



SENTINEL NEXUS

SENTINEL NEXUS | COMMVAULT

# FROM AI ADOPTION TO AI RECOVERY: STAYING CLEAN AND PROVABLE

The AI Resilience Benchmark 2026 • Field data from 107 security executives

## Independent, executive-level field data on whether organizations can recover the AI their business now runs on.

AI systems are now production infrastructure. They run critical business processes, make decisions inside operational workflows, and act autonomously across the systems that companies depend on. But an AI system is not a server. It is models, weights, training data, pipelines, and agent workflows, and it fails differently: silently, through its data, at machine speed. Traditional recovery assumes you can restore to a known-good state and verify you are healthy again. AI breaks both assumptions.

This report's authority comes from who answered. One hundred seven unique, validated respondents: CISOs and senior security executives from the Sentinel Nexus community, answering a 17-question instrument across eight domains. This is the right data from the right people: the executives accountable for AI resilience, describing their own organizations to their peers. Page 5 says more about why that matters.

The report is organized around one question that every act returns to: **can you get it back, clean?**

Three pressures make the question urgent. Machine speed: AI failures propagate in seconds, faster than any human-paced response can follow. A new surface: models fail in ways servers never did, and a poisoned model runs as smoothly as a clean one. A whole estate: recovering AI means recovering an interdependent set of components, and the business processes running on them.

The pages that follow will show that today, nine organizations in ten have never proven the answer is yes.

### CONTEXT

**94%**

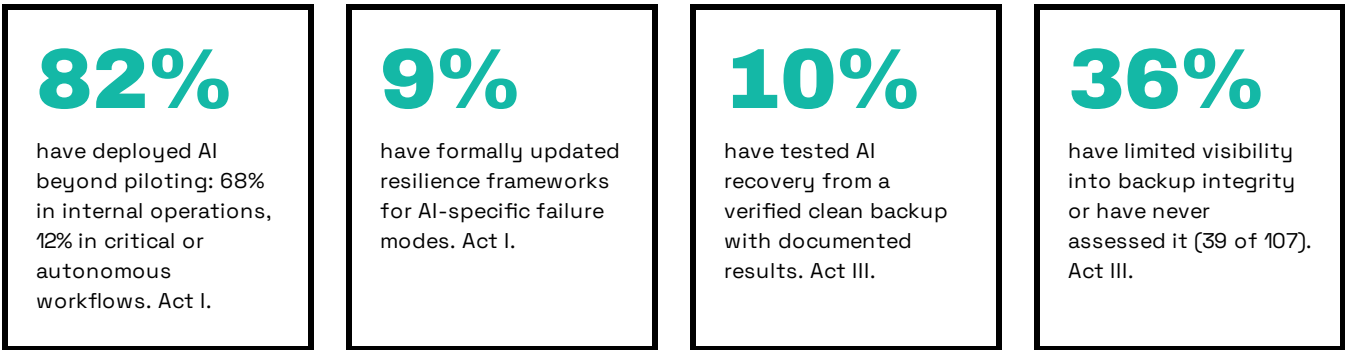
of global leaders cite AI as the most significant driver of change in cybersecurity.<sup>12</sup> Almost none of the frameworks behind that number address recovery.

### METHOD

107 validated respondents.  
17 substantive questions across eight domains: deployment, maturity, data integrity and governance, recovery confidence, testing, ownership, business continuity, and metrics.

# THE STORY IN FOUR NUMBERS

Eighty-two percent of organizations have deployed AI beyond piloting. Nine percent have a resilience program built for how AI fails. Ten percent have proven they can recover one. One in three couldn't tell you if the data they recovered was clean.



The gap those numbers describe is one of construction. What the organizations in this panel have not yet built, in overwhelming numbers, is the resilience program that production status demands. Frameworks predate the systems they are supposed to protect. Exercises rehearse everything except the failure class that moves fastest. And the capability at the center of it all, proving that a restored system is clean, remains immature: rare in these organizations, thin in the standards, and unsettled in the research.

The report unfolds in five acts.

- **Act I, The Estate**, establishes the reframe everything else depends on: AI is not a tool your organization uses; it is infrastructure your business runs on, and 82% have deployed it beyond piloting while only 9% have built the program that status demands.
- **Act II, Machine Speed**, puts a clock on the problem: failures propagate in seconds, and 54% of organizations would face them with pre-AI instruments and reflexes never trained on this failure class.
- **Act III, The Restore Problem**, is the reason this report exists: even a successful restore can bring back poisoned data no one can detect, and only 10% have proven they can recover an AI system at all.
- **Act IV, Whole Estate**, shows why even a solved verification problem would die inside today's organizations: ownership, planning, and the leadership chain are missing at every level where they must exist.
- **Act V, The Contribution**, states what these 107 leaders have put on the record, and why no one else could have.

Through it all, one question: **can you get it back, clean?**



The five-act bar above repeats at each section opener to mark your place in the argument.

# THE HEADLINE GAP

## Adoption outran the defensive half of the dual-use equation.

The research community spent 2025 establishing that generative AI is simultaneously an attack surface and a defense: a dual-use technology that expands what defenders can do and what attackers can reach in the same motion.<sup>2</sup> That framing was theoretical. This benchmark measures how far apart the two halves have drifted inside real organizations, at the desk of the executive accountable for closing the gap.

AI exposure is broad, deep, and already inside critical paths. Eighty-two percent of organizations have deployed AI beyond piloting. Sixty-eight percent run it in internal operations. Twelve percent have it embedded in critical or autonomous workflows. That 12% matters most: when an agent participates in a critical workflow, the workflow inherits the agent's failure modes. Agents act autonomously across systems, identities, and applications, which turns AI adoption from a software statistic into an operational dependency.<sup>3</sup>

The defensive half never got built. Nine percent of organizations have formally updated their resilience frameworks for AI-specific failure modes: data poisoning, model compromise, agentic failure. Thirty-eight percent still operate on frameworks that predate AI entirely. Only 23% have any formal cross-functional AI resilience program at all. The imbalance the taxonomy literature theorized across more than 100 publications now has a field measurement: an 82-to-9 ratio between the systems in production and the programs built to bring them back.

82% have deployed AI beyond piloting

Where it runs (multi-select): 68% internal operations · 25% customer-facing · 18% at scale · 12% critical or autonomous workflows

38% still operating resilience frameworks that predate AI

23% have any formal cross-functional AI resilience program

9% have frameworks formally updated for AI-specific failure modes

The cliff: adoption at 82%, purpose-built resilience at 9%.

### THE DUAL-USE FRAME

A taxonomy published in 2025, spanning more than 100 publications, established generative AI as both attack surface and defense.<sup>2</sup> This page is that imbalance's first measurement at the executive level: organizations built the offense-exposed half and skipped the defensive half, 82-to-9.

### WHY AGENTIC MATTERS

Adversarial research on autonomous agents shows compromise crossing identities, applications, and stacks without human checkpoints.<sup>3</sup> Agentic deployment inside critical workflows is dependency, not experimentation.

# WHO ANSWERED

Everything in this report comes from the people who own the problem.

The panel behind this benchmark is 107 unique respondents: CISOs and senior security executives drawn from the Sentinel Nexus community, the operators accountable for security and resilience at organizations where the stakes are highest. Respondents span the revenue bands, headcounts, industries, and geographies detailed in the full-findings appendix, alongside every question's exact wording and per-question sample size.

The instrument itself: 17 substantive questions across eight domains. Deployment, to establish what is actually running. Maturity, to establish what frameworks govern it. Data integrity and governance, to establish whether the inputs were ever trustworthy. Recovery confidence, testing, ownership, business continuity, and metrics, to establish whether anyone could bring it back and prove it.

Candor is the differentiator, and the reason these numbers diverge from the headlines. While a vendor-sponsored survey published in 2026 put confidence that disaster recovery plans cover agentic AI at 94%,<sup>13</sup> our panel reports 10% proven recovery. Those are different questions asked of different rooms, and that is the point. Confidence is what most surveys collect. Proof is what this panel was asked for, and proof is rare, because this instrument asked the harder question. Every number in this report should be read in that light. These are the figures security leaders give when they are asked as practitioners among peers, about capabilities only they are positioned to describe.

Confidence is also an unstable thing to measure. Proof is not, which is why this benchmark asked for it.

That candor is also the report's contribution. Academic research cannot reach a validated panel of accountable executives. Vendor research reaches a different audience, with a different frame. Only a peer community can put the accountable executives on the record, and this one did: the right data, from the right people.

## CONFIDENCE VS. PROOF

**94%** **10%**

confident  
their DR  
plans cover  
agentic AI, in  
a vendor-  
sponsored  
2026  
survey<sup>13</sup>

proven AI  
recovery, in  
this  
independent  
benchmark

## THE EIGHT DOMAINS

Deployment · Maturity ·  
Data integrity & governance  
· Recovery confidence ·  
Testing · Ownership ·  
Business continuity ·  
Metrics

## PANEL

107 unique validated  
respondents. CISOs and  
senior security executives.  
Demographic distributions  
published in the appendix.



MICHAEL MARSILIO • CISO, PARADIES LAGARDÈRE

**“The most successful organizations are treating AI resilience as an enterprise discipline rather than a technology initiative. As AI becomes embedded in workflows, productivity platforms, analytics, and decision support, it must be governed, monitored, secured, and recoverable like any other business-critical capability.”**

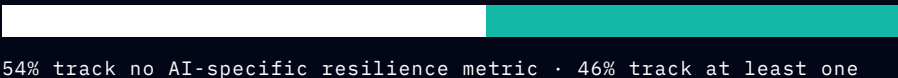
# FAILURE AT MACHINE SPEED, RECOVERY AT HUMAN SPEED

Agents act across identities, applications, and stacks in seconds. A single compromise spreads before a human ever sees the alert.<sup>3</sup> **Can you get it back, clean, in time?**

Recovery is one leg of resilience, not the whole of it. Cutting off a failing agent's access and falling back to manual or non-AI processes does not bring the estate back. Once the failure is contained, the question that remains is the one this report measures. Response speed comes down to two things: the instruments that let you see a failure in real time, and the reflexes that let you act on what you see. This act examines both, and finds both frozen in the pre-AI era.

Start with the instruments. Twenty-seven percent of the panel reported that they measure nothing at all for AI resilience. Thirty-seven percent reported legacy RTO and RPO targets, the recovery-time and recovery-point language of the server era, with no AI-specific metrics alongside. Counted at the respondent level, 54% of organizations track no AI-specific resilience metric of any kind, and only 46% track even one. A majority would have to detect, scope, and steer a machine-speed AI failure using instruments that cannot describe one. Their dashboards speak a language the failure does not.

**Metrics maturity across the panel** (respondent level)



**54%**

track no AI-specific resilience metric. Only 21% track training-data integrity and 21% track data lineage: the two measurements the restore problem in Act III turns on.

The literature reached the same verdict from the detection side. Peer-reviewed research on security operations concludes that human-centric operating models are untenable against machine-speed threats.<sup>6</sup> But detection at least has an AI-augmented roadmap: the industry is actively rebuilding the SOC around it. Recovery has neither roadmap nor metrics. The recovery language in the draft standards is the thinnest part of the entire stack.<sup>9</sup> Our data shows organizations are further behind on recovery than they are on detection.

## THE VERDICT ON SPEED

Peer-reviewed SOC research from 2025 and 2026 concludes that human-centric security operations cannot hold against threats moving at machine speed.<sup>6</sup> The operating model, not the analyst, is the constraint.

## TWO LAYERS OF THIS ACT

**Instruments** (this page): 54% could not see a machine-speed failure in the terms their dashboards speak.

**Reflexes** (page 9): the muscle that makes recovery fast was never trained on this failure class.

# THE FIELD NAMES ITS OWN GAP

We asked security leaders what worries them most about recovering from an AI compromise. The #1 answer is the most clear-eyed finding in this benchmark.

Analysts have spent two years diagnosing the AI readiness gap from the outside. This page is different: it is the field naming its own gap. When we asked security leaders to name their greatest concerns about recovering from an AI compromise, the most-cited answer, at 51% of the panel, was that recovery procedures exist for their IT systems but not for their AI. The focus is on the gap itself, which is exactly where it belongs.

**51%** · Recovery procedures exist for IT systems, but not for AI (most cited)

**34%** · Restoring to corrupted or poisoned data

**34%** · No way to verify the integrity of recovered data

**26%** · Never tested AI recovery

**22%** · Responsibility for AI recovery unassigned

**20%** · Don't know which AI systems are critical

Recovery concerns, share of panel citing each (multi-select). The two 34% findings are examined on page 12. The remaining concerns preview the rest of this report: untested procedures, unassigned ownership, unmapped criticality.

The concern is rational, because the objects of recovery changed and the procedures did not. RTO and RPO presuppose a system whose restoration you can watch and verify. A server boots or it doesn't; services come up or they don't; the checks go green or they don't. No equivalent acceptance test exists for a restored model. A model that comes back poisoned performs exactly like a model that comes back clean, and none of the frameworks these respondents operate under supplies a test that could tell the difference.<sup>9</sup>

The 37% of organizations running legacy metrics only, reported on page 7, is this same gap seen from the instrumentation side. Recovery procedures and the metrics beneath them were built for an earlier estate, and the debt has compounded.

## WHERE THE STANDARDS STAND

NIST's Cyber AI Profile (IR 8596) is a preliminary draft, released December 2025; its first comment period closed in January 2026. Its recovery guidance is the thinnest layer of the stack, and the draft explicitly solicited informative references from the field.<sup>9</sup>

## WHAT THIS MEANS

When a majority of accountable executives cite "we have no procedures for this" ahead of any named threat, the field is seeing its gap clearly. The problem is not knowledge of the adversary. It is that recovery practice has not caught up with the AI estate.

# REHEARSING FOR THE WRONG DISASTER

**Speed is trained, not installed. And AI isn't in the training.**

A Delphi study of incident-response experts reached a conclusion this benchmark now independently replicates at n=107: AI-era resilience is a practiced human capability, not a property of tooling.<sup>4</sup> You cannot buy your way to a fast recovery. Exercises are the mechanism that converts plans into speed, which makes the panel's exercise data the most predictive finding in this act.

The training regimen excludes the exact failure class that moves fastest. Thirty-nine percent of organizations run resilience exercises that exclude AI entirely. And 39% exclude AI from tabletops and post-incident reviews, the two forums where organizations are supposed to learn from failure before and after it happens. And even where AI is included, the cadence is built for the old clock: the most common exercise frequency is annual, at 19%, and only 21% exercise quarterly or better. An annual rehearsal, against a failure class that propagates in seconds.

Exercises that exclude AI

Exercises exclude AI entirely · **39%**



Tabletops exclude AI · **39%**



Exercise cadence (with AI)

Annual (most common) · **19%**



Quarterly or better · **21%**



Rehearsal, not confidence, is what separates capability from assumption, and the field's own confidence numbers prove it. The single most common recovery-confidence answer, at 37%, is "somewhat confident: procedures exist but are untested." Untested procedures are indistinguishable from wrong ones until the incident. In total, 88% of this panel (94 of 107) operates on AI recovery procedures that are missing, non-specific, or untested: 22% have no formal procedures for AI at all, 28% hold procedures that don't address it, and 37% hold procedures no one has tested. Proof lives with the remaining 13 respondents, the 12% who are very confident in tested, documented procedures. On a separate question, only 10% have tested AI recovery from a verified clean backup and documented the results.

**An organization that has never rehearsed AI failure will meet its first one at machine speed, unrehearsed.**

REPLICATED FINDING

A 2026 Delphi study concluded that AI-enabled incident response is a skills and capability problem: a practiced human capability rather than a tooling property.<sup>4</sup> This benchmark reaches the same conclusion from field data, independently, at n=107.

THE MODAL ANSWER · 37%

"Somewhat confident: procedures exist but are untested." The most common recovery-confidence response in the panel, and the definition of assumption mistaken for capability.

JONATHAN BARRIOS · CISO, VINDICIA

*"AI governance defines how AI should behave. AI resilience ensures it continues to operate safely, or fail safely when conditions aren't ideal. But to me the two are inseparable."*

# WHAT YOU'D ACTUALLY HAVE TO RECOVER

Recovering a server means restoring files and watching it boot. Recovering an AI system means restoring an estate: model weights, training data and retrieval corpora, fine-tunes, pipelines, agent workflows, and the identities they act through. Then proving all of it is clean.

Acts I and II established that recovery is necessary and that it would be slow. This act establishes something categorically worse: it may be impossible to verify. First, the anatomy, because the phrase “AI system” hides how many distinct things must come back for recovery to mean anything.

An AI system is a set of interdependent recoverable components, each with its own failure mode. The taxonomy below follows the agentic-threat and AI-lifecycle literature.<sup>3, 11</sup>

|                           |                                                                                                                             |
|---------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| Model weights             | Tampering. Modified parameters that alter behavior while the system appears healthy.                                        |
| Training & reference data | Poisoning. Corrupted training sets, RAG corpora, vector stores, and agent memory that silently reshape what the model does. |
| Pipelines & configs       | Corruption. Broken or altered build, training, and deployment paths.                                                        |
| Agent workflows           | Hijack. Autonomous actions redirected across the systems agents touch.                                                      |
| Identity & access surface | Privilege abuse. The credentials and permissions agents act through, turned against the estate.                             |

The anatomy holds whether an organization trains its own models or consumes them as a service. An API consumer never touches base weights but owns everything else here: the fine-tune data, retrieval corpora, pipelines, agents, and keys. The estate shifts; it does not shrink. The failures also differ in persistence: a hijacked agent leaves the model clean but the business state dirty, so the recovery object is everything the agent touched, while poisoning and tampering live in the components themselves. Both roads end at restore-and-verify, of different objects.

These are live attack surfaces, not hypotheticals. Joint NSA, CISA, and NCSC-UK guidance from May 2025 confirms adversaries are actively poisoning AI systems across sectors by corrupting the data that trains them.<sup>10</sup> The estate is already being targeted at exactly the layer that recovery must be able to verify: the data.

The next three pages walk the restore problem in causal order: captured, trusted, provable. **Can you get it back, clean?**

**ACTIVE TARGETING**

Joint NSA / CISA / NCSC-UK AI Data Security guidance, May 2025: adversaries are actively poisoning AI systems across sectors.<sup>10</sup> Data is the targeted layer.

**THE TAXONOMY**

The component groups at left follow the agentic-threat literature and ENISA's AI-lifecycle taxonomy, which maps threats across every stage from data collection through deployment.<sup>3, 11</sup>

# THE ESTATE ISN'T EVEN CAPTURED

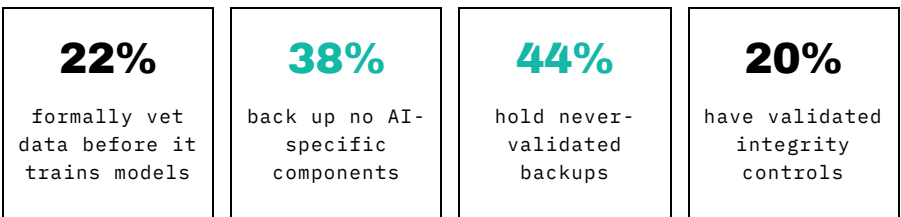
Before asking whether recovery would work, ask whether there's anything to recover from, and whether it was ever trustworthy to begin with.

The restore problem compounds across the full data lifecycle. The data is unvetted going in, uncaptured in backups, unvalidated where it is captured, and unverifiable at restore time. Each stage inherits the deficit of the one before it.

Start with capture. For a large share of organizations, the AI estate is absent from backups entirely: 38% back up infrastructure but no AI-specific components. Mapped against the anatomy on page 10, that means model weights, training data, and agent workflows, the components adversaries are actively targeting,<sup>10</sup> are the components most likely to have no restore point at all. For these organizations, the question “would your restore work?” is moot. There is nothing AI-shaped to restore. The pattern inverts priority. Model weights, the one component whose compromise is invisible in operation, are the least captured: 22%. Fifteen percent formally back up nothing at all.

The trust deficit starts upstream of the backup, because the data was never verified going in either. Only 22% of organizations have a formal, automated process to vet data before it trains or fine-tunes models. Forty-two percent rely on manual review. Twenty-eight percent have effectively nothing: data owners flag issues when they happen to notice them, or there is no process at all. Downstream, only 20% are very confident in their backups, with validated data-integrity controls; 44% are merely somewhat confident, holding backups they have never validated. This layer, unlike the model, is buildable with tools the field already owns: a cryptographic hash can prove a restored corpus matches its snapshot. What no hash can help prove is that the snapshot was clean when it was taken, which is why provenance has to be captured going in.

The lifecycle trust chain



The lifecycle runs on trust at every stage, and that is exactly what the Software Engineering Institute warns against: chain-of-custody and provenance controls must exist before compromise, because retroactive tamper-detection is becoming less effective.<sup>8</sup> The 21% of organizations tracking data lineage, reported on page 7, shows how far those controls are from existing. When the incident comes, trust cannot be reconstructed after the fact. It either was built into the lifecycle, or it wasn't.

BEFORE, NOT AFTER

Carnegie Mellon's Software Engineering Institute, 2026: chain-of-custody and provenance controls must precede incidents. Retroactive tamper-detection is becoming less effective.<sup>8</sup>

UPSTREAM VETTING

22% formal, automated vetting before training  
42% manual review  
28% effectively nothing: owner flags issues, or no process at all

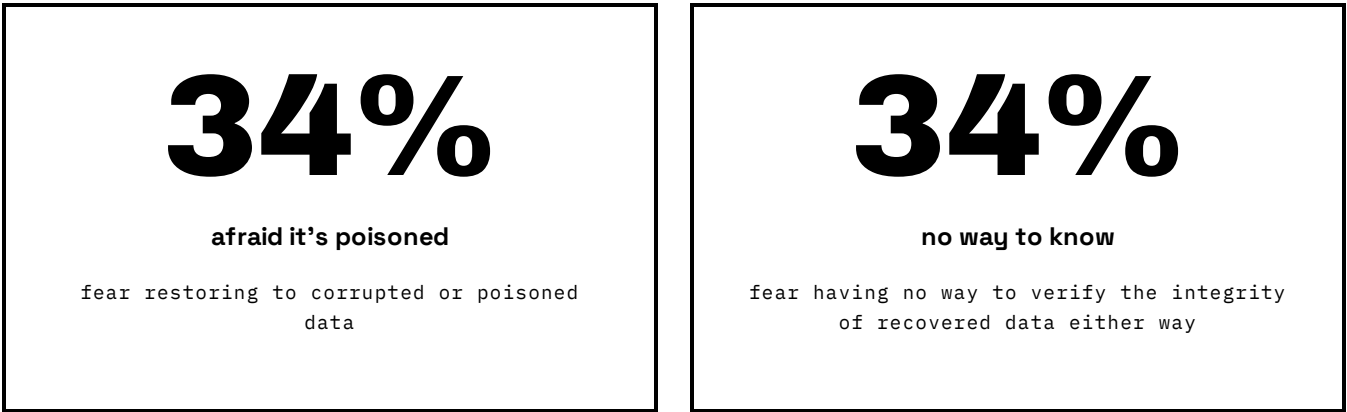
READ IT FORWARD

Unvetted → uncaptured → unvalidated → unverifiable. Each stage inherits the one before it. Page 12 shows where the chain ends.

# THE FEAR IS THE RESTORE

The scariest finding in this benchmark isn’t about being attacked. It’s about the moment after.

Poisoning is silent, and it may predate the oldest backup you’ve retained, so a clean point may not exist in anything you’ve kept. And a poisoned model runs as smoothly as a clean one, so the restore itself cannot tell you whether it worked.



10% have tested AI recovery from a verified clean backup, with documented results



Half the panel named at least one side of the same missing capability. One side fears the restore brings the poison back. The other fears there is no way to know either way. Both fears point at the same absent discipline, verification of a clean state at restore time, which the next page shows does not yet exist anywhere.

And almost no one has crossed from fear to proof. Ten percent of organizations have tested AI recovery from a verified clean backup and documented the results. Nine in ten organizations running AI in production have never demonstrated they can get it back, clean or otherwise.

FRED CLAYTON • CISO, AKUMIN

*“If [an AI system] goes down or something bad gets injected into it, how do you recover and get back the view you need? That’s the gap, in all honesty.”*

## WHY NOBODY CAN SAY “CLEAN”

**This is not a problem an engineering team can fix on its own. The research community has tested the available remedies, and they fall short.**

If verifying a clean state were merely hard, some party would be closing the gap: researchers with an algorithm, standards bodies with a discipline, or vendors with a product. This page lays out what each offers, and where each stops.

### ACADEMIA • THE MODEL-LEVEL PATH HAS BEEN TESTED AND DOES NOT WORK

The most rigorous evaluation to date, presented at ICLR 2025, ran eight state-of-the-art machine unlearning algorithms against data poisoning across language and vision tasks. None fully reverted the poisoning. Full retraining, the only reliable remedy, is frequently impractical at modern model scale.<sup>7</sup> Carnegie Mellon’s Software Engineering Institute reached the same verdict independently: recovery from poisoning remains meager at best, and the field’s posture has collapsed to prevention-only.<sup>8</sup> If the poison is in the model, science currently offers no way to get it out short of starting over.

### STANDARDS • THE DISCIPLINE IS UNWRITTEN

The standards stack contains no formalized restore-time verification. NIST’s Cyber AI Profile is a preliminary draft whose first comment period closed in January 2026 with no restore-time discipline in it.<sup>9</sup> ENISA’s threat landscape maps the attacks across the AI lifecycle but supplies no recovery discipline to answer them.<sup>11</sup> Nowhere in the stack is there an acceptance test for a restored model: a procedure by which an organization could demonstrate, to itself or to a regulator, that what came back is clean.

### VENDORS • ACCESS WITHOUT INDEPENDENCE

Existing survey data on organizational recovery practice is almost entirely vendor-sponsored.<sup>13</sup> Nearly every datapoint in circulation about whether organizations can recover their AI comes from companies selling recovery. Whatever its accuracy, it cannot be independent. And it measures the wrong thing: the same survey carries both 94% confidence and, on another question, a majority not confident or only somewhat confident about the same scenarios. Confidence data bends with the questionnaire. Capability data is what the field is missing.

Model-level recovery fails. Standards are unwritten. Vendor data is conflicted. The discipline of verifying a clean state at restore time does not exist, and the 107 executives in this benchmark are the first independent evidence of what that absence looks like inside organizations. It looks like half a panel afraid of the restore or unable to know, and one in ten with proof.

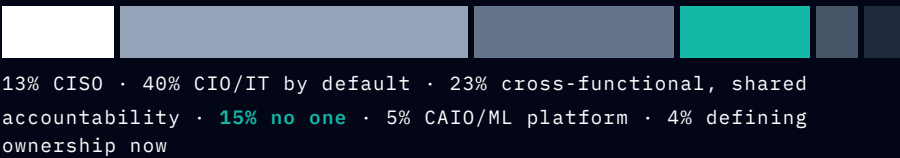
# AN ESTATE WITH NO OWNER

Suppose the verification problem were solved tomorrow. Who in your organization would implement it? **Can you get it back, clean?**

Act III ended at the edge of what science can currently do. This act begins with a thought experiment that moves the problem somewhere less comfortable: assume the clean-state verification discipline arrived tomorrow, fully formed. In most of the organizations in this panel, it would die on arrival, because nobody is cast in the role of implementing it.

Accountability for AI resilience has not been assigned. It has been defaulted. Thirteen percent of organizations place AI resilience with the CISO. Forty percent place it with the CIO or IT operations, by default rather than by decision. Fifteen percent have assigned it to no one at all. Twenty-three percent report a cross-functional team with shared accountability. Five percent place it with a Chief AI Officer or an ML platform team. A default is not an assignment, and shared accountability is not assigned accountability: until a named owner is set, the cross-functional answer is another form of the same vacuum. The 40% who inherited the duty hold it without the mandate, the budget, or the security context to execute it.

Who owns AI resilience



The gap between where the literature puts ownership and where the field puts it is the governance gap, made concrete. The CISO-governance research positions the security executive at the center of generative AI oversight.<sup>5</sup> Ownership of AI resilience is the natural extension of that role. The systematic-review literature concludes that governance mechanisms are lagging AI capability,<sup>1</sup> a finding usually read abstractly. This data gives it a number: the nominated owner actually holds the role in roughly one organization in eight. The governance gap is not an abstraction. It is an ownership vacuum, and everything else in this act flows downhill from it.

THE NOMINATED OWNER

Peer-reviewed governance research places the CISO at the center of generative AI oversight.<sup>5</sup> In the field, the CISO holds AI resilience in 13% of organizations.

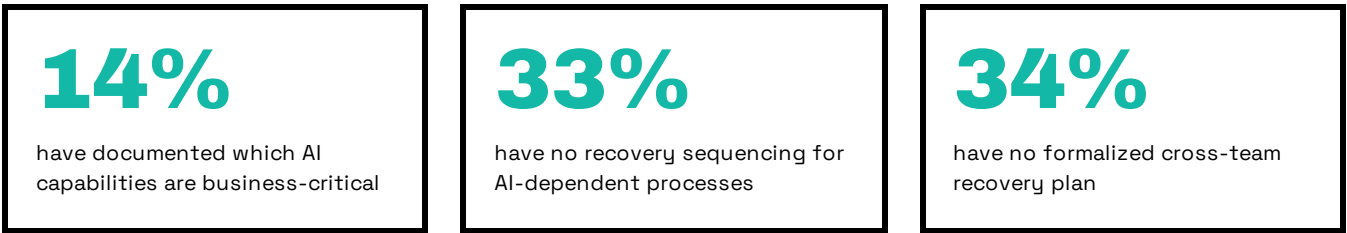
THE GOVERNANCE GAP

A 2025 PRISMA systematic review concluded that generative AI capability is outpacing governance.<sup>1</sup> In this panel, that gap has a shape: fragmented or absent ownership.

# NO MAP, NO SEQUENCE

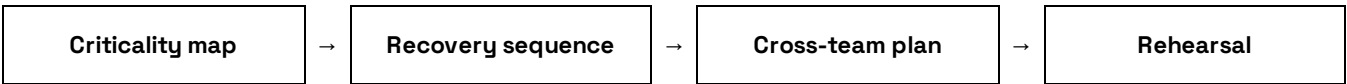
Recovery starts with knowing which AI systems must come back first. Few organizations have written that down.

Recovering AI systems is ultimately recovering the business processes running on them. Those processes have not been mapped to the AI they depend on, which means the operational layer of recovery is missing along with the executive layer the previous page described.



The prerequisite artifacts of estate recovery are missing at every step, and the deficits chain. Only 14% of organizations have documented which of their AI capabilities are business-critical, so 86% would begin an incident without knowing what matters most. Thirty-three percent have no recovery sequencing for AI-dependent processes, so even organizations that know what matters have not decided what comes back first. Thirty-four percent have no formalized cross-team recovery plan, so the teams that would have to execute the sequence — security, infrastructure, data, and the business — have never agreed on who does what. No criticality map means no sequence. No sequence means no plan. No plan means improvisation, at machine speed.

The missing discipline is mundane, which is the good news. Unlike the verification layer in Act III, the ordered-restore chain requires no unsolved science. It is standard business-continuity method: identify what is critical, decide the order of restoration, assign the teams, and rehearse it. The only thing novel about it is the surface it has never been applied to. Every organization in this panel already runs this discipline for its conventional infrastructure. Extending it to the AI estate is a decision, not a research program.



The ordered-restore chain. Intentionally plain: standard continuity method, applied to a new surface.

**JONATHAN BARRIOS · CISO, VINDICIA**

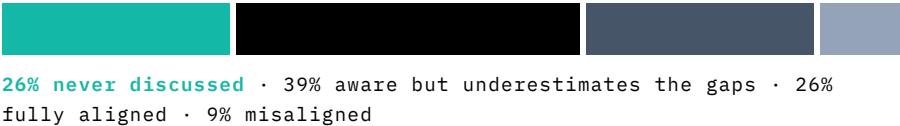
*“In practice, business owns the risk, security governs it, IT operates it and makes it available and recoverable. The key is moving away from siloed ownership. Success is measured by the business continuity capabilities.”*

# THE CHAIN THAT WOULD HAVE TO APPROVE ALL THIS

Every framework in this space assumes an informed reporting chain. Ours found it missing or degraded in two-thirds of organizations.

Ownership and planning both depend on a third level: the board-to-CISO chain that receives the findings, understands the stakes, and resources the response. In a quarter of the organizations in this panel, that conversation has never happened at all. Twenty-six percent have never had a substantive leadership conversation about AI resilience readiness. Thirty-nine percent report that leadership is aware but underestimates the gaps. Only 26% report full alignment between what security knows and what leadership acts on.

Leadership alignment spectrum



This finding explains, in advance, why framework uptake will lag. Standards like the Cyber AI Profile presuppose an informed reporting chain as their adoption mechanism: findings travel up, mandates and budget travel down.<sup>9</sup> Where that chain is absent or degraded, which describes roughly two-thirds of this panel, frameworks arrive with no one to catch them. The best-written guidance in the world does nothing inside an organization where the conversation it depends on has never been held.

**Whole-estate resilience is organizational before it is technical, and the fix starts with an org chart and clear board conversations rooted in the consequences of failure.**

THE DORA SIGNAL

Organizations subject to DORA or NIS2, the two regimes that specifically mandate operational resilience testing, outperform everyone else on exactly the practiced capabilities this report finds missing:

Tested AI recovery, documented



28% under DORA/NIS2 vs. 5% of everyone else

Exercise quarterly or better



40% vs. 15%

Generic compliance shows no lift. Organizations under NIST CSF, HIPAA, SOX, or ISO 27001 updated their resilience programs at 9%, no better than the rest of the panel. Checkbox compliance doesn't build the capability. Resilience-specific mandates with testing requirements correlate with it.

For US readers: NIST's resilience-specific guidance is taking a similar shape.<sup>9</sup>

*Caveat: n=25 for the DORA/NIS2 group, skewing toward large financial institutions, so resources confound regulation. A pattern, not a proven effect.*



MICHAEL MARSILIO • CISO, PARADIES LAGARDÈRE

**“Leaders are asking fewer questions about AI capabilities and more questions about reliability, accountability, governance, data protection, vendor dependence, and operational continuity. They want confidence that AI can scale safely and continue supporting the business when disruptions occur. Resilience is becoming a foundational pillar of AI governance. Effective governance must address not only acceptable use, risk, and compliance, but also availability, recoverability, vendor concentration, data dependencies, and human oversight.”**

# THE EVIDENCE THAT WAS MISSING

Three groups could have produced this evidence. Two structurally cannot. **Can you get it back, clean?**

Every claim in this report rests on a kind of evidence the field has been missing: independent measurement of organizational recovery practice, taken from the executives accountable for it. It is worth being precise about why that evidence did not exist before now, because the answer explains why it was this community's to produce.

Each existing evidence source stops exactly where its structure forces it to stop. Academia measured the algorithm and found it fails: unlearning cannot revert poisoning, and recovery from it is meager at best.<sup>7,8</sup> Reaching a panel of accountable security executives is a different kind of access, built on a standing community and years of trust, and without it the organizational practice those findings imply had gone unmeasured. Vendors have the access but not the independence: nearly every existing datapoint on organizational AI recovery comes from companies selling recovery.<sup>13</sup> And standards bodies have the mandate but not the data. NIST's Cyber AI Profile is a preliminary draft that issued a call for informative references,<sup>9</sup> which is a standards body saying, in its own procedural language, that it asked for exactly this.

## THE STANDARDS GAP

NIST IR 8596, the Cyber AI Profile: preliminary draft, December 2025. First comment period closed January 2026; the profile is still in development.<sup>9</sup>

## WHAT ONLY THIS PANEL COULD SAY

107 accountable executives answering the questions only they can: the right data, from the right people.

### Academia

measured the  
algorithm, not  
the practice

### Vendors

access without  
independence

### Field evidence

this benchmark:  
independent data  
from the  
executives who  
own the problem

## CONCLUSION

# THE GAP IS PROOF, NOT RECOVERY

**Preparedness has been defined wrong.**

Every discipline this community has mastered — backup, failover, restore — rests on an assumption AI silently breaks: that you can tell when a system is healthy again. Restored servers announce their health. The machine boots, services come up, checks go green, and the incident commander can say with a straight face that the environment is recovered. Restored models announce nothing. A poisoned model performs exactly like a clean one, and half of this panel named that fact as a central fear, from opposite sides: afraid the restore brings the poison back, or afraid there is no way to know either way.

The science offers no exit after the fact. Machine unlearning cannot revert poisoning once it is in the model.<sup>7</sup> The controls that make a clean state provable must exist before the incident, not after it.<sup>8</sup> Which means the gap this benchmark measures is not, at its root, a recovery gap. It is a proof gap. Organizations cannot say “clean” because nothing in their estate was built to make clean a provable claim.

**PREPAREDNESS FOR THE AI ERA IS NOT ONLY THE ABILITY TO RESTORE. IT IS THE ABILITY TO PROVE WHAT YOU RESTORED.**

The work ahead is definable and assignable, and every piece of it maps to a gap this report has already measured. The verified clean state must be defined: what it actually means for an AI system, component by component. It must be instrumented, with the AI-specific metrics 54% of organizations currently lack. It must be owned, filling the executive vacuum in which 15% of organizations have assigned the duty to no one and 40% have defaulted it to teams without the mandate. It must be mapped, producing the criticality map 86% of organizations lack and the sequencing and cross-team plans a third have never written. And it must be rehearsed, at the cadence machine speed demands rather than the annual cycle that is most common today.

None of that requires waiting on new science, but none of it is solved by an org chart alone. Naming an owner is where the work starts, not where it ends: the definitions, the instrumentation, the maps, and the rehearsals each remain a program to build, not a memo to send.

**The 107 leaders in this benchmark have named the problem. Defining and proving the verified clean state is the work ahead, and it runs through the people accountable for the answer.**

**DR. LOVELIE MOORE · BISO, TOYOTA FINANCIAL SERVICES**

*“Resilience must be built into AI governance from the beginning, not added after an incident. Our governance conversations should address how AI systems will remain secure, reliable, and recoverable when models fail, data becomes compromised, vendors experience outages, or outputs cannot be trusted. Effective AI governance is not only about responsible use; it is also about ensuring the organization can continue operating when AI does not perform as expected.”*

# NINE IN TEN OF US CANNOT PROVE WE CAN RECOVER THE AI OUR BUSINESSES RUN ON. WHAT WOULD IT TAKE TO SAY “CLEAN,” AND MEAN IT?



## SOURCES

1. Radanliev P, Santos O, Ani UD. Generative AI cybersecurity and resilience. *Frontiers in Artificial Intelligence*, 8:1568360, 2025. doi:10.3389/frai.2025.1568360.
2. Mirtaheiri SL, Movahed N, Shahbazian R, Pascucci V, Pugliese A. Cybersecurity in the age of generative AI: A systematic taxonomy of AI-powered vulnerability assessment and risk management. *Future Generation Computer Systems*, 175:108107, 2026. doi:10.1016/j.future.2025.108107.
3. Radanliev P, Santos O, Maple C. Threats and vulnerabilities in artificial intelligence and agentic AI models. *Frontiers in Artificial Intelligence*, 9:1731566, 2026. doi:10.3389/frai.2026.1731566.
4. Senarak C. Cyber resilience skills for AI-enabled incident response in port operations. *Transportation Research Interdisciplinary Perspectives*, 2026. Two-round Delphi study.
5. Christiansen J. Navigating the GenAI frontier: The role of the CISO in GenAI governance. *Cyber Security: A Peer-Reviewed Journal*, Vol. 8, No. 4, pp. 321–336, 2025.
6. Mallu YR. SOC 2.0: Less triage, more treachery. *Cyber Security: A Peer-Reviewed Journal*, Vol. 9, No. 4, 2026.

## RESEARCH REPORT SUPPORTED BY



7. Pawelczyk M, Di JZ, Lu Y, Kamath G, Sekhari A, Neel S. Machine unlearning fails to remove data poisoning attacks. *ICLR 2025*. arXiv:2406.17216.
8. Metcalf R, Churilla M. Data poisoning in AI models: the case for chain of custody controls. SEI Blog, Carnegie Mellon University Software Engineering Institute, April 27, 2026. doi:10.58012/25yz-9b86.
9. NIST IR 8596, Cyber AI Profile. Preliminary draft, December 2025; initial comment period closed January 2026.
10. NSA AISC, CISA, FBI, ASD's ACSC, NCSC-NZ, NCSC-UK. AI Data Security: Best Practices for Securing Data Used to Train & Operate AI Systems. Joint Cybersecurity Information Sheet, May 2025.
11. ENISA Threat Landscape 2025 and ENISA AI Threat Landscape 2026 (two documents, cited jointly).
12. World Economic Forum. Global Cybersecurity Outlook, 2026.
13. Peer Insights on AI Adoption and the Disaster Recovery Gap. *CIO Market Pulse*, sponsored content by Keepit; Foundry online survey of 301 IT decision-makers (AU, FR, DE, NZ, UK, US; companies of 1,000+ employees), fielded Nov 19–Dec 15, 2025. Distinct from an earlier Keepit/Foundry survey wave with different fielding dates.



THE AI RESILIENCE BENCHMARK 2026

# FULL-FINDINGS APPENDIX

All 23 survey questions with exact wording, full response distributions, and n per question.

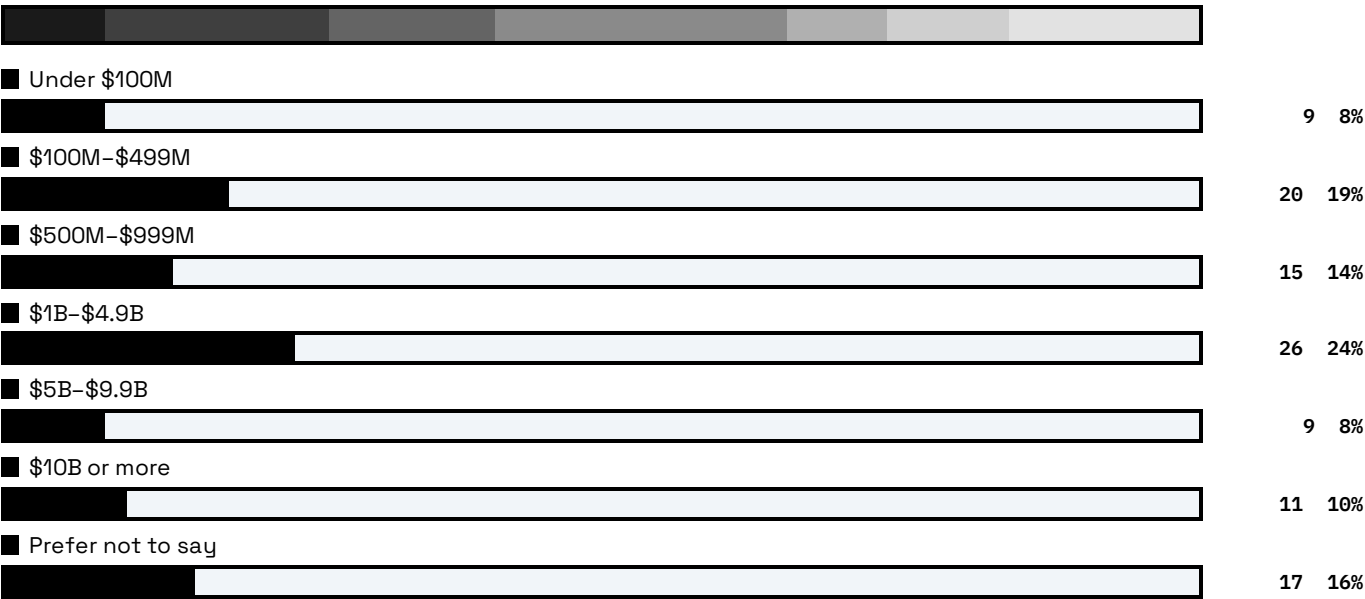
# FULL-FINDINGS APPENDIX

All 23 survey questions with exact wording, full response distributions, and n per question. Panel: 107 validated respondents, CISOs and senior security executives from the Sentinel Nexus community. Questions 1–6 cover demographics and regulatory context; questions 7–23 are the substantive instrument. Bars are scaled to 100% of respondents answering each question; single-select questions carry a distribution strip showing the full split at a glance.

■ Answers highlighted in teal are the figures featured in the report.

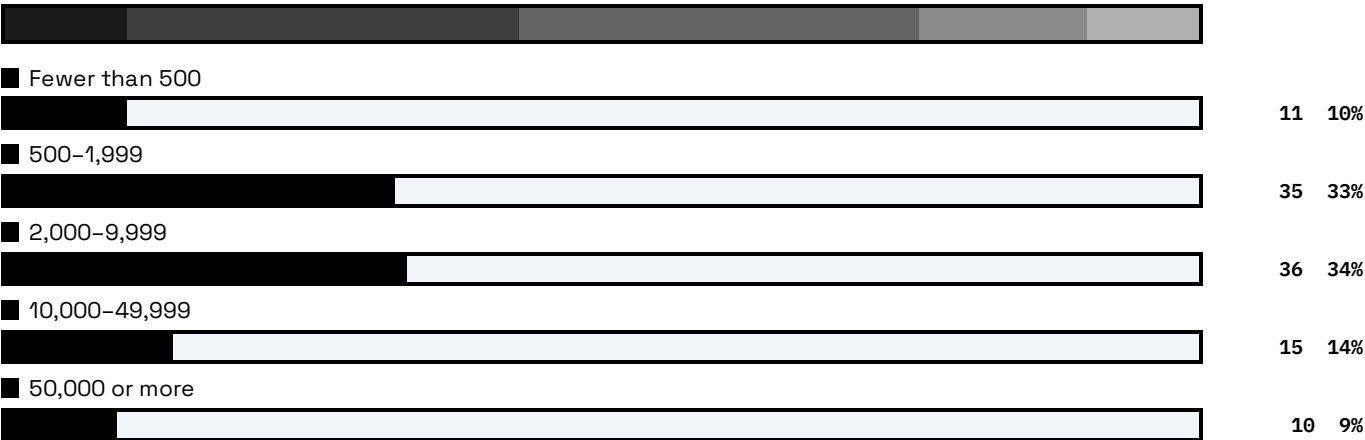
DEMOGRAPHICS • N = 107

1. What is your organization’s annual revenue?



DEMOGRAPHICS • N = 107

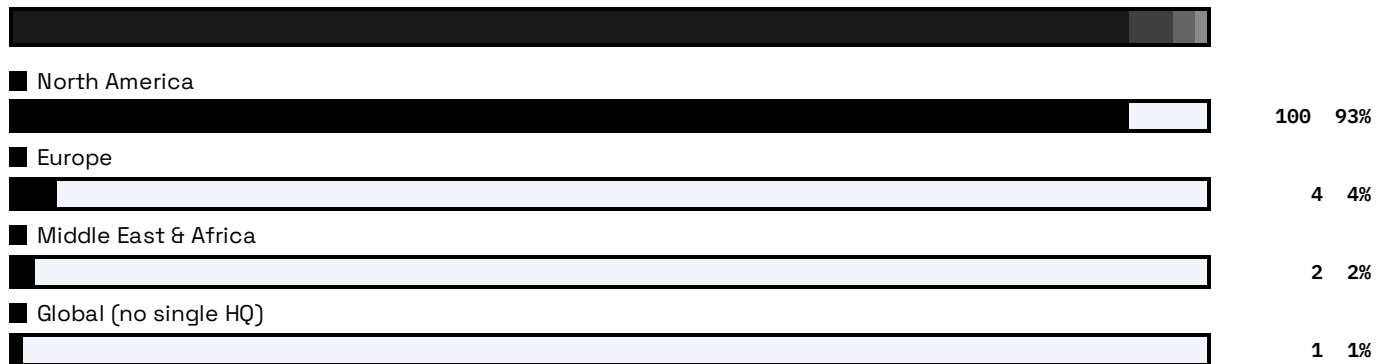
2. How many employees does your organization have globally?



---

DEMOGRAPHICS • N = 107

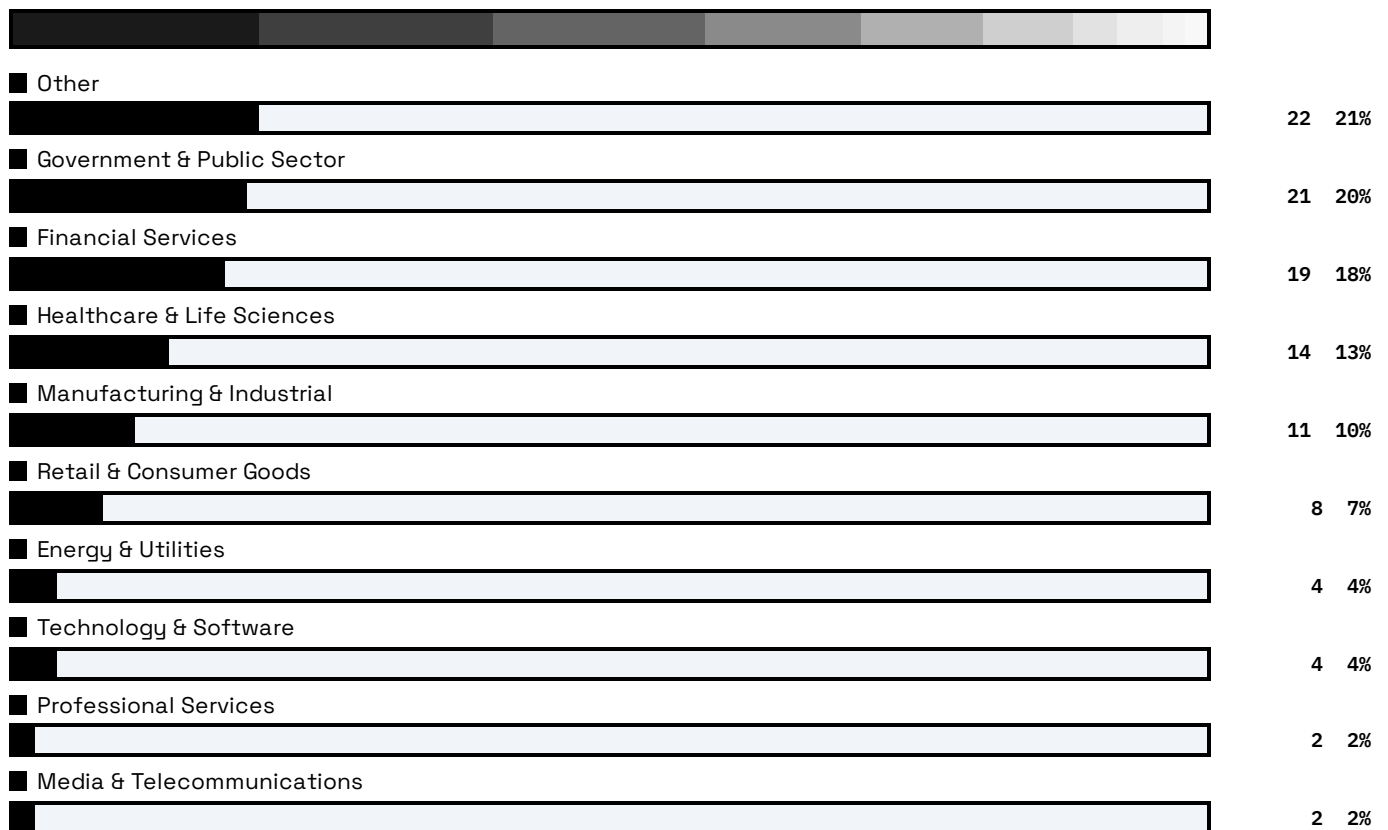
### 3. Where is your organization headquartered?



---

DEMOGRAPHICS • N = 107

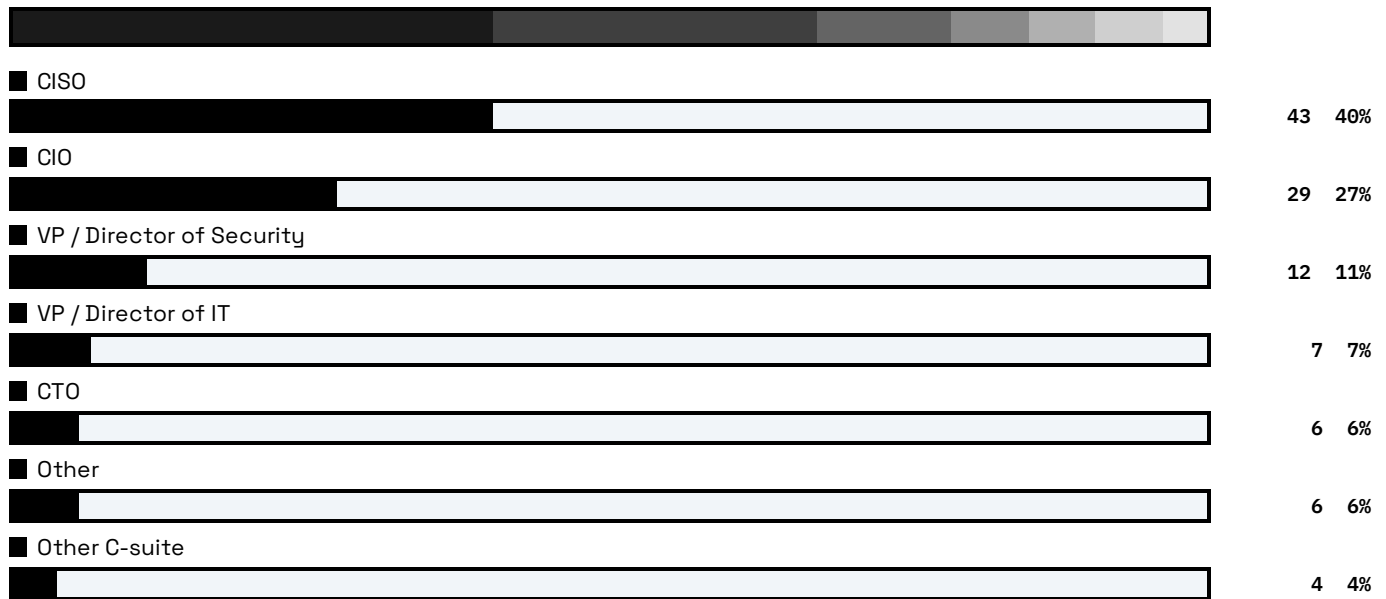
### 4. What is your organization's primary industry?



---

DEMOGRAPHICS • N = 107

## 5. What is your job title?

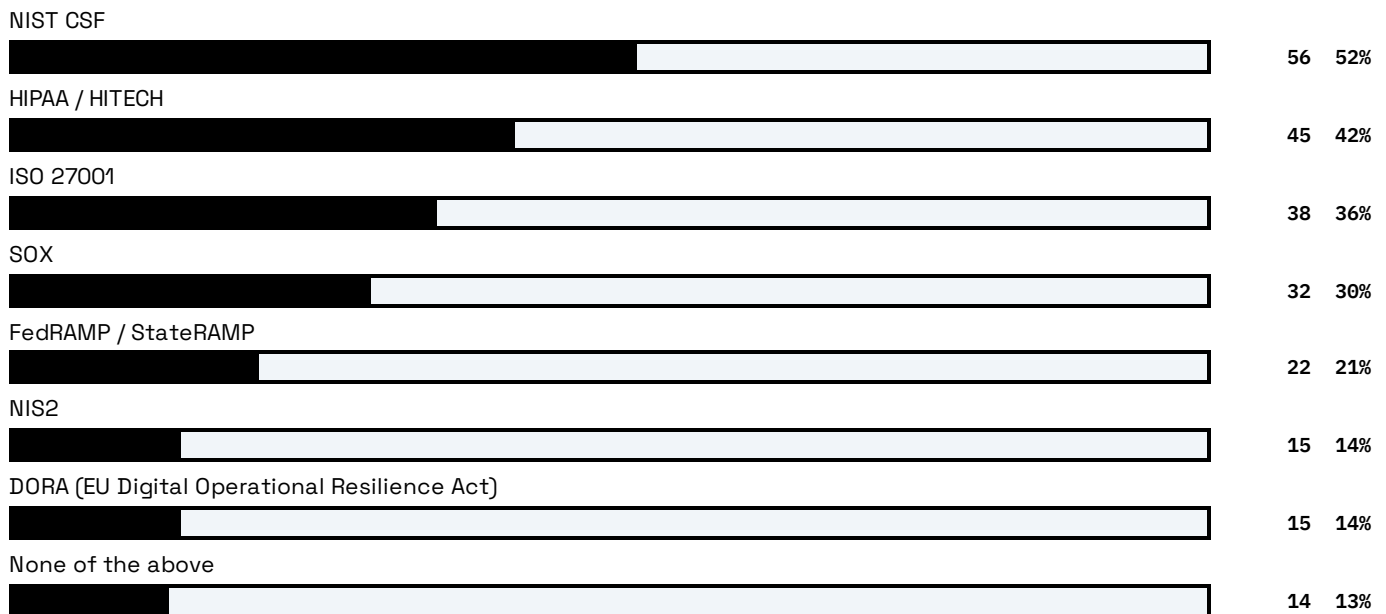


---

DEMOGRAPHICS • N = 107 • MULTI-SELECT

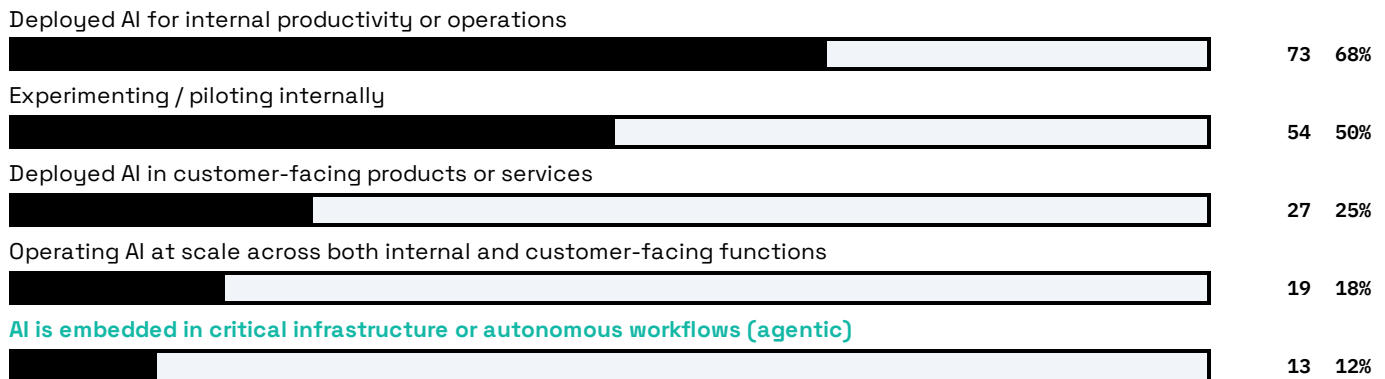
## 6. Is your organization subject to any of the following regulatory or compliance frameworks?

*Multi-select: respondents could choose more than one option, so shares sum to more than 100%.*

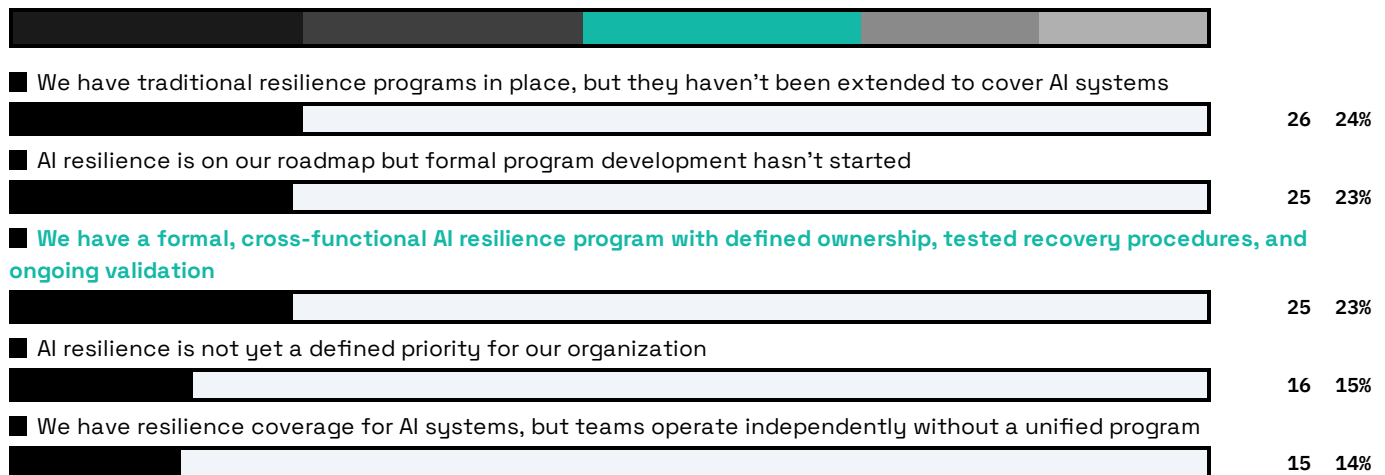


## 7. Which of the following best describes where your organization is in its AI deployment?

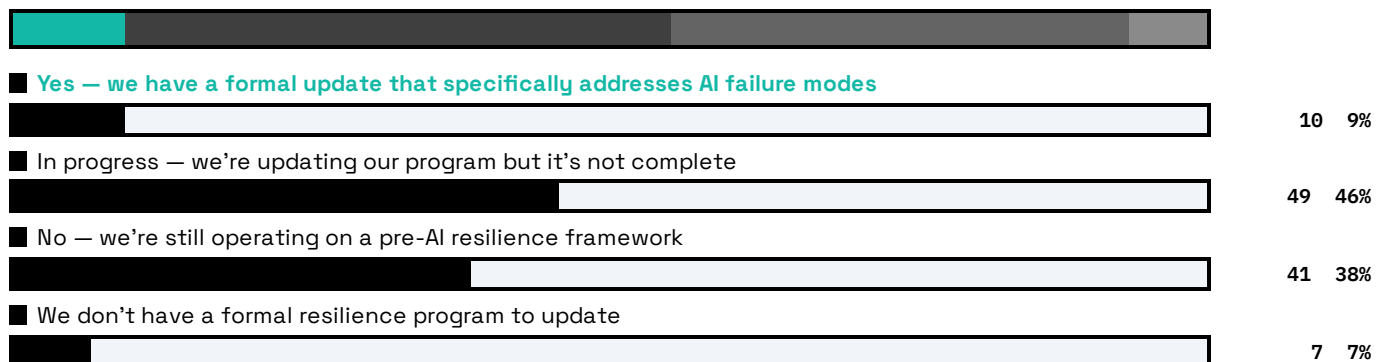
Multi-select: respondents could choose more than one option, so shares sum to more than 100%.



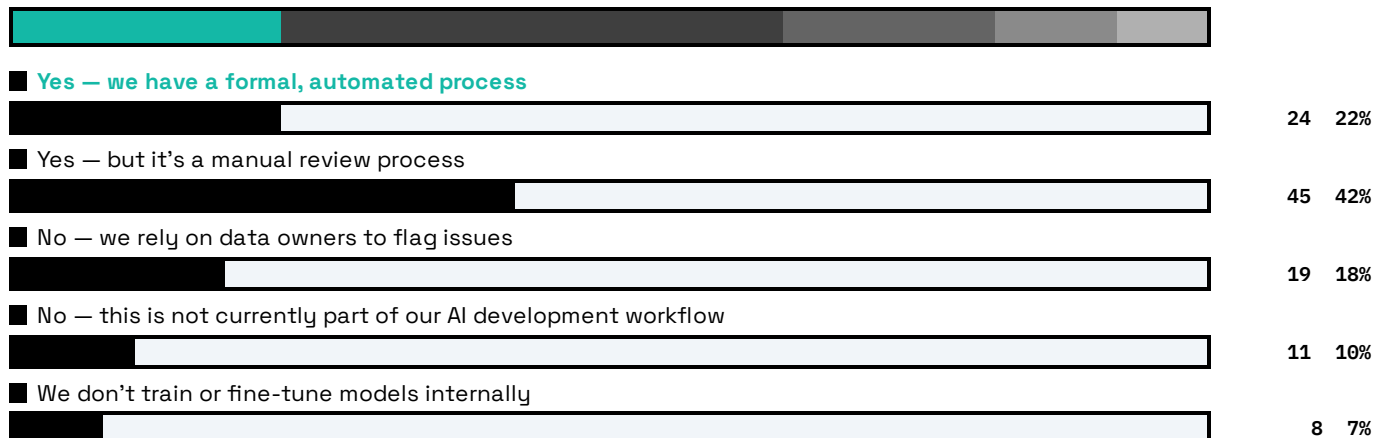
## 8. Which of the following best describes the current state of AI resilience in your organization?



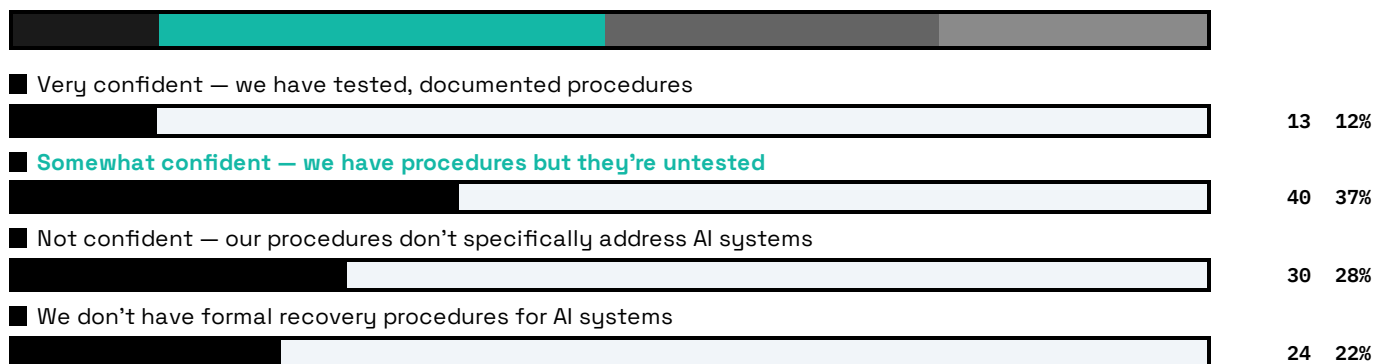
## 9. Has your resilience and recovery program been formally updated to account for AI-specific failure modes such as data poisoning, model compromise, or agentic system failures?



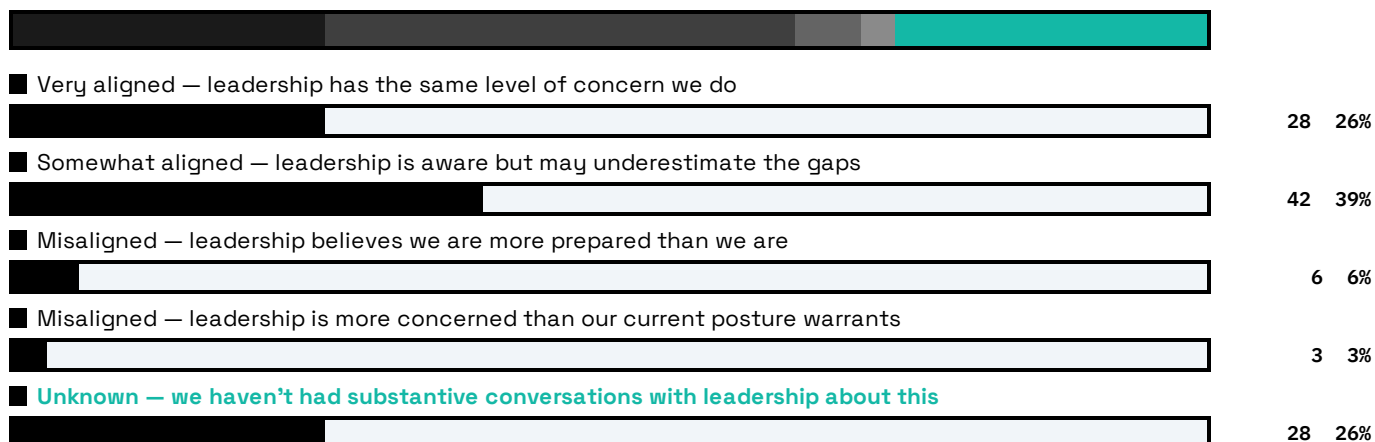
### 10. Does your organization have a process to vet data before it is used to train or fine-tune AI models (for example, to identify PII, trade secrets, or regulated content)?



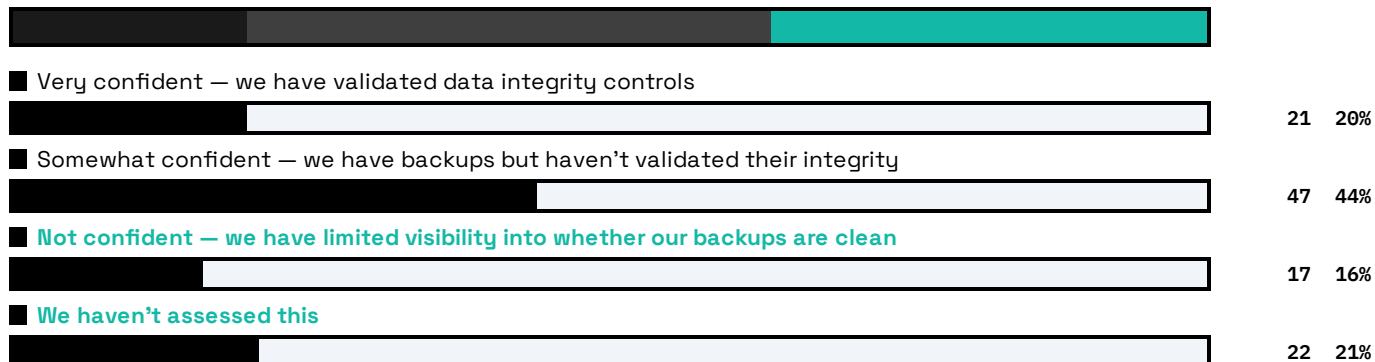
### 11. How confident are you in your ability to recover a critical AI system if it were compromised today?



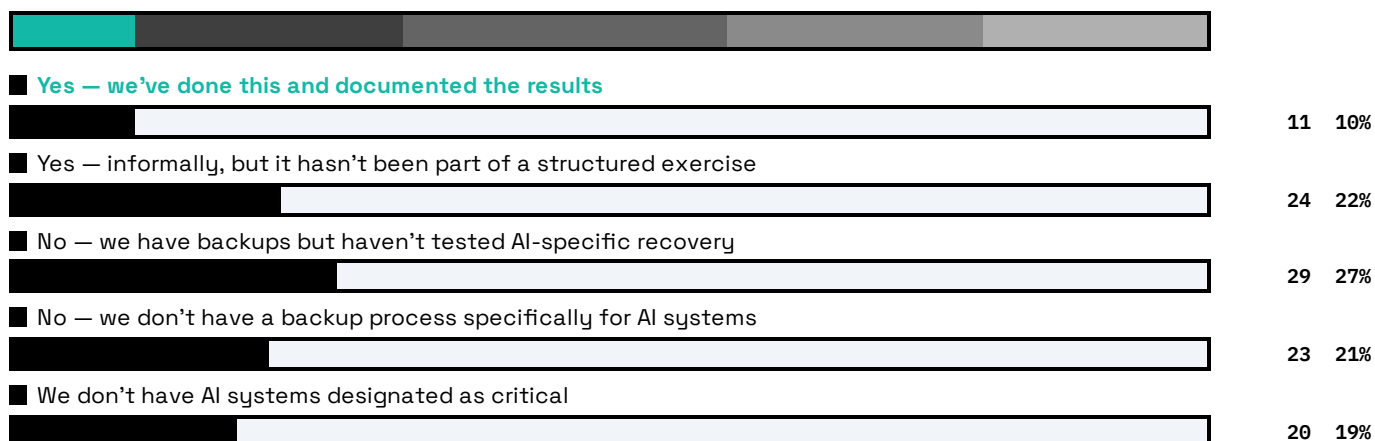
### 12. How aligned do you believe your organization's leadership (board/CEO) is with your assessment of AI resilience readiness?

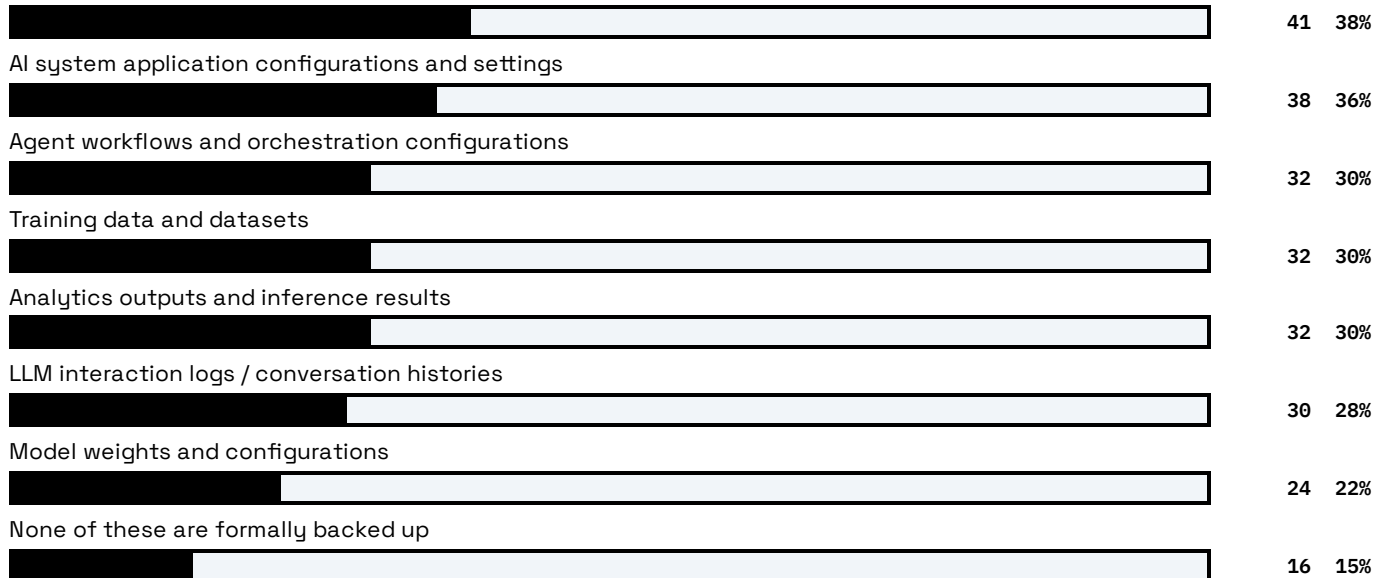
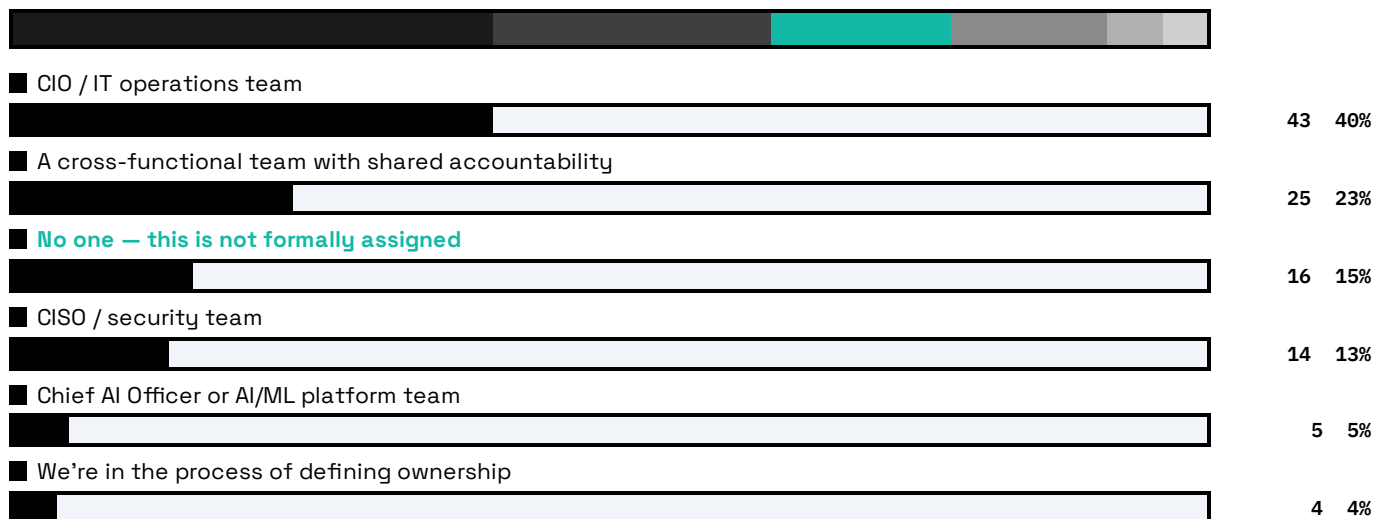


**13. If you recovered a critical AI system after a compromise, how confident are you that the data it would be restored to would be clean and untampered?**



**14. Have you ever tested recovery of a critical AI system or AI-dependent process from a verified clean backup?**



**15. Which of the following does your organization currently back up as part of your AI resilience program?***Multi-select: respondents could choose more than one option, so shares sum to more than 100%.***We back up underlying infrastructure but not AI-specific components****16. Who in your organization is formally accountable for the resilience and recoverability of AI systems and AI-accessed data?**

## 17. Which of the following are major concerns for your organization when it comes to recovering AI infrastructure after a disruption?

Multi-select: respondents could choose more than one option, so shares sum to more than 100%.

That recovery procedures exist for IT systems but not for AI specifically



That we'd have no way to verify the integrity of recovered data



That the data we'd recover to has been corrupted or poisoned



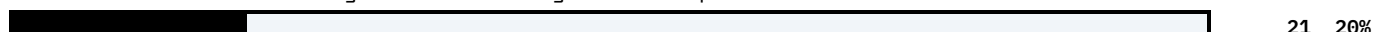
That we've never tested it



That responsibility for AI recovery isn't clearly assigned



That we don't know which AI systems are actually critical to operations



## TESTING & EXERCISES • N = 107

## 18. How frequently does your organization conduct resilience exercises that specifically include AI systems or AI-dependent processes?



Continuously or as part of ongoing monitoring



Quarterly or more frequently



Annually



Less than annually



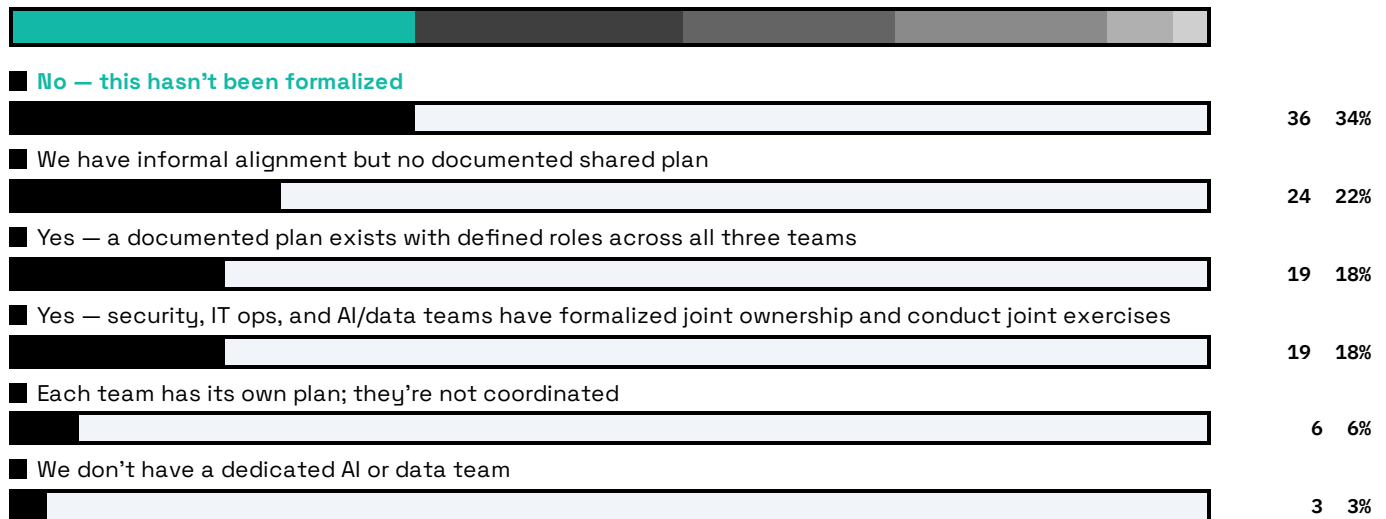
We conduct resilience exercises but they don't include AI systems



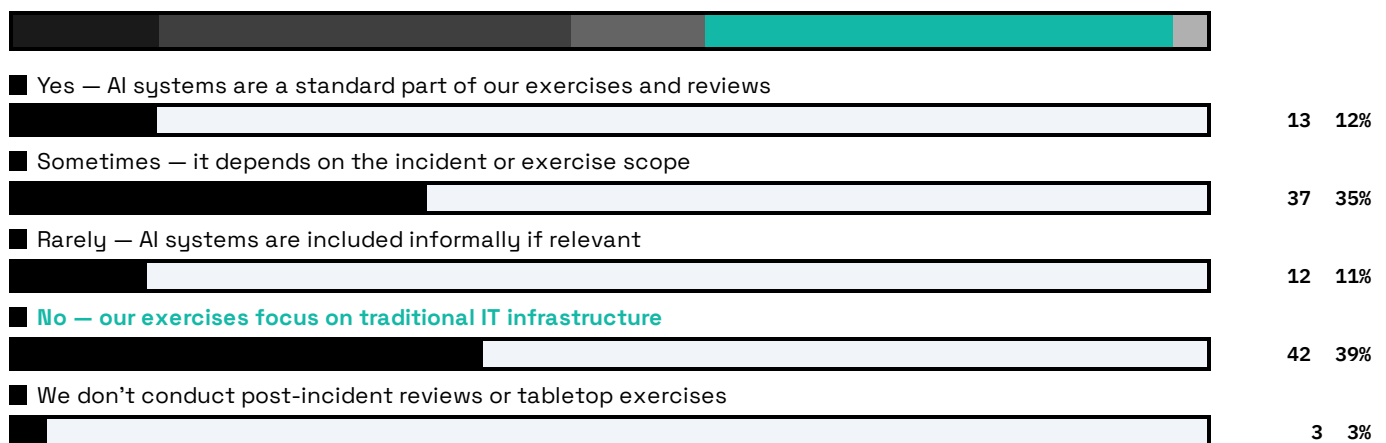
We don't conduct resilience exercises



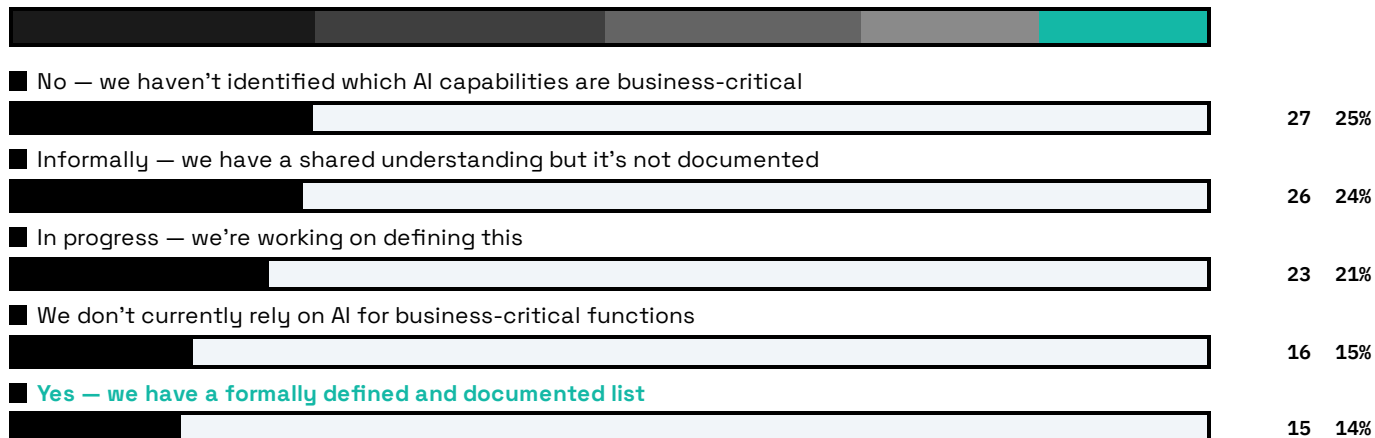
### 19. Do your security operations, IT operations, and AI or data teams have a defined, shared recovery plan for AI systems?



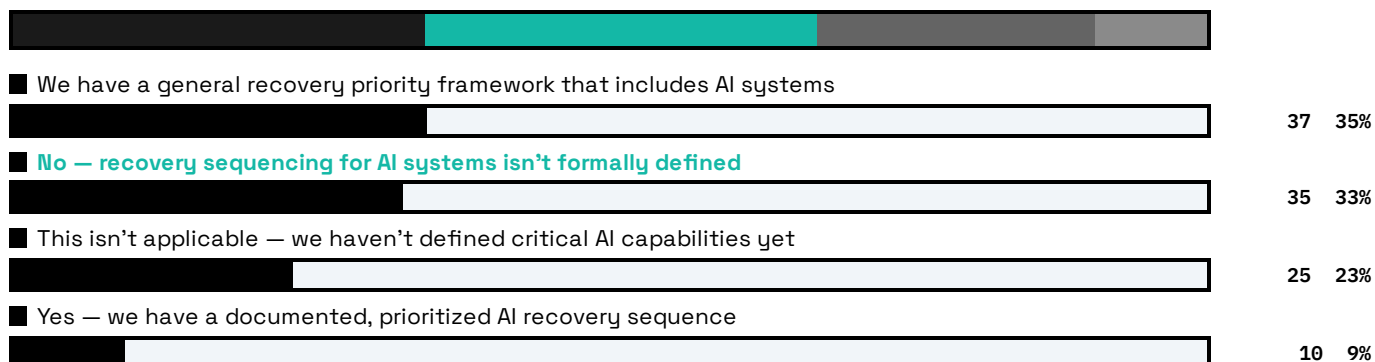
### 20. When your organization conducts post-incident reviews or tabletop exercises, are AI systems and AI-dependent processes explicitly included in the scenario?



## 21. Has your organization identified a 'minimum viable' set of AI capabilities that must be operational for the business to function?



## 22. If a major disruption occurred today, do you have a plan to prioritize recovery of those critical AI capabilities first?



23. Beyond traditional RTO and RPO metrics, which of the following AI resilience metrics does your organization currently track?

Multi-select: respondents could choose more than one option, so shares sum to more than 100%.

