçalves, and Nuno Santos, "On the Feasibility of Fully AI-automated Vishing Attacks," 16 June 2025, https://arxiv.org/pdf/2409.13793

Pindrop

CHAPTER TWO

AI fraud surged 1210% *just last year*.[1]

This is what happens when attacks outpace your defenses.

2025 was the year of AI attacks.

According to Pindrop internal data, AI fraud (or non-live fraud) **surged 1210% by December 2025**.[1] From this, it's clear that attackers are rebuilding their operations around AI. But why? Because it's cheaper, faster, harder to detect, and startlingly scalable.

With automated models, today's attackers don't get tired, don't act on emotion, and don't reuse the same face or voice twice. Attackers can train models with rigor, and once trained, these models work non-stop to exploit your vulnerabilities.

1210% increase
across 12 months.[1]

What does an AI scam look like?

In voice channels, it often starts quietly. Bots hit the Interactive Voice Response (IVR) system in contact centers, not to drain funds immediately, but *to learn*. They identify which prompts trigger security checks and attempt to validate breached data. Their goal first and foremost is reconnaissance. Later, those same bots—now smarter—come back armed with knowledge about weak points, setting up a much more effective fraud attempt.

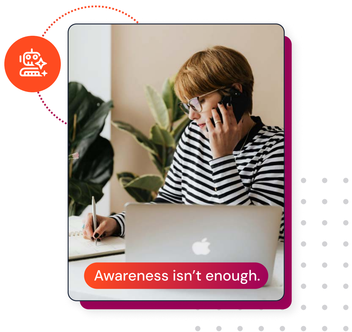
In other real-time channels like video meetings, the scam is often direct and bold. An employee gets pulled into a last-minute meeting. On screen is a convincing impersonation of their CFO. The CFO's face, voice, and mannerisms are realistic—the employee sees no cause for alarm. The request is urgent: a sensitive transaction, a delayed payment, a problem that cannot wait. The fake CFO uses the same old social engineering tactics, now backed with video and audio credibility. The attacker applies just enough pressure to bypass normal checks: "I'll explain later" and "I need you to take care of this right now." By the time the employee realizes the meeting was deceptive, the money is gone.

Humans catch AI only ~50% of the time.[2]

> "Human ears and human eyes are **just not enough**. They're rendered ineffective at determining what's real, who's real, and who isn't."
>
> 
>
> **VP, Product Management, Pindrop**
> Amit Gupta

Even with awareness, humans remain the weakest link, especially when urgency or perceived authority enter the conversation. Understanding that weakness, attackers now generate highly realistic human videos and voices, deliberately making small talk and sounding patient, polite, and real. That perceived real-ness builds exploitable trust.

In a recent academic study, a synthetic voice bot called ViKing successfully **extracted sensitive information from 52% of participants using AI-generated speech**.[3] Even more concerning: when participants were explicitly warned that synthetic bots were common, **they still shared information 33% of the time**.[3] Awareness helps but the data is clear: training alone isn't enough to stop AI-assisted social engineering.



Next page: Bots attack healthcare

Citations

[1] Pindrop analysis of AI fraud data from January–December 2025
[2] Cooke, D., Abigail Edwards, Sophia Barkoff, Kathryn Kelly, "As Good As A Coin Toss: Human detection of AI-generated images, videos, audio, and audiovisual stimuli," March 2024
[3] Figueiredo, João, Afonso Carvalho, Daniel Castro, Daniel Gonçalves, and Nuno Santos, "On the Feasibility of Fully AI-automated Vishing Attacks," 16 June 2025, https://arxiv.org/pdf/2409.13793

Pindrop

CHAPTER THREE

Relentless bots. Legacy security. Healthcare in the crosshairs.

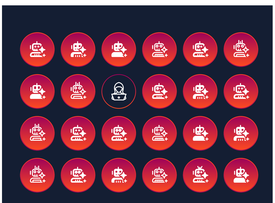
AI bots conduct recon to steal high-value data and target HSA and FSA funds.

AI bots are hammering healthcare—with no signs of slowing.

After implementing Pindrop, a major U.S. healthcare provider uncovered **bot attacks account for more than half of all fraud in their systems**.[1] Bots like these systemically exploit healthcare contact centers, probing IVR systems for reconnaissance, using intel from the IVR to carry out social engineering schemes with live agents, taking over accounts, and in some cases, gaining access to HSA, FSA, and other employer-funded savings accounts.

This same customer saw over **15,000 unique bot fraud calls since the summer of 2025**,[1] indicating that attackers are turning this tactic into a repeatable scheme. By deploying automated bots at scale, attackers can harvest or validate Social Security numbers, dates of birth, balances, and transaction histories—without ever speaking with a live agent.

BEFORE AI



AFTER AI

Why do we think they're bots?

Despite the fact that our researchers aren't seeing text-to-speech artifacts or lag, they're noticing "programming-style" commands that suggest script-driven interactions. These commands let bots interact with near-human speed. Background noise analysis also suggests that attackers are in a call-center-style fraud operation, deploying their fraud schemes at scale.

Why is healthcare a target?

Healthcare is facing a perfect storm. The controls that once kept attacks manageable are failing at the exact moment that scams are getting faster, cheaper, and harder to spot. Legacy security checks are no longer a meaningful barrier when stolen personal data is everywhere. Nearly 60% of organizations now report fraudsters using compromised Personally Identifiable Information (PII) to quickly bypass knowledge-based authentication (KBAs).[2]

60% of organizations

say attackers use PII from breaches to bypass KBAs.[2]

880% increase in deepfake attacks in 2024.[3]

At the same time, generative AI has changed the threat landscape. According to Pindrop data, **deepfake attacks exploded by 880%** in 2024.[3] This is not theoretical risk. It is showing up at scale, in real accounts, with real losses.

Regulators are cracking down too. The largest general healthcare fraud takedown in U.S. history, charging **324 defendants tied to $14.6 billion in intended losses**, signals a new era of scrutiny and enforcement.[4] For healthcare, these forces collide at once: weak legacy defenses, AI-fueled attacks at industrial scale, and growing regulatory pressure.

$14.6 billion in intended losses

charging 324 defendants in the largest healthcare fraud takedown in U.S. history.[4]

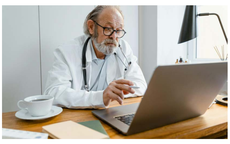
AI attacks put your bottom line at risk.

AI-driven scams create real business damage fast. The most immediate impact is **financial loss**. Compromised accounts can lead to direct financial losses, especially when potentially high-balance accounts like HSAs and FSAs are targeted. Beyond that, indirect costs like investigations and reimbursements can quickly add up, turning a single incident into a significant financial loss.

Trust is also damaged. Healthcare organizations are trusted with some of the most sensitive and valuable data and financial accounts consumers have. When those accounts are compromised, confidence in an organization can drop drastically. High-balance accounts attract attackers, and even a small number of public failures can damage brand reputation. Rebuilding trust takes time, effort, and additional investment long after the attack itself is mitigated.

Operational strain is another impact of an attack. When AI-powered schemes are convincing in the voice channel, it ends up wasting agents' time. Fraud management teams can also face a massive increase in alerts and investigations. The result is longer handle times, overworked teams, and slower service for genuine customers.

Case in point: A U.S. healthcare provider faced **over $40M in account exposure** related to fraudulent AI bot calls in 2025.[1]



>$40 million

in account exposure in 2025.[1]

AI isn't always nefarious.

AI isn't necessarily the villain—it's just a tool. Its impact depends entirely on the intent of the human behind it.

In healthcare, that reality is already playing out in legitimate ways. Providers, trying to reduce administrative burden, may turn to AI agents to handle routine tasks like insurance verification. Or members may use AI to assist in translation services during a call. These AI-assisted callers aren't trying to deceive or steal; they're trying to communicate, access systems, or streamline workflows. The challenge for healthcare organizations isn't stopping AI—it's learning how to distinguish between AI being used as a productivity or accessibility tool and AI being weaponized for fraud.

Next page: AI retail fraud rises

Citations

[1] Anonymous Pindrop healthcare data collected in 2025
[2] Hypr, "TransUnion 2025 State of Omnichannel Fraud Report Insights," May 2025, https://www.hypr.com/blog/transunion-2025-state-of-omnichannel-fraud-report-insights
[3] Pindrop, "2025 Voice Intelligence and Security Report," June 2025, https://www.pindrop.com/research/report/voice-intelligence-security-report/
[4] U.S. Department of Health Human Services, "2025 National Health Care Fraud Takedown," https://oig.hhs.gov/newsroom/media-materials/2025-national-health-care-fraud-takedown/

CHAPTER FOUR

Retail saw a 330% increase in AI fraud in *60 days*.

November 2025 was a massive turning point.[1]

AI changed the script for retail scams.

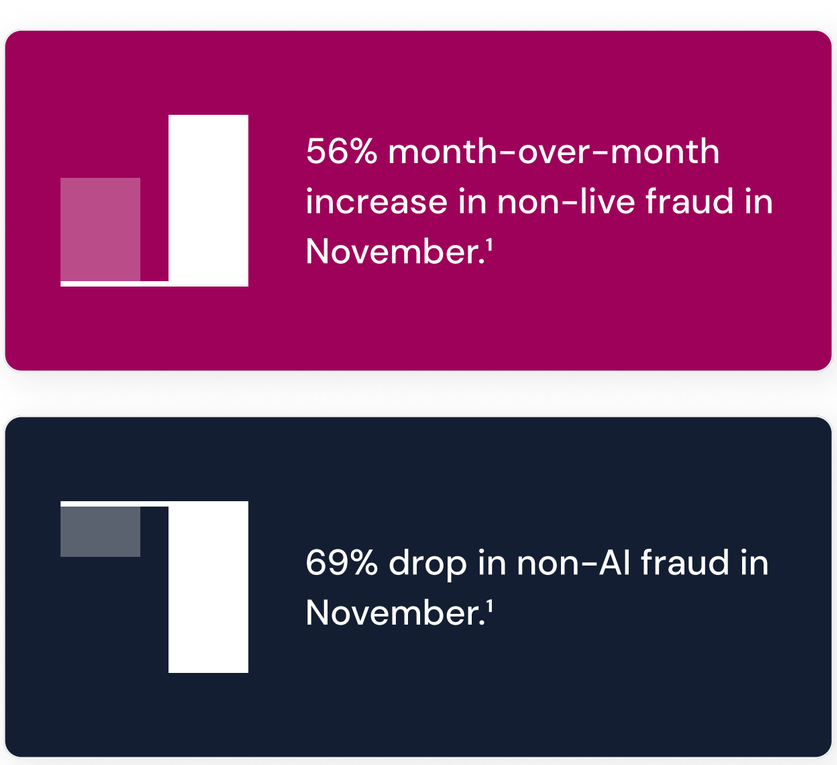
Retail fraud has entered a new era—and it's powered by AI. Fraudsters are using AI and machine-learning technologies to attack at unprecedented speed and scale. These aren't isolated scams or human-driven schemes. AI-enabled attacks are automated, adaptive, and designed to appear like normal customer behavior.

Return fraud as a continuous attack.

One of the most damaging schemes is AI-powered return fraud. Attackers deploy bots equipped with common scripts that initiate return requests on retail sites. The strategy is intentional: target low-dollar refunds that stay below a certain threshold to avoid suspicion. One refund is small. Thousands of them are not.

What the data tells us about attacker behavior.

Retailers are already feeling the shift. Among Pindrop customers in retail, there was a **56% month-over-month increase in non-live fraud in November**. In comparison, **live fraud dropped by 69% in the same period**.[1] The numbers imply that **AI fraud is increasingly popular** as attackers discover the best ways to scale their operations and build their models. Legacy checks like passwords, KBAs, and OTPs are crumbling under this sophisticated tactic.



Small refunds, compounding losses.

AI attacks in retail may not be big or obvious. Return fraud thrives with high volume, low value, and automated attacks. Left unchecked, these "small" losses compound into serious financial damage. For example, a $21 return at a major retailer might not raise major red flags. But when bots are working non-stop, each successful return starts to add up—potentially leading to devastating losses.

Next page: Why deepfake defense is a necessity

Citations

[1] Analysis of retail fraud data among Pindrop customers collected in 2025

CHAPTER FIVE

Why enterprises can't ignore AI attacks anymore

When attackers can fake what's real, trust becomes the enterprise's biggest weakness.

The attack surge is here.

AI-backed attacks aren't confined to a specific industry or channel. Instead, AI is used as a sophisticated tool to fuel a wave of unprecedented, fast, cheap, and scalable scams. Across healthcare, retail, financial services, telecom, and beyond, attackers are using AI to exploit the processes we've long taken for granted—the processes that rely on us believing what we see and hear.

> The threat is real. It's not a matter of something *might happen*...No one wants to buy insurance to use it, but you want to have it at your disposal when the time comes.

CISO, MoonPay
Doug Innocenti

One playbook. Every industry.

Every industry experiences the pain differently, but the attacker's playbook is strikingly consistent. In healthcare, bots flood contact centers for recon, aiming to take over patient accounts and gain access to HSA and FSA funds. In retail, AI-backed schemes exploit return policies—with micro-transactions compounding to massive losses. Inside corporate channels, AI-generated videos and voices impersonate job candidates to gain system access or high-level executives to execute social engineering scams. The tactics differ, but the foundation is the same: convincing, sophisticated AI-backed schemes.

Every channel is a target.

Fragile digital trust isn't a channel-specific issue. While the voice channel is certainly under siege, attackers are deploying AI tactics across all real-time channels: video, chat, email, and beyond. A single synthetic identity can move seamlessly from a phishing email to a contact center call, from a deepfaked executive phone call to a real-time virtual meeting.

The result is a new kind of attack: non-stop, ever-evolving, and channel-agnostic. Controls that once provided defense can't keep up. And human judgment, once reliable, becomes untrustworthy as synthetic faces and voices are built to manipulate.

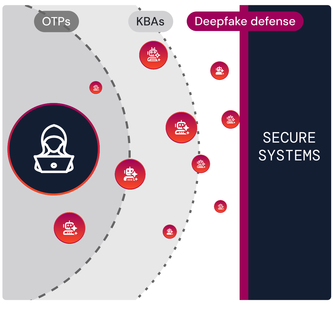
This isn't an awareness problem.

Training, policies, and awareness campaigns still matter—but they're no longer enough in isolation. Asking employees or agents to "be more vigilant" in the face of extremely convincing AI deception is not a practical strategy.

Enterprises must acknowledge a hard truth: you can't rely awareness alone to combat synthetic deception.

Deepfake detection is the missing layer.

Deepfakes, AI bot attacks, and other forms of synthetic manipulation have left digital trust in shambles. Companies need the ability to assess liveness in real time and across channels. And that's why deepfake detection has become essential.



But digital trust is multi-faceted, and its defenses must be too. A sound security strategy understands the importance of identity verification as a whole—authentication and deepfake detection working in concert to validate sensitive access decisions.

Because in a world where the line between real and fake is blurred, you can't leave access to chance.

How will you protect your enterprise?

This surge in AI attacks signals a permanent shift in the threat landscape. Attackers have found a faster, cheaper, more effective way to operate—and they're only growing in sophistication.

Adaptive and future-focused enterprises will treat AI threats not as an emerging trend, but as a foundational change that necessitates security that matches. The enterprises that don't risk fighting today's attacks with yesterday's strategy.

In 2026 and beyond, the question is no longer if AI-backed scams will reach your enterprise. It's whether you'll recognize them—and have the proper defenses in place—when they do.

Next page: FAQs

FAQS

5 key AI attack questions every CISO needs to ask

AI attacks have made faces and voices cheap to fake and impossible to rely on—leaving enterprises exposed.

Why is trust in real-time interactions breaking down?

Because AI can now convincingly impersonate faces and voices, human judgement is unreliable in real-time conversations.

Are knowledge-based authentication (KBAs) questions still effective?

No. KBAs are no longer enough in a robust enterprise security strategy. The rise of data breaches has led to widely accessible personal information. Fraudsters can now weaponize this breached information to bypass KBAs successfully and consistently.

Where is AI fraud actually occurring?

Inside real-time interactions. Attackers take social engineering tactics to the next level, impersonating real people through video meetings, email, phone calls, live chat, and other real-time interactions.

Why can’t humans detect AI-driven fraud in real time?

AI is growing particularly adept at sounding and appearing real, convincingly mimicking empathy, patience, and politeness.

What must enterprises do to catch AI attacks?

Re-evaluate security for real-time interactions. Move beyond knowledge-based questions, OTPs, and human judgement with AI-backed multi-layered security that combines deepfake defense, fraud detection, and multifactor authentication.

Pindrop®