READINESS REPORT ISSUE 6 JUNE 2026

The Agentic Blind Spot: Why AI Resilience Demands a System of Record

For CIOs and CISOs navigating the next era of enterprise AI

READIVERSE

Commvault

Executive Summary

AI adoption has crossed a threshold. The first wave – retrieval-augmented generation, stateless assistants, light automation – gave enterprises a taste of what was possible. What's emerging now is categorically different: agentic systems that plan, remember, and act. They don't answer questions. They execute workflows, coordinate with other agents, and operate continuously across the enterprise.

This shift changes the risk calculus entirely. Stateful agents introduce new failure modes – corrupted memory, compromised training data, ungoverned identity, cascading decisions built on bad state.

The security and resilience frameworks enterprises rely on weren't designed for this. Neither were the point tools most organizations have assembled to manage data protection, identity, and recovery. No collection of point tools can reconstruct the relational context that separates an AI system that appears recoverable from one that actually is.

A system of record (SOR) is **the authoritative, centralized data source for a specific domain within an organization, maintaining data consistency and accuracy.** Every major enterprise technology era has produced an SOR – enterprise resource planning (ERP) for operations, customer relationship management (CRM) for customers, IT service management (ITSM) for service. Each SOR emerged because fragmentation became an existential liability. The agentic era is no different.

This report explores what readiness may look like for the agentic enterprise: the architectural shifts underway, the gaps organizations may not yet have recognized, and considerations for a unified SOR as a potential foundation for AI resilience.

The Architecture Has Shifted

The move from Agentic 1.0 to Agentic 2.0 is not an incremental upgrade. It is a structural change in what enterprise AI does and how it might fail.

Agentic 1.0 was characterized by stateless retrieval. Tools based on Retrieval Augmented Generation answered questions. Copilots suggested next steps. Automation handled discrete, bounded tasks. Failures were local and recoverable.

Agentic 2.0 is stateful and autonomous. These systems maintain memory across interactions. They plan and execute multi-step workflows without step-by-step human instruction. They coordinate with other agents – spawning subagents, delegating tasks, synthesizing outputs across sources. They act on production systems: writing to databases, sending communications, making procurement decisions, modifying code.

This introduces four architectural layers most organizations haven't yet designed for:

Agent memory: Vector databases and session state that give agents continuity – and a new attack surface.

Runtime control: Dynamic workflows that are harder to monitor and harder to stop once compromised.

Agentic observability: Agent-to-agent interactions that happen faster than human-scale logging was built to capture.

Multi-agent coordination: Emergent failures that arise not from any single misconfigured agent, but from how they depend on one another and interact.

These four layers are not only unowned by existing teams, they're also the layers that determine whether AI can be trusted.

Only 1 in 5 companies has a mature model for governing autonomous AI agents.[1]

By 2028, 33% of enterprise software applications will include agentic AI – up from less than 1% in 2024.[2]

The architecture has shifted. The governance, observability, and resilience infrastructure has not.

The shift is from AI as a tool you use to AI as a system that runs – continuously, autonomously, and across the enterprise.

[1] State of AI in the Enterprise, Deloitte

[2] Gartner research cited by Datagrid

The New Attack Surface

Every architectural layer that Agentic 2.0 introduces also represents a threat vector. The security frameworks most enterprises have in place were built for a different kind of AI – not for systems that act autonomously, maintain persistent state, or coordinate decisions across agent networks.

A number of threat categories have emerged, and others will continue to surface. The highest-impact threats so far:

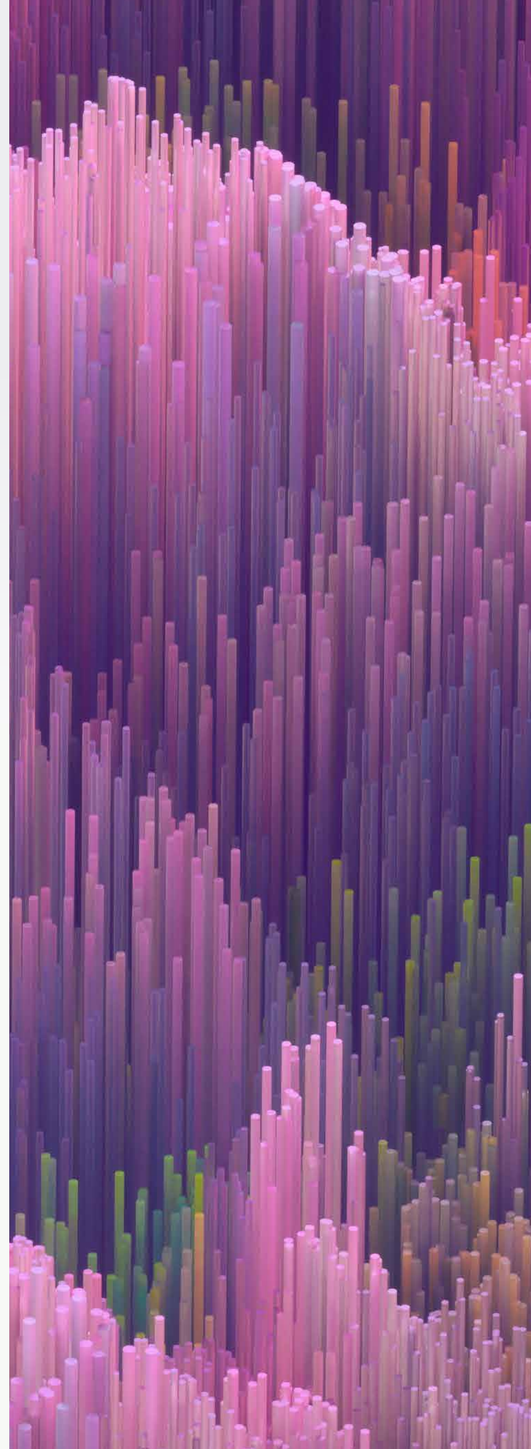
- **Poisoned training data:** Adversarial actors who can influence training pipelines can shape model behavior at scale, with effects that may not surface until long after deployment – and that often evade standard benchmarks.

- **Compromised vector databases:** Injected or manipulated embeddings can redirect agent behavior, exfiltrate data, or cause agents to take actions that appear legitimate but serve an attacker's goals.

- **Ungoverned agent identity:** In multi-agent architectures, a stolen or escalated credential gives an attacker not just access to data but to the autonomous decision-making layer of the enterprise – at machine speed, with blast radius far exceeding any human attacker.

- **Frontier AI models:** New models like Anthropic's Mythos are now accelerating this exposure further – autonomously discovering and chaining vulnerabilities at a speed and scale that compresses the window between discovery and exploitation from weeks to hours.

- **Cascading decisions built on bad state:** Because agents share context and coordinate outputs, a failure in one layer propagates through others before detection.



The board question – is our AI trustworthy? – currently has no defensible answer for most organizations:

37%

Securing AI agents is the top concern for CISOs, cited by 37% of CISOs in a 2025 survey.[3]

38%

Only 38% of organizations monitor AI traffic end-to-end.

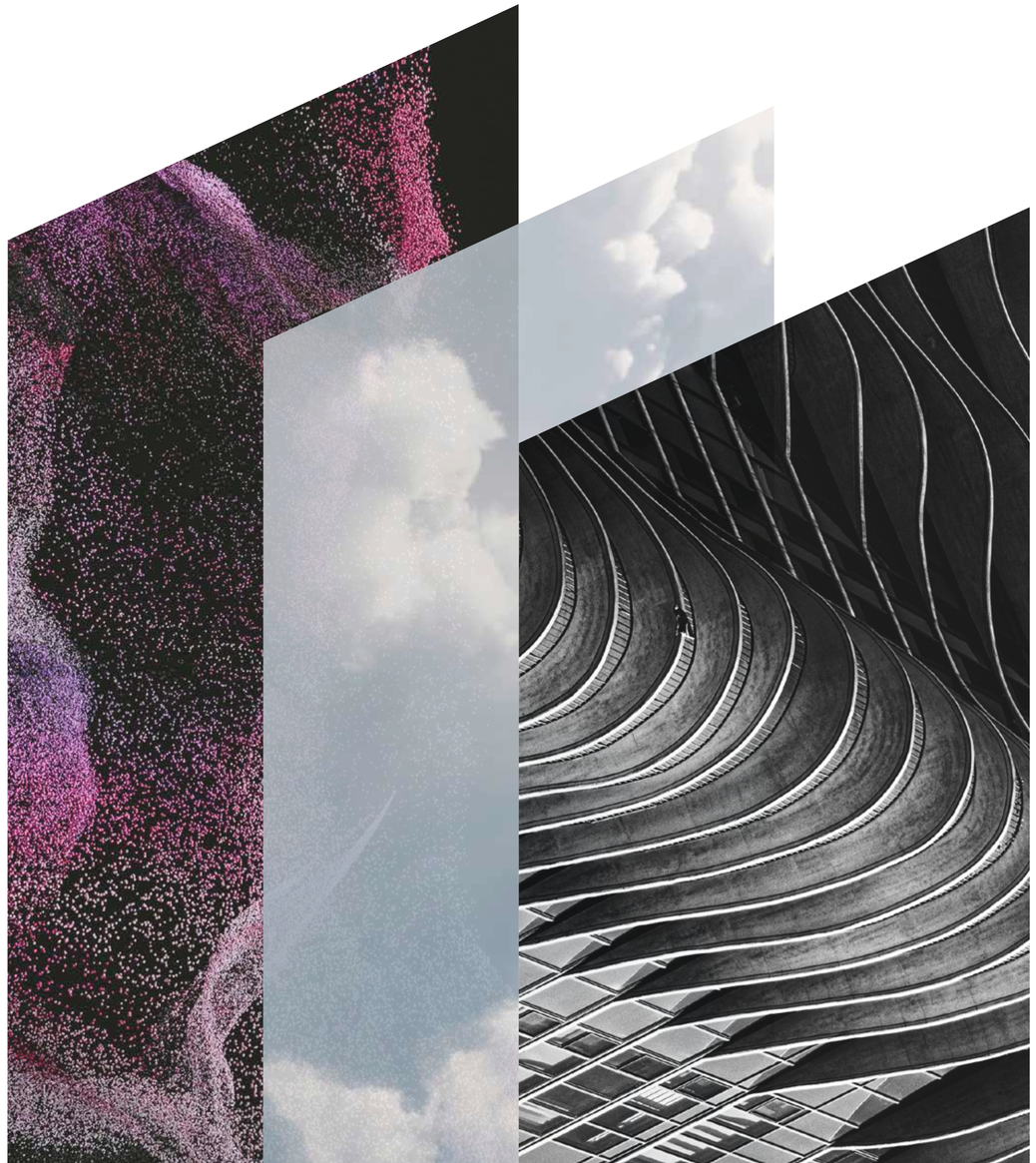
17%

Only 17% continuously monitor agent-to-agent interactions.[4]

100%

Nearly all CISOs now report that AI governance and risk management fall within their responsibilities – which represent significant increases in their responsibilities.[5]

[3] Key Findings from Team8's 2025 CISO Village Survey
[4] The State of Agentic AI Security 2025, Akto
[5] The CISO Report: From Risk to Reliance in the AI Era, Splunk



Where Enterprises Are Exposed

In conversations with CIOs and CISOs managing agentic AI deployments, four pressure points surface consistently.

Workloads Where Hyperscaler-Native Protection Ends

The data stores, orchestration layers, and model infrastructure that agentic systems depend on frequently run outside managed hyperscaler services: OpenShift clusters, MongoDB document stores, Snowflake data warehouses, self-hosted model registries. Native cloud protection doesn't follow workloads into these environments.

The result is a protection gap at exactly the layers where AI state lives – and where integrity must be demonstrable at any time.

Cyber as the Primary Resilience Differentiator

Traditional disaster recovery planning assumed the primary threat was infrastructure failure. That assumption no longer holds. Targeted attacks on AI infrastructure – poisoning training pipelines, corrupting vector databases, compromising agent identities – are the types of threat profiles resilience planning must now be built for.

Organizations that have updated their AI resilience strategy for cyber-specific failure modes maintain a meaningful advantage over those still planning for the incidents of the previous era.

Identity as an Underrecognized Recovery Vector

When organizations plan AI recovery, they focus on data: restoring training sets, checkpoints, model weights. Identity is rarely part of the playbook. Agent identity – the credentials, all-important context, roles, and permissions that govern what agents can do – must be restored as part of any coherent recovery.

A recovered AI system rumnning on clean data with compromised identity configurations is not actually recovered. It is a clean system with a poisoned access layer.

TCO Discipline as Agentic Deployments Scale

Resilience infrastructure is not a cost center – it is the foundation that makes AI investment defensible. Organizations that don't establish it early find themselves unable to justify continued AI spending when the first major incident occurs, or when the board asks for a return.

The Context Gap

When an agentic AI system fails – through attack, corruption, or infrastructure failure – the first question isn't "What went wrong?" It's "What do we recover to?"

Every organization managing enterprise AI has some collection of tools watching over it: a model registry, an artifact store, a pipeline configuration manager. Each can confirm its own slice of the picture: The model registry reports Version 3.2; the artifact store has a checkpoint file – belonging to Version 3.1; the pipeline configuration looks current – but references a preprocessing script updated after the last backup.

What no tool can confirm is whether those pieces belong together – whether they reflect the same moment in time, the same training run, the same operational state.

Pull on any one of those threads and the whole picture unravels. You don't have a recoverable AI system. You have four confident reports describing parts of something that no longer coheres as a whole.

This is the context gap: The absence of a coherent, time-stamped record of how all the components of an AI system fit together at any given point in time. It is not a tooling problem. Adding more point solutions doesn't close it – each new tool adds another slice that can be confirmed in isolation and trusted in nothing.

Backup data – properly air-gapped, immutable, and versioned – is the only asset in the stack that is independently provable as clean. More importantly, it captures state relationally: what version of what model was running against what data, with what configuration, at what moment.

That relational record is what separates a recovery that appears complete from one that actually is – and the natural anchor for a unified system of record.

79% of enterprises operate with significant blind spots – agents invoking tools, touching data, or triggering actions outside the scope of any monitoring or governance framework.[6]

[6] The State of Agentic AI Security 2025, Akto

The SOR for AI Resilience

Every prior enterprise technology era produced a moment of reckoning – where technology sprawl made it impossible to know which version of reality was authoritative, and a unified SOR emerged to resolve it. ERP brought coherence to operations. CRM made customer relationships knowable. ITSM gave IT a single source of truth.

AI has reached that inflection point. And the SOR that emerges from it must do three things no previous system of record has been asked to do:

Provide Continuous Visibility Across Every Data Asset, Identity, and Agent Interaction

An SOR for AI resilience maintains a unified view of the relationships between layers – not just what model version is running, but what training data it was built on, what pipeline configuration it depends on, what agent identities have access, and what the state of all those components was at any given point in time. This includes agent-to-agent interactions that existing tools leave unobserved.

Resilience for AI Stack

AI's output is non-deterministic, so unlike a classic database, businesses cannot just restore and correct it. The rebuild and recreation of all the layers of an AI stack involves all dimensions that constitute the SORs, such as data lineage, model artifacts, weights, config and/or orchestrate state, prompts and RAG indices, inference and audit logs and, most importantly, the vector stores.

Resilience for AI stack is predicated on restoring the provenance chain to demonstrate that the system's behavior after recovery is authentic.

The complexity of the task of rebuilding provenance should be acknowledged and planned for accordingly. Nevertheless, it is important to embark on the process for resilience for AI stack sooner than later.

Prove Recovery to a Known-Good State

Recovery from a compromise of AI infrastructure isn't like restoring a database. You're not restoring data. You're restoring a system – and you need to prove that what came back is what was running before the incident – not just clean, but also coherent. You need a defensible answer to the board, to regulators, and to auditors.

Organizations with comprehensive AI governance policies are nearly twice as likely to report early adoption of agentic AI – 46% vs. 25% for those with only partial guidelines.[7]

[7] The State of AI Security and Governance, Cloud Security Alliance

Readiness for the Agentic Enterprise

Organizational readiness for agentic AI resilience can be assessed across four dimensions. The diagnostic question for each is not whether controls exist – it is whether those controls were designed for the failure modes that agentic AI introduces.

Observability

Can you see what your AI is doing – in real time?

- End-to-end monitoring of AI traffic
- Agent-to-agent interaction logging
- Anomaly detection across the AI stack
- Understanding why agents made the decisions they did

Data Integrity

Can you prove your AI is running on clean data?

- Air-gapped, immutable backup of AI assets
- Relational capture of model/data/pipeline state
- Verifiable chain of custody for training data

Governance

Does your framework cover what agents do, not just what models say?

- Agent identity management and access control
- Policy coverage for autonomous agent behavior
- Audit trail for agent decisions and actions

Recovery

Can you recover to a verifiable, coherent AI state?

- Recovery to known-good AI state (not just data)
- Proof of integrity under board/regulatory scrutiny
- Tested playbooks for AI-specific incident scenarios

Most organizations entering the agentic era have partial coverage across these dimensions. The priority sequence follows the risk:

1. Observability first (you cannot govern what you cannot see).

2. Data integrity as the recovery anchor.

3. Governance, extended to cover agent behavior and not just model outputs.

4. Recovery tested against AI-specific failure scenarios – poisoned training data, compromised vector databases, identity-layer failures – not just the infrastructure disruptions of the previous era.

The window to establish this infrastructure on deliberate terms is narrowing. Organizations that build AI resilience foundations now – before a significant incident – will be better positioned to scale with confidence. Those that wait will be building under pressure, with less time and higher stakes.

AI that runs on untrustworthy data produces untrustworthy outcomes. No enterprise can afford to find that out mid-incident.

About This Report

This Readiness Report is published by Commvault as part of the Readiverse thought leadership series. Readiness Reports are designed to help CIOs and CISOs navigate the strategic and operational dimensions of cyber resilience.

Access additional resources at www.readiverse.com/ai-resilience.