outtake

2026

Digital Risk Report

Protecting Digital Trust in the AI Era

Research by

Cybersecurity INSIDERS

Executive Overview

Digital risk has crossed a threshold. As enterprises hardened endpoint, identity, cloud, network, and email, adversaries shifted to the open internet where business trust is exposed. What once appeared as isolated impersonations and one-off fraud has become a coordinated, industrial-scale operation. The asset under attack is trust: trust in executives and employees, trust in brands, trust in the workflows that move money. Attackers now run campaigns end to end, while most digital risk programs still respond one incident at a time. That gap is visible across this survey of more than 1,100 security and risk leaders.

Behind every visible artifact, from a spoofed domain or fake social account to a deepfake video or synthetic persona, sits an operator directing the campaign. These campaigns move through a recognizable kill chain: reconnaissance, infrastructure setup, trust exploitation, target engagement, credential capture, account takeover, impact and fraud, and monetization. Yet most programs still intercept visible artifacts case by case, without following the chain back to the operator behind them.

Key findings:

- **Digital risk is a business risk category, not a security tool problem:** 84% of organizations experienced material digital risk incidents in the past year, yet only 7% describe their program as leading. Costs spread across staff hours, customer support, legal response, and executive time, yet 21% have no single owner for digital risk at all.
- **People are the most exposed and least protected attack surface:** Executive or employee impersonation hit 53% of organizations this year. Yet protection remains narrow: 77% limit workforce coverage to executives, a few high-risk roles, or reactive case-by-case response, and 43% run no person-of-interest threat profiling at all.
- **Detection, investigation, response, and measurement are broken across the kill chain:** Only 7% have end-to-end visibility from reconnaissance through fraud. 42% say attacks now move faster than detection, 34% close cases at takedown without pursuing the adversary behind them, and 28% run no takedown SLA to measure response performance. Failures at one stage cascade into the next.
- **AI is opening a second front:** 47% have already encountered confirmed or suspected synthetic-media impersonation of an executive or brand representative, while AI-generated attacks that look like real activity now top the survey's visibility gaps at 44%. On the exposure side, only 4% have full visibility with active controls over their agents' external interactions, and 96% have no automated way to stop an AI agent manipulated through its external inputs - the AI Trust Gap most programs haven't yet measured.
- **The market is at an investment inflection point with no category leader:** 58% plan to increase digital risk investment in the next year, yet 82% still lack a purpose-built platform. 69% place themselves below an established response model. The next 12 to 24 months will decide whether digital risk consolidates into a purpose-built platform category or stays fragmented across point tools and manual workflows.

Each finding traces back to the same problem: a coordinated, AI-amplified threat operating across channels and surfaces is being met by a fragmented response that no single team owns and no single platform defends against. Investment is rising into that gap, but the architecture hasn't yet consolidated to close it. At this scale, and at machine speed, response has to be an agentic loop: AI agents running pre-staged detection, investigation, attribution, takedown, verification, and learning at machine speed, with humans setting policy and judgment.

Digital Risk Has Outgrown Security Operations

Industrial-scale digital risk hits every function. The heaviest costs fall outside security operations, and the data shows exactly where.

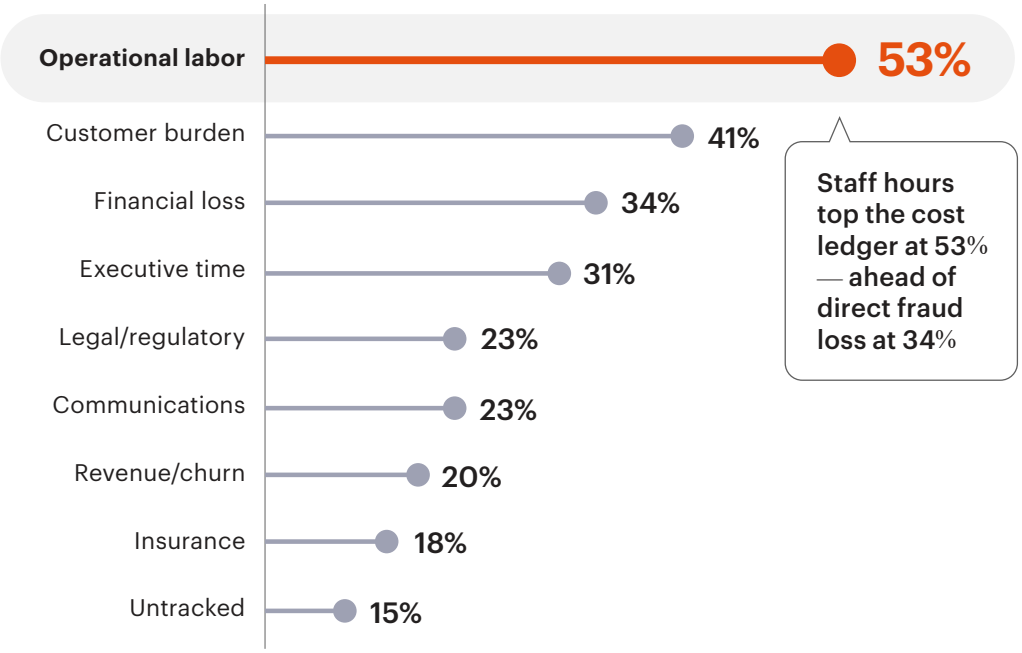
84% of organizations experienced material digital risk incidents in the past year, yet only 7% describe their program as leading. Behind that gap is coordinated deception in many forms at once: 65% saw lookalike or homoglyph domains, 53% had an executive or employee impersonated, and 47% faced coordinated multi-channel campaigns, often within the same twelve month window. The business impact runs in parallel: credential theft (34%), brand harm (32%), direct fraud (30%), customer confusion (29%), and payment diversion risk (26%).

The cost profile tells the same story. Staff hours lead the cost categories at 53%, ahead of customer support (41%), executive time (31%), and legal and regulatory response (23%); another 23% don't track the cost at all. The largest single cost is the labor of cleanup, spread across functions outside security operations. That is what turns digital risk into a P&L event, even when the cost never lands in a single budget line.

Take an executive impersonation campaign that runs for six weeks before discovery. By takedown, the cleanup has pulled in five functions. Marketing rewrites public statements. Customer support fields confused inquiries. Legal coordinates removal. The executive's office scans for further fakes. Risk and Insurance file an incident report. None of the five are in security operations. Digital risk now sits at board level. The governance required to manage it (including ownership, accountability, and cross-functional authority) has not yet caught up.

Where the Cost of Digital Risk Lands

What has been the true cost of digital trust incidents your organization experienced in the past 12 months?



84%
experienced material digital risk incidents

Only 7%
describe their program as leading

Where this is working, every team that touches digital risk operates from the same connected workflow, with shared evidence, a single operator view, and one accountable owner.

People Are Now the Attack Surface

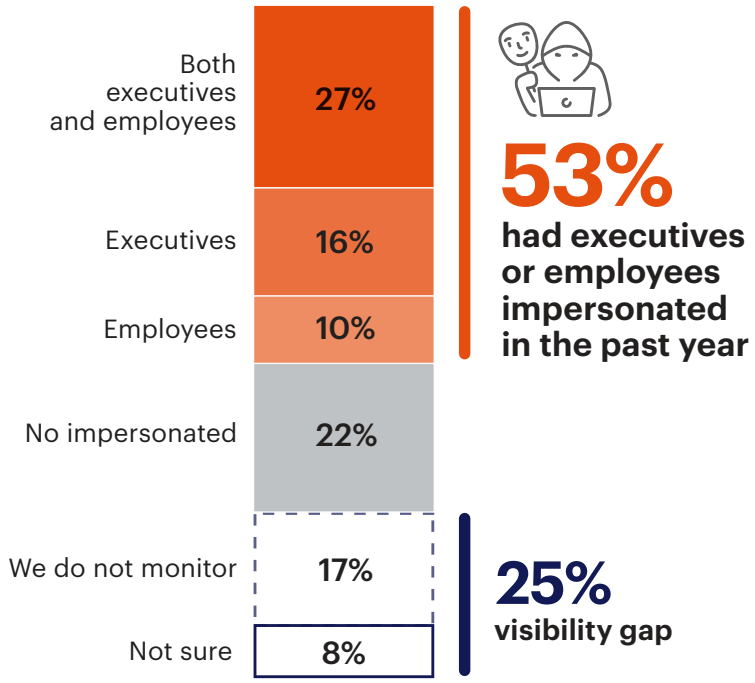
Digital risk turns personal. The individuals who carry authority, access, and credibility are among the most exposed attack surfaces, and often the least protected.

More than half of organizations (53%) had an executive or employee impersonated in the past year, with 27% seeing both and another 17% not monitoring at all. Executive impersonation operates through authority and urgency, while employee impersonation opens doors through access and familiarity. The activity spans spoofed email or messaging (41%), social media (36%), and professional networks (32%), the surfaces individuals actually live and work. Security has long been organized around technical attack surfaces such as networks, endpoints, credentials, applications. The 2026 data expands that priority list. The asset under attack is now a person's identity, and individuals live on open channels most brand-protection programs do not reach.

Personal exposure feeds the attack. Adversaries assemble target profiles, including home addresses, family details, and contact data, from broker sites, credential dumps, and public signals long before impersonation campaigns begin. The protection gap on this reconnaissance surface is severe: 43% of organizations have no Person-of-Interest threat profiling capability for actively targeted individuals, 35% run no active PII removal program across broker and people-search sites, and 52% lack visibility into encrypted or decentralized channels such as WhatsApp. The reconnaissance window stays open while protection focuses on what comes after.

Who Has Been Impersonated

In the past 12 months, has any executive or employee at your organization been impersonated online?



53% had executives or employees impersonated in the past year

25% visibility gap

43% have no POI threat profiling capability

35% run no active PII removal program

52% have no or limited visibility encrypted/decentralized channels

The protection layers that should catch this earlier are missing in almost half of programs

41%	36%	32%
Spoofed email/ messaging	Social media	Professional networks

Programs that have extended brand protection to people monitor executive and employee impersonation continuously, attribute it to the operator, and remediate on a measurable SLA. A program that can't name which executives and high-risk employees were impersonated this year, on which channels, and how quickly each case closed is still protecting brands while its people stay exposed.

Workforce Protection Is Still Too Narrow

By the time an attacker picks a target, they have already mapped which roles hold the access worth taking. Yet workforce protection rarely reaches past the executive suite.

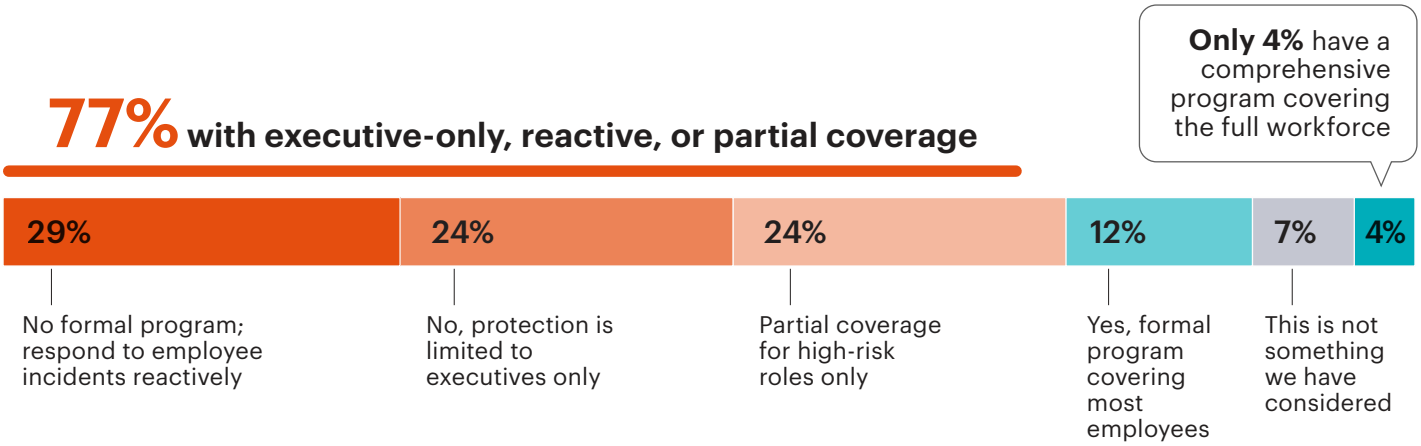
77% keep protection limited to executives, react case by case, or cover only a few high-risk roles. Within that group, 29% run no formal program at all and respond only after an incident lands. Another 12% cover most employees, and just 4% run comprehensive coverage across the full workforce.

Attackers move through the roles with the right access, regardless of title. Finance approvers release payments. IT admins hold the access keys. Customer-facing staff carry the brand trust. Heavy executive monitoring just routes the threat around the protected layer, toward the unmonitored seat one level down that still moves money or grants access. When one of those seats is hit, 19% have no defined owner for the response, so delay and duplication land exactly where speed matters most. Protection built around seniority watches the wrong list.

An attacker ignores the monitored CFO and goes after the accounts-payable manager who actually releases funds. No program covers that seat. The fraudulent payment clears.

How Far Workforce Protection Reaches

Does your organization have a formal digital protection program for the broader employee base, beyond executives?



77% keep protection to executives, react case by case, or cover only a few high-risk roles. 4% cover the full workforce. Attackers go through whichever seat holds the access — and most of those seats sit beyond the executive suite.

CFO | AP Manager | IT Admin | HR | Customer-facing staff

Programs ahead of this extend POI profiling, broker-site monitoring, and impersonation detection beyond the executive suite to the seats that actually hold the access: finance approvers, IT admins, and customer-facing staff. Protection that stops at seniority leaves the attacker's real targets uncovered.

AI-Generated Attacks Now Look Real

Attackers can now make impersonations that look and sound like the real thing. AI made that possible at industrial scale.

44% name AI-generated attacks that look like real activity as their biggest visibility gap, the top gap in the survey. The threat is already live: 47% have confirmed or suspected synthetic-media impersonation, including voice clones or deepfake video, of an executive or brand representative (18% confirmed, 29% suspected).

Defenders point at the same threat through three related visibility gaps: speed, hidden platforms, and cross-channel movement. 42% flag attacks moving faster than detection. 39% flag platforms they can't see into. 32% flag activity they can't connect across channels. 35% name AI-generated deception detection as a top buying priority. CISOs see the same threat from both sides: the largest defense gap is also one of the clearest investment priorities.

The old tells used to catch fakes were bad grammar, distorted images, and off-tone phrasing. AI has made those tells unreliable, and content-provenance standards like C2PA are not yet widespread enough to replace them. Defenders who still rely on something "looking off" are losing ground. Detection has to move earlier in the kill chain, to where the campaign is being built.

The AI Threat From Three Angles

What are the biggest gaps between your current threat visibility and what adversaries can actually do to your organization?

44%

Name AI-generated attacks as their #1 visibility gap

In the past 12 months, has your organization identified synthetic media or deepfake impersonation?

47%

Have confirmed or suspected synthetic-media impersonation in the past 12 months
(18% confirmed + 29% suspected)

Which capabilities are your top investment priorities for the next 12 months?

35%

Name AI-generated deception detection as a top buying priority

CISOs see the same threat from three angles: the biggest visibility gap, a confirmed live problem, and a top buying priority. The old tells that used to catch fakes — bad grammar, distorted images, off-tone phrasing — are gone.

The defenders staying ahead of AI don't wait for the attack to land. They run pre-staged detection: fake accounts seeded, lookalike domains registered, content libraries built - caught before a campaign goes live. Then they intercept the adversary behind them. If detection still starts after the fake content shows up, the defense window has already closed.

AI Agents Create a New Trust Boundary

That's the attack side. On the exposure side, organizations are deploying AI agents into the same open environment attackers are flooding with synthetic content. As agents begin to participate in communication, research, transactions, and decisions, they read external inputs and act on them, mostly without active oversight.

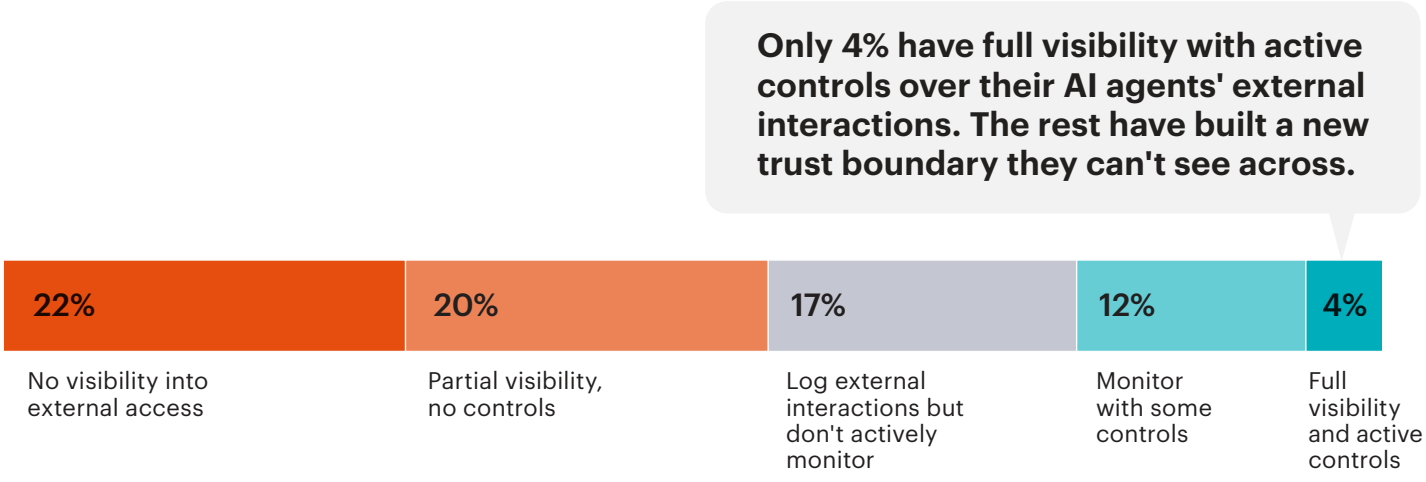
Only 4% have full visibility with active controls over their agents' external interactions, roughly 1 in 20. Another 12% monitor agent activity with some controls in place. Everyone else running these agents is more exposed: 22% have no visibility into what they access externally, 20% have partial visibility but no controls, and 17% log activity they never actively monitor. The remaining 15% don't run external agents at all.

The risk is concrete: adversaries can plant instructions in external content (such as emails, web pages, or documents) that an agent reads as part of its normal work. This technique, known as indirect prompt injection, is the top entry on OWASP's Top 10 for LLM Applications. The agent treats the planted input as legitimate and acts on it. Most programs have no visibility into when that happens.

An accounts-receivable agent reads an email asking about a payment. Hidden instructions in the message direct the agent to forward customer payment details to an external address. The agent acts. No one sees it for three days. The agent now stands on a new trust boundary: one foot in the untrusted outside world, one in the trusted internal system. A planted instruction crosses the boundary between them.

Visibility Into AI Agent Interactions

Which of the following best describes your visibility into what AI agents and automated workflows in your organization access or retrieve from external sources?



Do not use AI agents with external access 15% | This is not something we have considered 6% | Not sure 4%

Programs treating agents as governed identities monitor external inputs for planted instructions, log agent actions, and inspect outputs before they propagate. Without that loop, agent decisions become a path no one can see, interrupt, or reconstruct.

Few Can Stop a Hijacked Agent

Seeing the manipulation is only half the problem. Containment is the bottleneck. Once an agent acts on a planted instruction, most programs have no automated way to stop or roll it back.

14% can detect manipulation but can't contain it automatically. 25% rely on manual review of high-risk outputs. 34% know the risk and have built neither detection nor containment. 14% aren't aware of the risk at all. Between them, more than 9 in 10 have no automated way to stop hijacked agents before they act.

It comes down to speed. An AI agent acts in seconds. Manual review and human escalation take minutes or hours. The 14% that can detect without automated containment may find out only after the agent has moved the money, sent the data, or made the call. And the agent acts with privileged access: an unstopped one is a trusted insider carrying out an attacker's instructions at machine speed.

That is the AI Trust Gap: 96% have no automatic way to stop a manipulated agent before it acts. For them, the brake is a person catching it in time. The few with real containment build an automatic stop into the agent: output inspection, egress limits, and a halt that triggers without waiting for a human. An agent that can act until a person notices carries two gaps at once, a visibility gap and a containment gap.

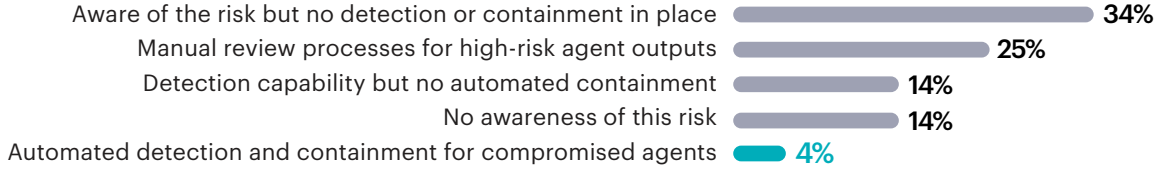
The AI Trust Gap

Does your organization have any mechanism to detect, isolate, or roll back an AI agent that has been manipulated through adversarially controlled external content?

96%
cannot automatically detect and contain a hijacked AI agent

4%
can detect and contain a hijacked AI agent

Between an AI agent acting in seconds and a human noticing in minutes, the brake is somebody catching it in time.



An additional 9% have not considered this risk (excluded from breakdown)

Programs treating agent security as infrastructure put governance on the same operational footing as identity and access management: continuous monitoring, manipulation detection at input-time, and automated containment that fires before the action lands.

Detection Depends on the People Being Harmed

For most organizations, the first signal of an attack arrives from outside the company - from customers, partners, or the people already being harmed.

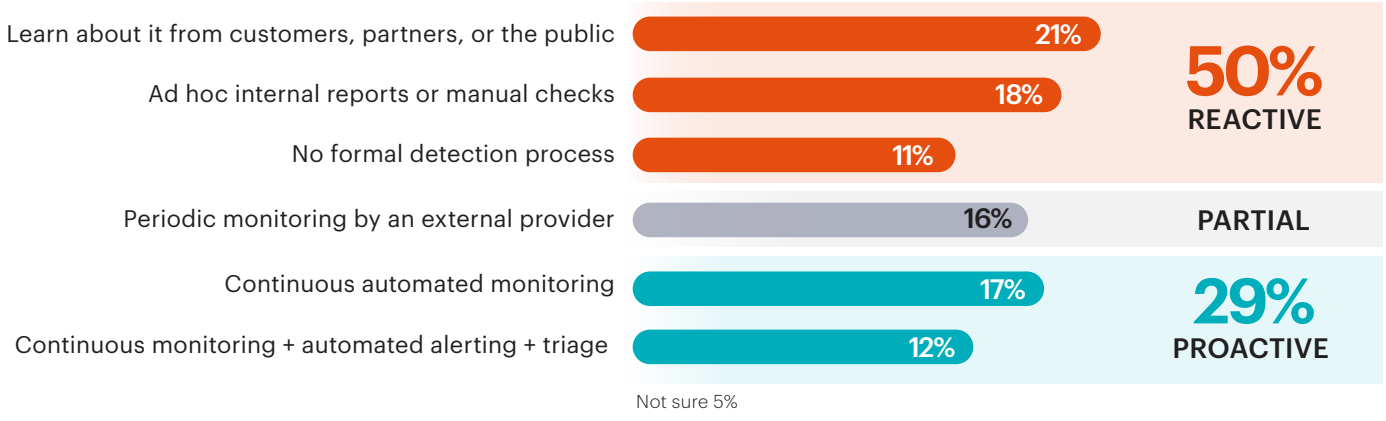
Digital risk attacks unfold across a kill chain: reconnaissance, infrastructure setup, trust exploitation, target engagement, credential capture, account takeover, impact and fraud, and monetization. Defenders have to cover the same path. Most programs cover only part of it, and at every stage more than half come up short.

21% of organizations learn about brand impersonation from customers, partners, or the public. 18% rely on ad hoc internal reports. 11% have no formal detection process. Only 29% run continuous monitoring, and just 12% run the full pipeline of monitoring, automated alerting, and triage. Half the field finds out from outside or doesn't look systematically.

Detection here runs on someone else's pain, a manual hunt, or an accidental discovery. For most organizations, the first signal is a customer or partner reporting harm, not the SOC or threat intelligence. Detection begins at the point of harm, long after the staging has run. Take social media impersonation. A fake account follows customers, builds an audience, and sends a phishing message. The brand finds out when a customer asks if the offer is real. Most teams stop at the artifact. 34% close cases at takedown without pursuing the operator. Only 16% map the broader campaign or attempt attribution. Just 5% conduct full campaign attribution. 9% continuously correlate activity to spot coordinated targeting. When the adversary becomes visible, most teams have already stopped looking.

Where Detection Starts

How does your organization typically discover brand impersonation activity?



50% of organizations learn about brand impersonation from customers, ad hoc internal checks, or no formal process at all. Detection runs on someone else's pain — long after the staging is done.

Programs that have closed this gap run pre-staged detection: monitoring, alerting, and triage at the reconnaissance and infrastructure stages, before campaigns launch. When most incidents surface through customers, detection is a downstream trigger rather than an early-warning system.

Coverage and Remediation Break at the Edges

Even programs that investigate well still leave channels uncovered. The channels they miss are often the hardest to see and the slowest to remediate.

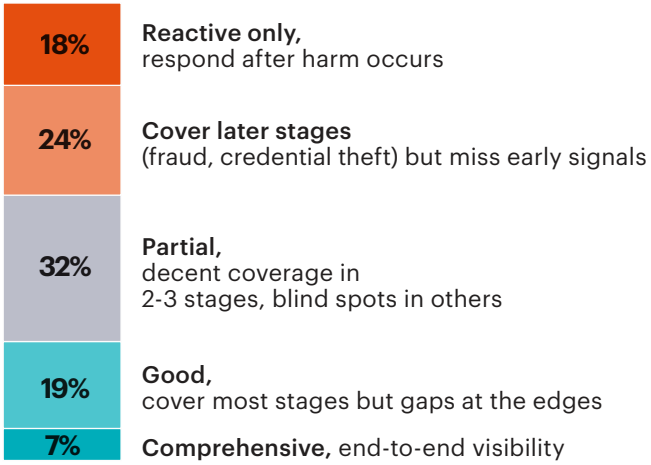
Only 7% have end-to-end visibility from reconnaissance through fraud execution. Another 19% cover most stages but still have gaps at the edges. The other three-quarters run partial coverage, miss early-stage signals, or respond only after harm.

Some channels are harder to clear than others. 22% report encrypted messaging such as Telegram as the slowest channel to remediate, followed by app stores and mobile ecosystems at 16% and social media at 14%. Another 19% don't track time-to-takedown by channel type at all. The slow channels are slow for a reason. Encrypted messaging and app-store ecosystems carry access, jurisdictional, policy, and approval barriers that slow removal. Campaigns stay live longest in exactly those channels, where defenders have the least leverage to shut them down. Defenders concentrate on the easy channels. Operators concentrate on the hard ones.

Defenders fall short on two more fronts. Only 5% correlate external threat signals with internal fraud data in real time. 28% have no defined takedown SLA, and another 24% carry informal targets they don't track. Without real-time correlation or a tracked SLA, a team can't tell whether takedown is getting faster or slower, which channels lag, or why the same campaign keeps coming back.

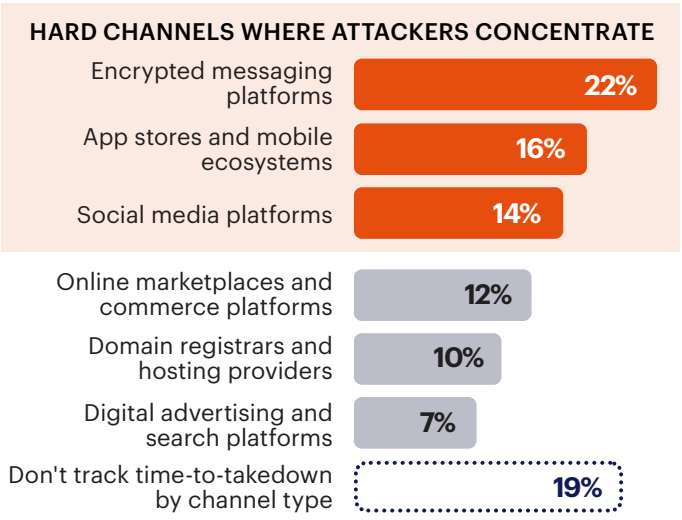
Coverage Across the Attack Lifecycle

How would you describe your threat intelligence coverage across the full attack lifecycle — from reconnaissance through to fraud execution?



Where Remediation Slows Down

Which channel type takes the longest to remediate?



Only 7% have end-to-end lifecycle coverage. The slowest channels to clear — encrypted messaging, app stores, social — are exactly where attackers concentrate while defenders cover the easy channels.

Programs with real coverage run cross-channel monitoring and coordinated takedown across the surfaces attackers actually use, including encrypted messaging, mobile ecosystems, and app stores. A program that covers email and social and stops there has secured the easy half of the surface and left the hard half open to longer-running campaigns.

No One Owns the Whole Response

Even where programs cover parts of the kill chain, responsibility still breaks at the handoffs. No one owns the whole response.

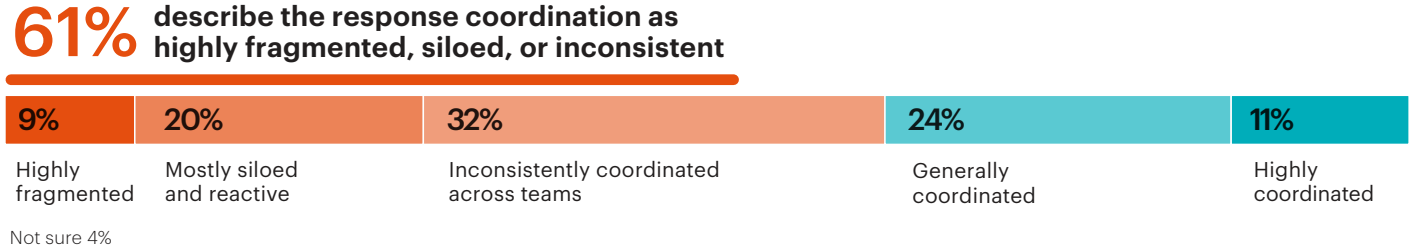
The most common answer to who owns digital risk is no one: 21% have no single owner. When organizations do name one, responsibility fragments across eight functions, none with more than 18% share. Accountability is spread thin before the incident even starts.

When an incident hits, fragmentation becomes operational drag. 61% describe response across teams as inconsistent, siloed, or fragmented. 45% have experienced a crisis where the social media narrative outpaced the company response. Teams confirm the gap themselves: only 7% describe their digital risk program as leading, while 69% remain below an established response model.

Without one accountable owner, each function holds a different piece of the evidence. A fake account lands with one team, an impersonation report with another, a leaked credential with a third. Each team sees its own artifact. Few see the campaign. The adversary moves through the seams.

Who Owns Digital Risk Today

Which team has primary responsibility for digital trust risk in your organization today?



Mature programs assign digital risk to one owner with authority across artifacts, channels, and functions. Two questions expose the gap: Who owns digital risk? Who calls the shots when an incident hits multiple functions? If the answers vary, the response workflow is not mature yet.

Architecture Remains Fragmented

The ownership gap on the previous page shows up in the architecture. Buyers are increasing investment, but most still deliver digital risk response through stitched tools, manual workflows, and weak coordination.

58% plan to increase digital risk investment over the next year. Only 18% run a purpose-built digital risk platform. Most rely on multiple internal tools stitched together (27%) or manual workflows supported by point tools (24%). Another 12% have no dedicated tooling or formal program.

The buying motion is converging faster than delivery. Only 11% report highly coordinated cross-team response, the layer every other capability depends on. More budget poured into a stitched model creates more systems, more handoffs, and more places for an adversary campaign to move between teams. Point capability cannot create the connected workflow by itself. Fragmentation has become the risk.

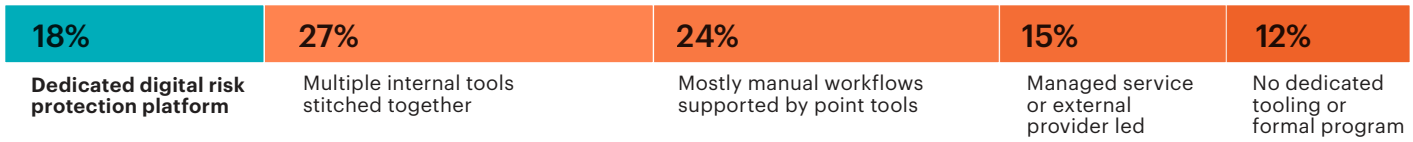
Verification shows the cost. 51% have built domain and sender authentication, while only 17% have moved to identity-based or cryptographic verification. The basic layer handles bulk spoofing. The advanced layer requires connected evidence, shared ownership, and coordinated response. Without that architecture, organizations harden the bulk channel while high-value interactions, including wires and executive accounts, remain exposed.

Investment Up - Delivery Model Still Fragmented

How will your organization's investment in digital trust protection change over the next 12 months?

SPEND IS RISING. THE ARCHITECTURE ISN'T.

Buyers are pouring budget into a fragmented delivery model that can't connect the pieces.



Not sure 4%

Organizations that escape the fragment trap consolidate the delivery model first, then add capability on top. The next investment should close seams, connect evidence, and reduce handoffs before adding another point capability.

Buyers Want the Whole Chain

Once the architecture problem is clear, the buying signal becomes sharper. Organizations are increasing spend, and their priorities point toward one connected response chain.

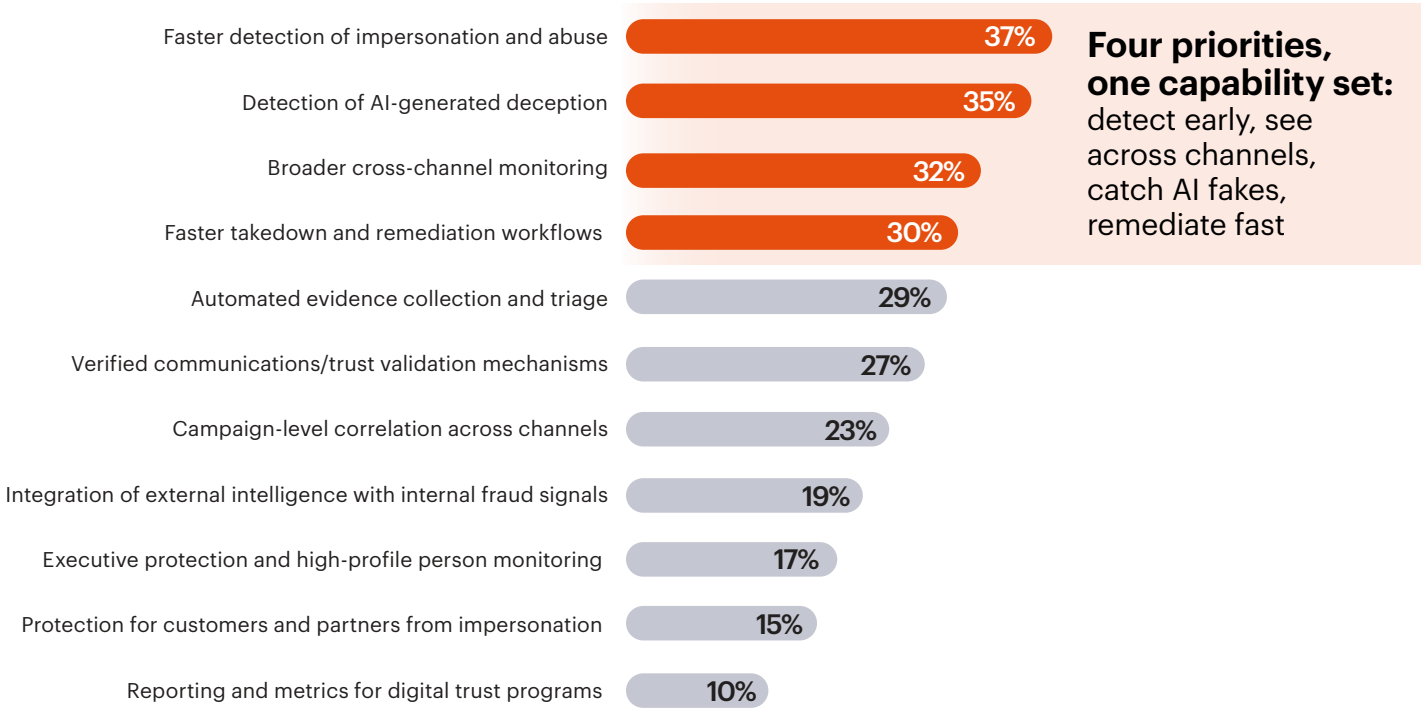
58% plan to increase digital risk investment in the next twelve months, including 19% planning significant increases. Only 7% plan to decrease.

The top priorities: faster impersonation detection (37%), detection of AI-generated deception (35%), broader cross-channel monitoring (32%), and faster takedown and remediation (30%). No single capability dominates because buyers need the chain to work end to end.

Those priorities map directly to the kill chain failures exposed throughout the report: late detection, synthetic-media deception, channel blind spots, and slow remediation. Together, they describe one capability set: detect early, see across channels, catch AI-generated fakes, and remove threats fast.

Top Investment Priorities

Which capabilities are your top investment priorities for the next 12 months?



Four priorities, one capability set: detect early, see across channels, catch AI fakes, remediate fast

The risk is that organizations fund these needs as separate projects and recreate the fragmentation they are trying to escape. Buyers getting ahead of this treat the four priorities as one connected requirement and evaluate tools on how well the capabilities link across detection, investigation, remediation, and learning.

The Response Model Has to Become Agentic

The pattern is clear. Buyers are increasing investment, but defenders still manage digital risk through disconnected workstreams.

Most organizations have parts of the required capability, but they do not operate them as a coordinated response model. Detection, investigation, takedown, verification, personal exposure monitoring, and AI agent oversight often sit with different teams. Attackers do not respect those boundaries. Their operations move across people, brands, infrastructure, channels, credentials, and now AI agents.

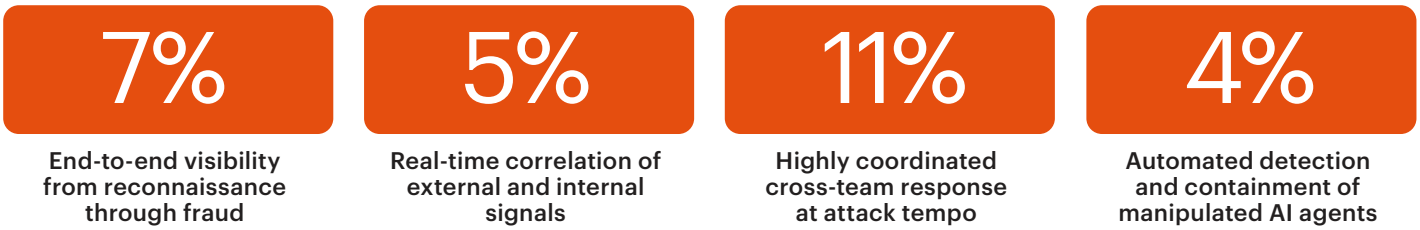
That is why many organizations fall short across the digital threat lifecycle. At reconnaissance, many do not profile executives or remove exposed PII from broker sites. At infrastructure setup, fake accounts and lookalike domains often appear without pre-staged detection. At target engagement, many organizations still learn about attacks from customers. At credential capture and account takeover, fraud lands, payment routing changes, and brand-trust attacks succeed. When defenders finally act, they often just remove the visible artifact while the operator rebuilds somewhere else. The isolated defenses exist. The workflow does not.

AI raises the stakes on both sides. Attackers use it to create better impersonation, faster campaigns, and more convincing social engineering at scale. Defenders' own AI agents introduce a new control surface that few organizations can fully monitor, govern, or stop before damage occurs.

More budget will not solve this by itself. Investment only delivers when detection, investigation, attribution, takedown, and verification operate as a connected loop. A signal must lead to investigation. Investigation must expose the operator. Attribution must accelerate takedown. Takedown must feed the next detection cycle.

Manual coordination cannot sustain that loop. 96% cannot automatically stop a manipulated agent before it acts. Only 7% have end-to-end visibility from reconnaissance through fraud. Only 11% sustain highly coordinated cross-team response at attack tempo. Only 5% correlate external signals with internal data in real time. The detection-to-response window has become too narrow for human-led processes alone.

The Capability Ceiling Across the Stages



The detection-to-response window has moved past what humans can close by hand.

At the scale and speed of AI, the response model has to become agentic. AI agents can run pre-staged detection, investigation, attribution, takedown, and verification as a continuous workflow, with each incident improving the next response. Humans set policy, judgment, and escalation thresholds. Agents execute the operational work that can no longer be sustained manually.

From Reactive to Agentic: A Digital Trust Maturity Matrix

Programs do not sit at one maturity level across the board. Most are stronger at some kill chain stages and weaker at others. This matrix maps maturity from reconnaissance through monetization across three tiers: Reactive, where coverage is minimal; Managed, where the program is structured but still mostly manual; and Adaptive, where agentic detection and response runs as a continuous loop.

KILL CHAIN STAGE	TIER 1: REACTIVE	TIER 2: MANAGED	TIER 3: ADAPTIVE
Reconnaissance	No POI profiling; broker exposure unaddressed.	POI for executives; broker removal periodic; encrypted channels limited.	Agentic exposure monitoring across executives and workforce; signals trace to the adversary.
Campaign Infrastructure Setup	No watch for lookalike domains, fake accounts, or content libraries.	Periodic scans for lookalike domains; partial fake-account detection; reactive cleanup.	Pre-staged detection across domains, accounts, and bot networks; early intercept feeds operator attribution from a single signal.
Trust Exploitation & Target Engagement	Detection from customers or external reports; cases close at takedown.	Active monitoring with gaps; inconsistent SLA; some campaign mapping.	Agentic cross-channel detection; operator attribution from a single signal; AI-deception detection at machine speed.
Credential Capture & Account Takeover	Credential exposure unmonitored; ATO discovered after impact.	Some credential monitoring; ATO via downstream signals; standardized response.	Agentic credential and dark-web monitoring; real-time ATO detection; verified high-value workflows.
Impact, Fraud & Monetization	Impact discovered late; manual takedown; ad hoc cross-functional response.	Standardized takedown; some cross-functional automation; partial impact measurement.	Agentic takedown with operator-network eradication; each investigation faster than the last.

Most programs today are not at Tier 3 at any stage. The few that are have built agentic detection and response across the stages: clear ownership, shared evidence, pre-staged detection feeding investigation, operator attribution driving containment, and containment feeding the next investigation. Buyers move their program up by investing in the agentic connective tissue. Adding capability alone leaves the program in place.

Use this matrix as a self-assessment. Mark your tier in each stage, then start your next move at the lowest one. Finally, ask the two questions the rows do not cover: who owns the response across stages, and whether those stages run as one agentic response loop.

Five Moves Toward Agentic Response

The maturity matrix names the gaps. Start where your program is weakest, then connect the five moves into one continuous response loop.

1

Cut Personal Exposure

Profile the people attackers map first: executives, high-risk workforce roles, and customer-facing roles. Remove exposed PII from broker sites, watch for re-publication, and monitor impersonation across channels before campaigns land.

2

Catch Campaign Infrastructure Before Launch

Run pre-staged detection across lookalike domains, fake social accounts, and content libraries. Attribute each artifact to its operator so the next campaign surfaces earlier.

3

Cover Every Channel

Extend coverage across email, social, mobile, encrypted messaging, app stores, and marketplaces. Run AI-deception detection at machine speed. Track takedown SLAs by channel so blind spots do not become time gaps.

4

Block ATO and Verify Workflows

Monitor credentials and dark-web sources continuously. Detect account takeover before impact lands. Require verification for high-value workflows: wires, payment-routing changes, and customer-account updates.

5

Pursue Beyond Takedown

Treat takedown as one step in the investigation, not the finish line. Pursue the operator's broader network. Coordinate response across security, fraud, legal, and customer-facing teams. Feed each incident into the next detection cycle.

The five moves cover every stage of the attack. Run together, they form one continuous loop. Detection feeds investigation. Investigation feeds attribution. Attribution feeds takedown. Takedown feeds verification. Verification feeds the next detection cycle. Each cycle runs faster than the last.

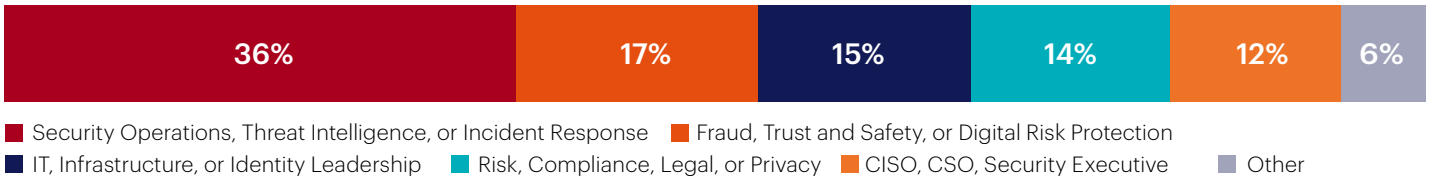
AI agents run the loop at machine speed. Humans decide what the loop is allowed to do and when to step in. Programs ahead of this curve are building the loop now. The next 12 to 24 months will separate them from the ones still trying to defend at human speed.

Methodology and Demographics

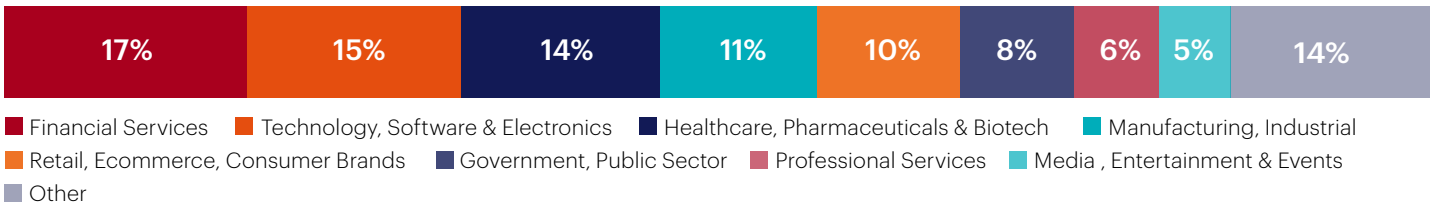
This report is based on a survey of 1,138 decision-makers responsible for cybersecurity, fraud, digital risk, and trust-related programs at their organizations, conducted in early 2026 across a wide range of industries and sizes. The research examines five dimensions of digital risk maturity: detection capability, response infrastructure, executive and employee protection, AI agent governance, and overall program maturity.

Using a stratified sampling approach, the survey achieved a 95% confidence level with a margin of error of ±2.90%.

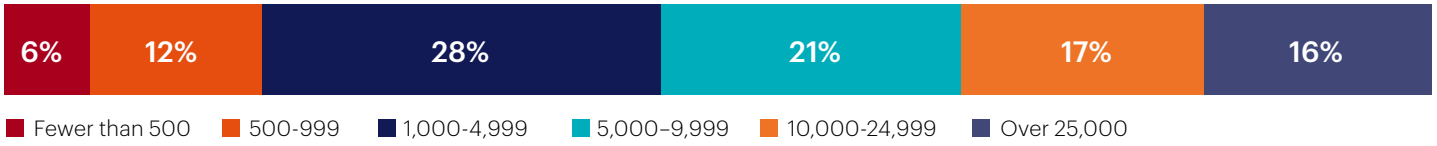
CAREER LEVEL



INDUSTRY



COMPANY SIZE



Cybersecurity
INSIDERS

BENCHMARK YOUR SECURITY MATURITY

Independent cybersecurity research revealing the gaps that shape cybersecurity strategy

Cybersecurity Insiders produces independent research based on surveys of cybersecurity leaders and practitioners worldwide. Our reports reveal where security strategies break down in practice — helping organizations benchmark their maturity, identify capability gaps, and prioritize the actions needed to close them.

For more information, visit

cybersecurity-insiders.com

June 2026 // Version 1.5

outtake

Outtake is on a mission to take out internet threats and restore digital trust.

As the AI-native digital risk protection platform, Outtake delivers unified detection, investigation, and response across the full threat surface — protecting brands, executives, products, and locations from impersonation, AI-generated deception, and AI agent security risks. In an era where coordinated, industrial-scale attacks move faster than human response, Outtake gives organizations the agentic capability to stay ahead of threats, not just react to them.

www.outtake.ai