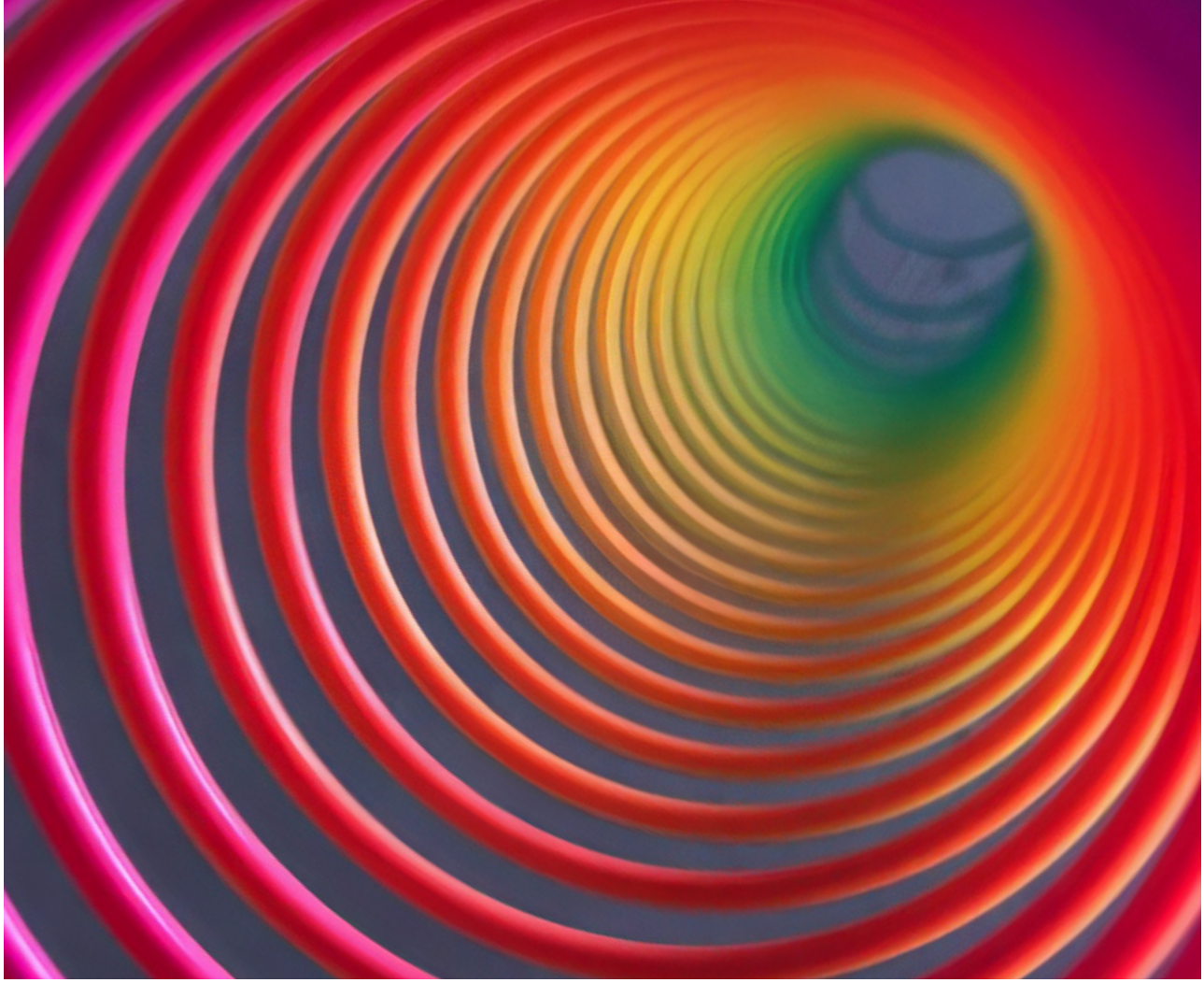


# Who's accountable when agentic AI acts alone?

IT governance was founded on human approval.
The rise of agents means a control plane, evidence layer, and ownership layer replace that permission.

OLIVER WYMAN
A MARSH BUSINESS

# Contents


Introduction 3

1. Governing what agents can do: the control plane 7

2. Proving what agents have done and containing it: the evidence layer 19

3. Deciding who’s responsible: the ownership layer 32

# Introduction

**This paper advances one argument told in three parts**

First, build a control plane. Stop approving agent actions one by one and replace that approval with guardrails enforced as agents work. Scale then becomes a matter of managing what already exists rather than a creation problem.

Second, prove it to the outside world. Your security officer wants to know what an agent can do. A regulator wants to know what it has already done and whether it was authorized to do so. Building an evidence layer answers both.

Third, name who is responsible. That means deciding which agents deserve capital and settling who answers when something goes wrong. That name must satisfy the board, the regulator, and the customers.

**What next?** Ask three questions of your own organization. Can you find every agent you have deployed, and explain what each one is permitted to do? Can you prove, after the event, what an agent did and who authorized it? Can you name the executive who answers to the board if one fails? Most large organizations answer "no" on all counts.[1]

---

The following chapters turn each of those answers from "no" to "yes."

---

1 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

**Exhibit 1: Most large organizations answer “no” three times**

**Chapter 1 — The control plane**
Can you find every agent you have deployed, and state what it can never do?




of CIOs can produce every production agent’s owner, scope, cost, and last certification date on demand

**24%**
can state the ratio of non-human to human identities in their estate

**8%**
of CIOs running agents in business as usual run a removal cadence with a named budget and owner

**Chapter 2 — The evidence layer**
Can you prove, after the event, what an agent did and who authorized it?

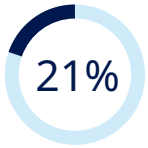



of CIOs can reconstruct, from a tamper-evident record, what an agent changed and who authorized it

**3%**
of CEOs are very confident they could stand behind what an autonomous system did

**63%**
only of CIOs running agents in business as usual enforce hard caps, throttles, or a kill switch

**Chapter 3 — The ownership layer**
Can you name the executive who answers to the board if an agent fails?




of CIOs have appointed a head of AI governance with cross-functional authority over agent policy, cost, and risk

**73%**
of boards or executive committees review agentic progress and risk quarterly

**57%**
expect freed capacity to be redeployed with no one named to do it

Source: The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026. Figures as presented in chapters 1 to 3.

## The assumption that just broke

From COBIT and ITIL to change advisory boards (CAB) and RACIs, every governance framework that enterprises run today rests on an assumption that until now nobody needed to write down: IT changes are made by a person who can be asked, before or after the fact, whether they have permission to act. Agentic AI removes that assumption and usually does so before anyone notices. An agent does not wait for a change window. It files no ticket and it has no manager to escalate to. It acts, then acts again thousands of times before a quarterly access review reaches its first decision.

© Oliver Wyman

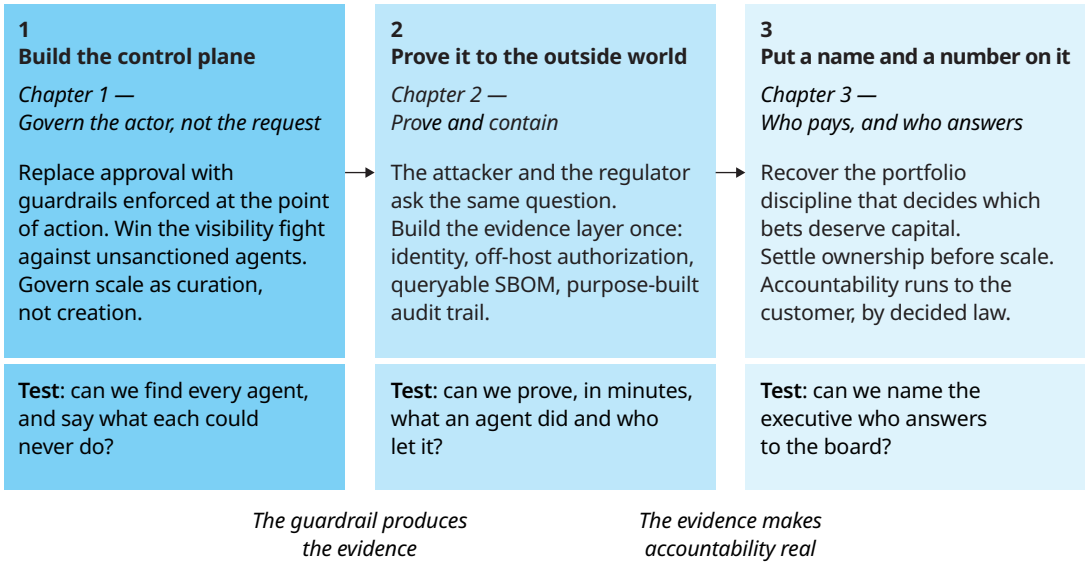
Most CEOs are ready to grant some authority: 83% of CEOs will let agents act in processes that touch customers, revenue, or financial controls — most of them doing so inside tight preset limits.[2] Ask the same executives whether they could reconstruct and stand behind what an autonomous system did in a customer or regulatory matter, and 90% say they are only "somewhat confident." Only 3% say "very confident."[3] The authority has moved. The means of accounting for its use has not.

None of this makes the argument for slowing down. Get this right and you will deploy agents faster than your competitors, because your governance is enforced at runtime rather than reviewed after the event.

## Governance moves from the meeting to the execution path

Approval, evidence, and ownership used to be three faces of the same mechanism: a person asked for permission, a record showed who granted it, and a manager answered for the result. Agents pull that mechanism apart. The following chapters rebuild each one in an order that creates the conditions for the next: the control plane sets the boundary an agent cannot cross, the evidence layer proves what happened inside it, and the ownership layer names who answers for the result. Remove any one layer and the other two weaken: guardrails without evidence cannot be audited, evidence without an owner is never acted on, and an owner without guardrails answers for what nobody could prevent (Exhibit 2).

**Exhibit 2: The guardrail produces the evidence, and the evidence makes accountability real**




2 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

3 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

© Oliver Wyman

## How to read this document

On first reading, take the three chapters in order. The guardrail creates the evidence on which the second chapter builds, and the evidence is what makes the ownership described in the third chapter enforceable. After that, each chapter stands on its own as a decision guide, and each closes with the risks a CIO must contain and the decisions only they can make. Start with the control plane if your program is stuck between pilots and scale, or if business units are building agents you cannot see. Start with the evidence layer if your CISO or a regulator is asking questions this quarter. Start with the ownership layer if the conversation with your board or your CFO is about funding, ownership, or liability.

Waiting for the technology to settle is not an option. Business units are already building invisible agents outside your line of sight. By available estimates, more than 80% of AI projects fail.[4] Governance designed now costs a fraction of governance improvised after an incident. The window is open today. It will not stay open.

4 Ryseff, De Bruhl, and Newberry. *The Root Causes of Failure for Artificial Intelligence Projects*, RAND Corporation, RR-A2680-1, 2024, available at rand.org/pubs/research_reports/RRA2680-1.html.

# The control plane

Governing what agents can do



1

CONTROL PLANE

Govern the agentic actors

**Asks: what may this agent do?**

2

EVIDENCE LAYER

Prove and contain

**Asks: can you prove who acted?**

3

OWNERSHIP LAYER

Deciding who is responsible

**Asks: who pays and who answers?**

© Oliver Wyman

The approval loop can't keep up with machine speed. The guardrail replaces it.

© Oliver Wyman

# Governing what agents can do: the control plane

## The operating reality

Of the three layers, the control plane is the one with the least time to spare. Agents are already in production, so the boundary on what they can do has to be built around a live estate rather than designed ahead of deployment.

The problem already exists across the estate. A business unit can build an agent with a chat prompt and an API key. Each one arrives with its own identity and its own reach across your systems.

An inability to count them is a common first failure: 24% of CIOs can state the ratio of non-human to human identities in their estate, 61% hold a count without a ratio, and 13% have not inventoried them at all.[5] Scaling only widens that gap. Some 6% of CIOs describe agents running across multiple functions as standard practice, and 60% are still in pilots.[6] Where no control mechanism exists, autonomy is granted by default. The first violation surfaces after several thousand executions rather than before the first one.

---

24% of CIOs can state the ratio of non-human to human identities in their estate. Where no boundary is enforced, autonomy is granted by default.

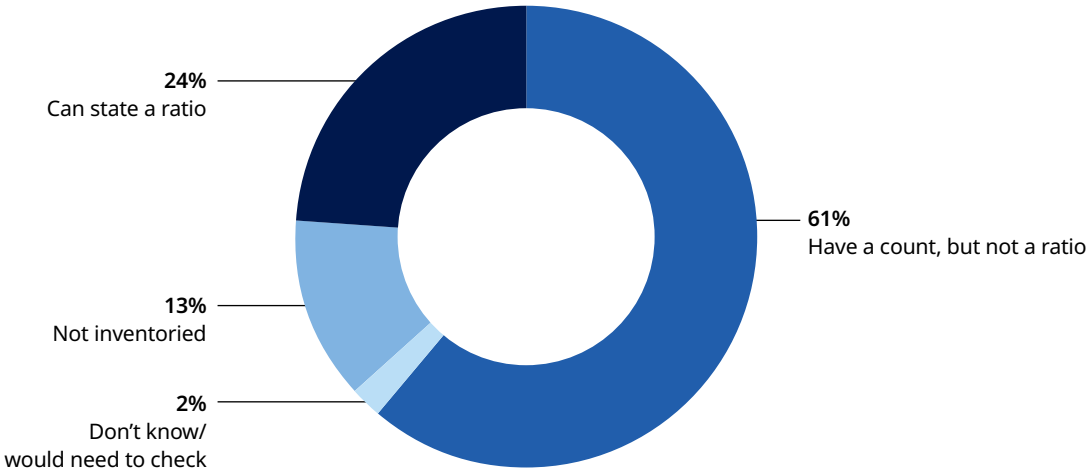
---

5 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

6 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

**Exhibit 3: Three in four CIOs cannot state their non-human-to-human identity ratio**

What is your ratio of non-human identities (agents, service accounts, workload identities) to human identities? % of CIOs




Base 126 CIOs

Source: The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

### Our point of view

Make the guardrail the unit of governance: versioned, testable rules enforced at the point of action, defining what an agent is prohibited from doing regardless of who asks. Keep self-service creation open and make the governed path faster than the workaround. Hold the enforcement boundary outside the agent's runtime. That way, a business unit builds inside the scope that central governance granted without widening that scope itself. At scale, run hub-and-spoke: a central team sets the rules and each business unit builds within them. Tie delegation to demonstrated readiness. Close the telemetry loop before adding units. And treat the registry as continuous curation rather than a creation gate. Follow this path and you execute this move faster, because your control system keeps pace with your agents.

**What next?** Ask two questions. First: can you find every deployed agent, each with a named owner? Second: for any one of them, can you list the actions an enforced control would block? The first capability is rare: 10% of CIOs can produce every production agent's owner, scope, cost, and last certification date on demand, 58% hold some attributes, and 29% run no central registry.[7] If you cannot answer the first, discovery (finding out what already exists) is your first priority. If the second list is empty, that agent runs fully autonomous, whatever your policy document says.

7 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

**Exhibit 4: Governance shifts from approving requests to bounding actors**

| Current usage | What agentic changes |
|---|---|
| Approval queues and change boards for agent actions | Guardrails enforced in the execution path |
| Policy documents a human interprets | Policy-as-code: versioned, testable, auditable rules |
| Prohibition, human tickets, and disciplinary memos | A governed path faster than the do-it-yourself workaround |
| Shared service accounts; autonomous by omission | A named owner and a distinct identity per agent |
| Agents trusted to police themselves | Enforcement boundary outside the agent's runtime |
| Per-agent logs with no cross-unit view | Telemetry[8] closed-loop across the whole federation |
| Registry as a one-time creation gate | Registry[9] as continuous curation of what exists |

## Put the guardrail in the execution path

**Approval is a means, not an end in itself.** ISO/IEC 38500, the standard from which COBIT and ITIL draw of their governance vocabulary, organizes governance around evaluation, direction, and monitoring, with stakeholder engagement added in its 2024 edition.[10] Human approval was a practical way to direct and monitor activity while a person executed each change. Once an agent applies policy faster than a reviewer reads it, it will take a different mechanism to guarantee the same level of assurance.

**Without an enforced guardrail, agents inherit whatever discretion the surrounding system allows.** Discretion can be described along four governance patterns, from Directive through Guided and Federated to Autonomous, each measuring how much latitude a governed unit holds. Most IT organizations sit between Guided and Federated for their human teams. Absent an enforced boundary, their agents default toward Autonomous, whatever setting the policy intends. In continuous delivery pipelines, 5% of CIOs report agents holding control-plane authority, over deploy policy and approval gates, with no human-owned control and no recorded decision behind it.[11] Nobody chose that setting. It is what remains when no boundary is enforced.

8 Telemetry is defined as the continuous collection and analysis of operational data (for example, activity, performance, and security events) across all agents to provide real-time visibility.

9 Registry is defined as the inventory of agents and their metadata (for example, ownership, capabilities, and permissions) that serves as the authoritative source for governance.

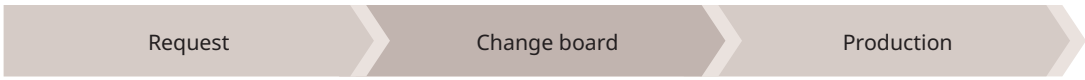
10 ISO/IEC. ISO/IEC 38500:2024 Governance of IT for the Organization, 2024.

11 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

**The mechanism already exists in a discipline many IT organizations run today.** In the more useful architectures, policy is enforced in the agent's execution path, where interception points stop, redirect, or log an action before it completes.[12] Policy-as-code is mature in cloud infrastructure and transfers directly to agents. The same rule language, versioning, and audit trail that once governed what a script could provision now govern what an agent may do.[13] Organizations already running infrastructure-as-code with policy gates are extending a discipline already in place, rather than buying a new one.

### Exhibit 5: From the approval gate to the guardrail in the path

**Yesterday — the approval gate**
Built on one assumption: the executor is a person who can be asked.




*Waits days to weeks*

**Now — the guardrail in the execution path**
Versioned, testable rules define what the agent can never do, whoever asks.




*Checked before the action completes*

*The question changes: from "who authorized this?" to "what could this system never have done, no matter who asked?"*

**This is not a slower way to run IT.** ITIL has preauthorized low-risk, well-understood standard changes for well over a decade.[14] Agentic AI forces the question of which changes ever needed a meeting in the first place. The answer becomes the preauthorization a machine can enforce. Existing practice already points that way: 96% of CIOs set autonomy per workflow rather than at one blanket level, and 70% sit at the level where the agent runs the steps and a human approves.[15] Those are intended settings. The guardrail is where each of those per-workflow decisions gets recorded and enforced; without it, the intended level and the effective one drift apart.[16]

12 Shlomov S, Shoham I, Oved A, et al. "Governance by Construction for Generalist Agents," arXiv:2605.20874v1, (May 2026).

13 Romeo F, Arena L, Blefari F, et al. "ARPaCCino: An Agentic-RAG for Policy as Code Compliance," arXiv:2507.10584v2, (November 2025).

14 AXELOS. *ITIL Foundation: ITIL 4 Edition*, The Stationery Office, 2019.

15 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

16 Anthropic. "The AI-Native SDLC Playbook," August 2026, available at claude.com/blog/the-ai-native-sdlc-playbook.

© Oliver Wyman

**Guardrails trade supervision for autonomy, but they do not circumvent accountability.** Responsibility has to stay traceable across four parties: the model provider, the deployer, the guardrail designer, and the policy owner. Keeping that chain visible is your job as CIO; no tool will do it automatically. Structures and processes are the part of governance now being automated.[17] The part still missing is a substitute for human judgment when an edge case falls outside the rulebook.

**Set where the guardrail intercepts an action by how easily the action can be undone, not only by its category.** Distributed-transaction design already has a usable taxonomy for this. Idempotent actions, those that can be repeated without side effects, do not need a gate. Reversible actions can run inside a light gate. Actions that only have a fallback fix, rather than a true undo option, need the interception to log that fix before executing. Irreversible actions are the pivot point, at which the guardrail holds the line regardless of how routine the request looks. Most guardrail catalogs get this wrong, classifying by action type or system touched. That misses the routine-looking action that is about to cross the pivot.

## The first test comes with the agents that nobody sanctioned

**What distinguishes shadow agentic AI is its unsanctioned judgment.** CIOs have seen shadow IT before: the SaaS tool bought on a corporate card, the macro nobody in IT ever saw. Treating this as similar to those examples is the most common mistake we see. Classic shadow IT was a person using a tool IT had not approved. Shadow agentic AI is a machine making decisions that IT never sanctioned.[18] An unapproved subscription creates a visibility gap and an unapproved agent with write access to customer records creates a decision gap, at a speed and volume no visibility program was built to catch.

**Prohibition pushes creation beyond the reach of your telemetry.** If registering an agent takes longer than building one with a personal API key, the policy loses whatever it says. Governing the decision to create an agent is targeting the wrong part of the problem. Govern instead the paths by which those agents become visible. That is the version of this problem you can solve without killing the self-service capability that makes agents valuable. The sanctioned path has not won that contest yet: 37% of CIOs call it the path of least resistance, and 55% say it is no easier than the workaround.[19]

**Visibility matters only if the business unit cannot switch it off.** Two architectural controls hold that boundary. A policy engine outside the agent's runtime lets a unit extend what its agents may do inside a scope that the center granted. That engine structurally prevents the unit from granting itself further scope. An authorization gateway checks what an agent is allowed to do from outside the agent's own host, so a compromised agent cannot talk its way past its

---

17 Van Grembergen W, De Haes S, and Guldentops E. "Structures, Processes and Relational Mechanisms for IT Governance," Idea Group Inc., 2004.

18 Shunmugasundaram M, Gangadharan S, Dave R, et al. "Shadow AI as the New Shadow IT: Governance Blind Spots in Autonomous Enterprise Systems," EDPACS, (June 2026).

19 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

own limits.[20] Sequence works against you here: 60% of CIOs running agents in production built the harness first and left the governance layer last, while 27% bought the governance platform before any agents existed.[21] Retrofitting a boundary around a live fleet costs more than building it first, and the market shows what the delay buys: 92% of organizations lack full visibility into their own agent identities, and 95% doubt they could contain a compromised agent once found.[22]

## At scale, governance means managing what already exists

**Scale creates two governance problems, and mixing them up is a common cause of failure in scale programs.** One is organizational: how to delegate the authority to build across business units of very different readiness, without a central bottleneck and without losing control. The other is volume: what happens once the number of agents passes the point where nobody, including the center, can count them.

**Hub-and-spoke answers the organizational problem as long as two conditions are present.** The hub owns the nonnegotiables: guardrail standards, the identity scheme, the shared registry, and security baselines. Business units then build inside that boundary at their own pace. First, demonstrated readiness gates delegation, assessed in verifiable stages rather than through a subjective maturity conversation; readiness-level frameworks from aerospace autonomy show how.[23] Second, the federation runs a closed telemetry loop across every unit. Without it, a violation spanning two units will remain invisible and the model reproduces the vacuum it was built to prevent, with better paperwork.[24] Close that loop before adding units. One issue remains open: no protocol yet exists for how a unit requests an extension to its delegated scope.

**Volume creates a different class of problem.** Agent sprawl follows five recurring patterns: duplicated agents, ghost agents outside any registry, orphaned agents whose owner has left, permission creep across a population, and delegation chains nobody traces end to end. One agent's error can propagate through the agents downstream of it faster than a human-paced incident process can react.[25] Containment therefore has to be a design property, not an operating habit: narrow scopes, approval tied to material consequences, detection at machine speed, and telemetry on the interactions between agents rather than per-agent logs alone. Runaway resource consumption is a financial problem as much as an operational one; chapter 3 picks up that thread.

---

20 Joshi A, Finin T, Joshi K, et al. "Deontic Policies for Runtime Governance of Agentic AI Systems," arXiv:2606.19464v1, (June 2026). Kodathala SV. "aiAuthZ: Off-Host, Identity-Bound Authorization for AI Agents," arXiv:2607.05518v1, (July 2026).

21 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

22 Cloud Security Alliance. "The AI Agent Governance Gap: What CISOs Need Now," CSA Lab Space, April 3, 2026.

23 Hobbs KL, Lyons JB, Feather MS, et al. "Space Trusted Autonomy Readiness Levels," *2023 IEEE Aerospace Conference*, (2023).

24 Pathak A and Jain N. "Governance-Aware Agent Telemetry for Closed-Loop Enforcement in Multi-Agent AI Systems," arXiv:2604.05119v1, (April 2026).

25 Acharya V. "Governing the Agentic Enterprise: A Governance Maturity Model for Managing AI Agent Sprawl in Business Operations," arXiv:2604.16338v1, (March 2026).

Once agents are cheap to create, governance has to move downstream of creation. When building was slow, creation was the event worth governing; now the pressure sits on what happens afterward, and three questions follow. Which agents are worth keeping? How are their actions monitored? And how quickly can the organization accept or reject what has already been built? Decommissioning agents is where operating models are thinnest: among CIOs running agents in business as usual, 8% run a removal cadence that has a named budget and owner, 44% run one without either, and 40% hold a maintenance cadence only.[26] At the acceptance gate, deterministic checks outperform a smarter agent watching the first one. The best available benchmarking shows rule-based guardrails in the execution path recovering 19.9% of agent failures at a 0.5% false-positive rate, and outperforming prompt-based defenses on all six models.[27] Still, the hardest cases need a person. That is where most organizations already sit: 83% rely on human review for consequential outputs, and 52% run automated evaluation against a reference.[28]

26 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

27 Naik VI, Xu C, Dong D, et al. "GuardianAgentBench: Where Agents Fail and How to Guard Them," arXiv:2607.20982v1, (July 2026).

28 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

# The material risks

**Exhibit 6: Four ways the control plane fails, and what contains each**

| Risk | Mechanism and mitigation | Residual | Owner |
|---|---|---|---|
| **Autonomous by omission** | No enforced boundary exists between an agent and the systems it touches, so a violation is discovered after thousands of executions instead of being blocked at the first.<br>**Solution**: Classify the fleet against the four governance patterns and build interception before increasing autonomy. | **High** | CIO |
| **An agent nobody can stop** | An orphaned or unregistered agent keeps acting after its creator has left. When it misbehaves, nobody can find who owns it or how to switch it off.<br>**Solution**: Require a named owner and a distinct identity as the price of the fast path, and fund discovery of what already exists. | **High** | CIO + BU heads |
| **A federation that sees nothing** | Hub-and-spoke without end-to-end telemetry cannot see a violation that spans business units. The model looks correct on paper and fails where it matters.<br>**Solution**: Build federation-wide telemetry before adding units. | **High** | CIO + CISO |
| **Cascade and sprawl** | One agent's error propagates downstream at machine speed, while ghost, orphaned, and duplicated agents accumulate unwatched. Registry coverage is thin.[29]<br>**Solution**: Narrow scopes, approval tied to material consequences, interaction-level monitoring, and a registry run as continuous curation. | **Medium—high** | CIO + CISO |

Source: Oliver Wyman analysis. Residual ratings reflect likelihood and impact after the stated mitigations.

29 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

© Oliver Wyman

# The decisions only the CIO can make

**Exhibit 7: Seven control-plane decisions only the CIO can make**

| Decision | Recommendation | What must be true |
|---|---|---|
| **Unit of governance** | The guardrail, enforced at the point of action, not a control bolted onto a human approval process. | Policy stops or redirects an action before it completes. |
| **Posture on self-service** | Enable and register rather than prohibit; treat unsanctioned creation as a service-design failure first. | The governed path is measurably faster than the workaround. |
| **Enforcement architecture** | A policy engine outside the agent runtime; authorization off the agent's host for consequential actions. | Business units can extend within scope, never beyond it. |
| **Identity** | Every agent carries a distinct identity and a named owner. | No shared service accounts across agents. |
| **Operating model** | Hub-and-spoke: hub owns guardrails, identity, registry, and baselines; units build inside the boundary. | The boundary is enforced in architecture, not policy. |
| **Delegation gate** | Readiness levels with explicit completion criteria per business unit. | Assessment is verifiable, staged, and repeatable. |
| **Registry and acceptance** | Continuous curation of what exists, with simple deterministic checks at the acceptance gate. | Verification capacity scales with creation volume. |

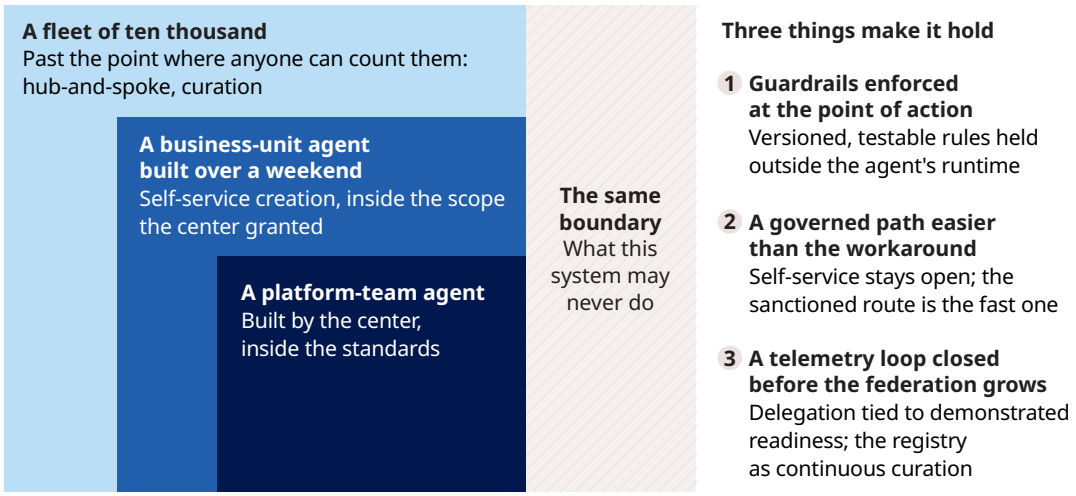
Source: Oliver Wyman analysis

© Oliver Wyman

# Conclusion: One boundary, three scales

The governance question shifts from after-the-fact authorization to what the agent was allowed to do in the first place. The same boundary has to be able to hold for a platform-team agent, a business-unit agent built over a weekend, and a fleet of 10,000. Three things allow it to hold: guardrails enforced at the point of action, a governed path easier than the workaround, and a telemetry loop closed before the federation grows. Build that boundary now and you will grant your agents more autonomy than your competitors can safely allow, and grant it sooner. The rest will learn where their limits were from an incident report.

**Exhibit 8: One boundary, three scales: what makes it hold**




Source: Oliver Wyman analysis (schematic). The three conditions are developed in this chapter: guardrails in the execution path, the governed path, the telemetry loop.

# The evidence layer

Proving what agents have done and containing it



1

CONTROL PLANE

Govern the agentic actors

**Asks: what may this agent do?**

2

EVIDENCE LAYER

Prove and contain

**Asks: can you prove who acted?**

3

OWNERSHIP LAYER

Deciding who is responsible

**Asks: who pays and who answers?**

The security team wants
to know what an agent can do.
The regulator wants
to know what it did, and under
what authority.
Both need the same evidence.

© Oliver Wyman

# Proving what agents have done and containing it: the evidence layer

## The operating reality

**The attacker moved first.** Mean time from vulnerability disclosure to working exploit has fallen to an estimated minus seven days: a working exploit now routinely exists before the patch is released.[30] Machine identities already outnumber human ones by a wide margin, while most security leaders still define "privileged user" as human-only.[31] Few organizations can even state their own position: as chapter 1 showed, fewer than one CIO in four can give the ratio of non-human to human identities in their estate.[32]

**The regulators moved second, and in opposite directions.** The European Union has written one answer to agentic risk into primary law that is uniform across sectors and member states. The United States, meanwhile, has roughly one answer per state and per regulator, and its banking supervisor has said outright that the current model-risk rulebook does not apply to agentic systems.[33] Despite the divergence, security and compliance draw on the same evidence, asking what happened, who acted, and under what authority. That evidence is thin today, with 25% of CIOs able to reconstruct from a tamper-evident record what an agent changed and who authorized it, and 72% holding partial logs.[34] The common foundation is an operating-model layer that can be built today. The security operations center (SOC) and incident response stay with the CISO, and legal strategy stays with the General Counsel. This chapter covers the layer beneath both.

---

30 Kutscher J. "M-Trends 2026: Data, Insights, and Strategies From the Frontlines," Google Cloud Blog, March 23, 2026, available at cloud.google.com/blog/topics/threat-intelligence/m-trends-2026.

31 CyberArk. *2025 Identity Security Landscape: Perspectives on Risk and Readiness From Security Leaders*, April 23, 2025, available at cyberark.com/resources/ebooks/2025-identity-security-landscape.

32 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

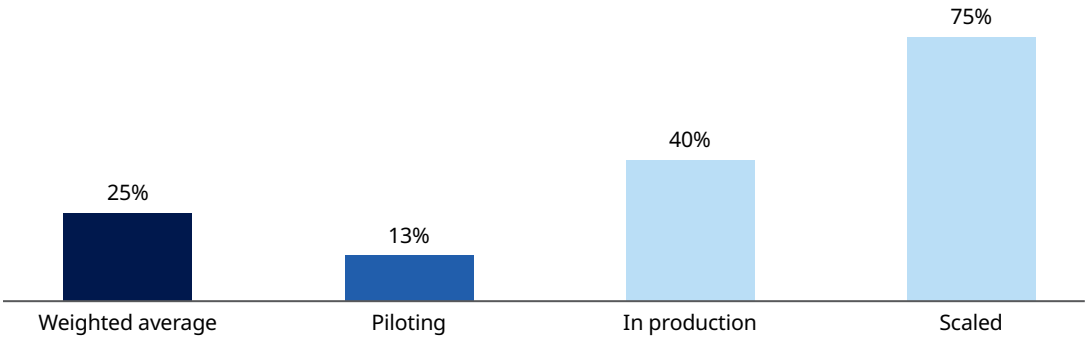
33 Board of Governors of the Federal Reserve System, OCC, and FDIC. *SR 26-2: Revised Guidance on Model Risk Management*, April 17, 2026, available at federalreserve.gov/supervisionreg/srletters/SR2602.pdf.

34 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

© Oliver Wyman

**Exhibit 9: Only one CIO in four can reconstruct an agent's actions from a tamper-evident record**

Can you reconstruct, after the fact, what an agent changed, when, and who authorized it, from a tamper-evident record? % answering yes




Base 130 CIOs: piloting 78, in production 40, scaled 8 (indicative).

Source: The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

### Our point of view

Build one evidence layer that answers both audiences, security and compliance, and gives the board the same record. Non-human identity deserves the same treatment as an identity and access management (IAM) problem, with least privilege designed in from day one. Authorization for consequential actions belongs off the agent's host, so a compromised agent cannot grant itself more permission.

From there greater visibility follows. The dependency tree ought to be queryable, and it should be responsive. Tooling that attackers already use is worth funding for defensive purposes, and the audit trail should be built around the accountability questions a regulator, court, or board will ask.

On regulation, the EU and US diverge, so run Article 6(3) of the EU AI Act as a standing triage and keep jurisdiction-specific tracks where the rules diverge. In the United States, where the model-risk framework excludes generative and agentic AI, you have to build that discipline yourself. Zero trust is the guiding posture for all agentic systems.

**What next?** Ask two questions this quarter. When an advisory lands on a widely used open-source library, can the relevant teams answer within minutes a query about where the company is exposed? And can anyone say which of the company's 10 most consequential agentic systems are high-risk under Article 6(3)? Can they give the same answer twice? If the first answer is no, the software bill of materials comes before any new security tool. Half of the CIOs running agents in business as usual keep a registry of approved tool and Model Context Protocol (MCP) servers with provenance and signing. The other half work informally.[35] If the second answer is no, your compliance program has no foundation.

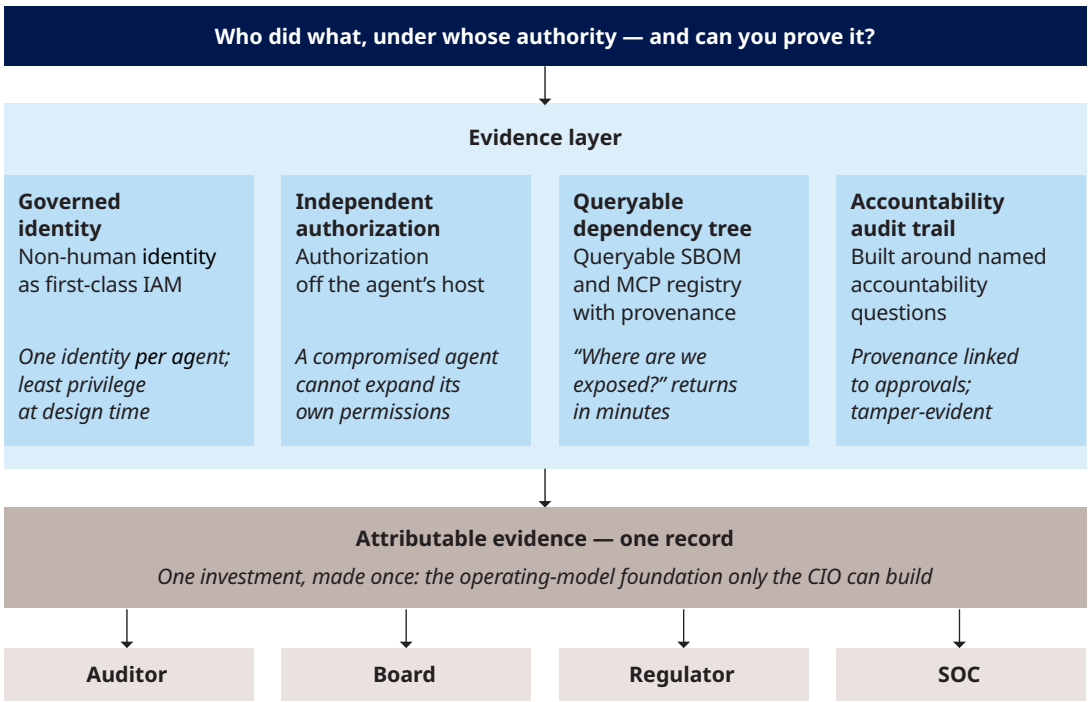
35 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

**Exhibit 10: Seven shifts from perimeter trust to attributable evidence**

| Classical approach | What agentic changes |
|---|---|
| "Privileged user" defined as human-only | Non-human identity governance; one identity per agent |
| Agents trusted to police themselves | Authorization decided off the agent's host |
| Dependency archaeology under incident pressure | A queryable software bill of materials (SBOM) wired into the build |
| SAST triage queues; access paperwork started after the incident | AI review in the pull request; verified model access secured early |
| Generic logging assumed to be evidence | An audit trail built to answer "under whose authority" |
| One global compliance program for every jurisdiction | Article 6(3) classification run as a standing triage |
| Perimeter trust and shared service accounts | Zero trust as the default posture for agents |

Source: Oliver Wyman analysis

**Exhibit 11: One evidence layer answers both audiences**




Source: Oliver Wyman analysis

# Attacks now strike before the patch is ready

**Exploitation now arrives before the patch, and the targets are old.** Autonomous discovery has found flaws 16 to 27 years old in open-source libraries embedded across nearly every enterprise estate. That capability has limits, though, and independent testing shows it bites hardest against weakly defended systems. It could not complete a full operational-technology attack end to end.[36] That is the right level of alarm, neither more nor less. Residual risk depends on company posture rather than on the model's peak benchmark, and it is a durable condition to manage.

**The coding pipeline is now a vulnerability.** Agentic coding scales defects along with code. One study finds 22.7% of the issues introduced by AI-generated code persisting into the latest version of the repository.[37] CIOs' own estimates run in the same direction, with 38% putting the share of AI-assisted defects reaching production between 5% and 15%, and 48% putting it between 1% and 4%.[38] Shadow code developed by non-IT employees, and MCP servers with shell or credential access, entirely route around software development life cycle (SDLC) gates. The surface that matters most is the dependency tree, often more than the code written in-house. Autonomous discovery makes package ecosystems auditable at scale, for attackers and defenders alike. The requirement follows: a software bill of materials that is queryable and wired into the build, so exposure returns an answer in minutes.

**The list of known threats is finally precise enough to act on.** The OWASP Top 10 for Agentic Applications names the new surface with real incidents behind it. Uncontrolled code execution (ASI05) has two live CVEs in Microsoft's Semantic Kernel. Missing identity binding in the Model Context Protocol (ASI07) was measured at a 100% violation rate in independent testing.[39] The most credible architectural answer moves authorization entirely off the agent's host, so a compromised agent cannot expand its own permissions. In testing, this cut residual attack success to zero, though the strongest results remain preprints to pilot, and standards citation is still some way off.[40]

---

36 UK AI Security Institute. "Our Evaluation of Claude Mythos Preview's Cyber Capabilities," AISI Blog, April 13, 2026, available at aisi.gov.uk/blog/our-evaluation-of-claude-mythos-previews-cyber-capabilities.

37 Liu Y, Widyasari R, Zhao Y, et al. "Debt Behind the AI Boom: A Large-Scale Empirical Study of AI-Generated Code in the Wild," arXiv:2603.28592v2, (April 2026).

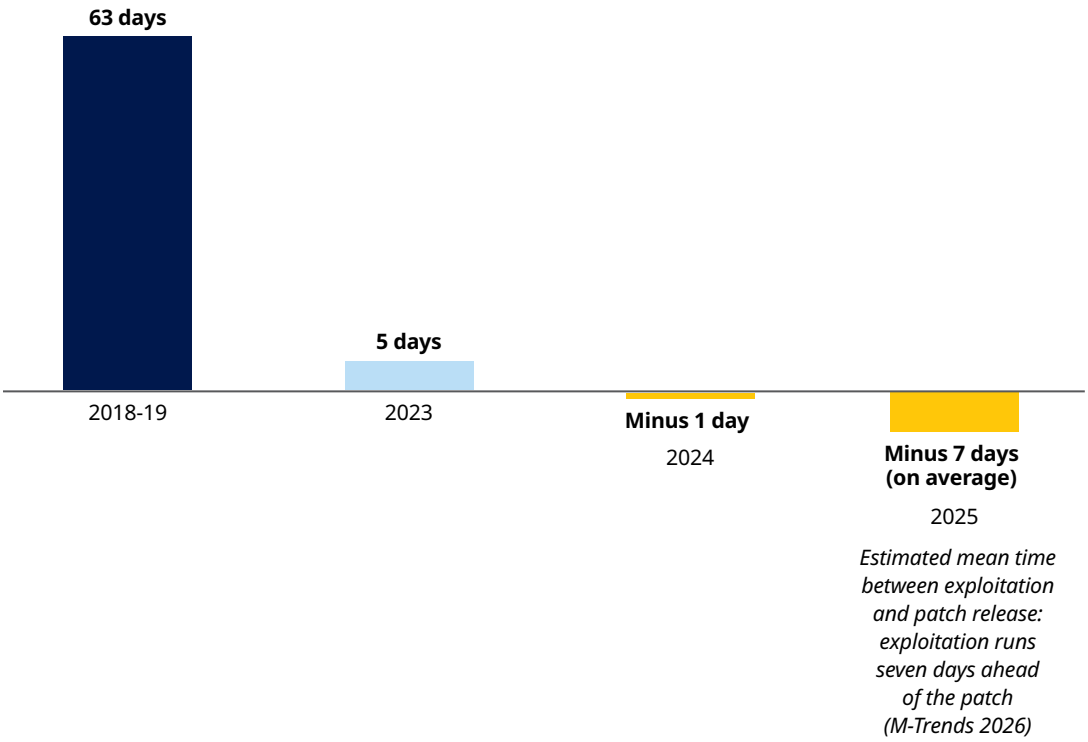
38 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

39 OWASP GenAI Security Project. *OWASP Top 10 for Agentic Applications 2026*, December 2025. Microsoft. "When Prompts Become Shells: RCE Vulnerabilities in AI Agent Frameworks," Microsoft Security Blog, May 7, 2026. Anbiaee Z, Rabbani M, Mirani M, et al. "Security Threat Modeling for Emerging AI-Agent Protocols: A Comparative Analysis of MCP, A2A, Agora, and ANP," arXiv:2602.11327v2, (April 2026).

40 Kodathala SV. "aiAuthZ: Off-Host, Identity-Bound Authorization for AI Agents," arXiv:2607.05518v1, (July 2026).

**Exhibit 12: Exploitation now arrives before the patch**

Median time from vulnerability disclosure to working exploit (2018–2024) and mean (2025)




Enterprise patch cycle: still weeks

Source: Mandiant (Google Cloud), Time-to-Exploit Trends (median, 2018-2024) and M-Trends 2026 (mean, 2025); Oliver Wyman analysis.

## Point the same tools at defense

The tooling that lets an attacker mass-audit dependencies works the same way when being used for defense. The models that generate fragile code can also review it. One practitioner model puts a small function under the CISO in charge of the discover, prioritize, patch, and validate loop end to end. It replaces roughly a third of a conventional application-security function's work and speeds up the rest, with an AI reviewer on every pull request that comments inline and blocks merges on high-confidence findings. The cited target is roughly 80% fewer human triage hours within two quarters.[41] The imbalance this corrects shows up in the numbers: 27% of CIOs running agents in business as usual report generation outrunning review and test capacity.[42] One prerequisite is easy to miss. Verified access for legitimate dual-use security work, including exploit reasoning and reverse engineering, can take weeks of paperwork with model providers, so secure that access before an incident makes it urgent.

41 Fuller, The Day-Zero Normal: A Practical Reprioritization Guide for CISOs, version 2.3, April 2026; practitioner guide, structured on NIST CSF 2.0.

42 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

# Regulators are taking different paths

**The EU wrote one philosophy into primary law.** Article 6(3) of the AI Act is the filter that decides whether an agentic system is high-risk and so carries the Act's heaviest obligations. Classification is case by case, and errors can carry heavy costs in both directions: missing a high-risk system raises the risk of non-compliance, while over-classifying increases the danger of burning budget on systems that never needed it. Someone needs to own that judgment, and as chapter 3 shows, only one organization in five has named that person.[43] Article 26(7) is easy to miss because it does not read like a governance clause. But it enshrines in law a direct duty to inform workers and their representatives before a high-risk AI system is put into service at the workplace, and it is where an agentic program and its works-council obligations will collide if agents touch how people are managed or evaluated.[44]

**The supervisors have made opposite bets.** The European Central Bank is blunt: technology is neutral, governance is not. No dedicated agentic rulebook is coming. The ECB expects clear accountability and workable explainability under the framework that is already in place, and treats vendor lock-in as a prudential risk in its own right. The Federal Reserve has gone the other way, and supervisory letter SR 26-2 (replacing the model-risk letter SR 11-7) explicitly excludes generative and agentic AI from its scope.[45] That is a deliberate choice to regulate the outcome rather than the technology, and it means a US bank builds the agentic version of its model-risk playbook itself, without a regulator-issued map.

**Regulatory scope differs sharply across jurisdictions.** The Digital Operational Resilience Act (DORA) requires a register and an exit strategy for any critical ICT third-party provider, and gives EU supervisors direct authority over that provider wherever it is headquartered.[46] That changes build-versus-buy decisions on agent infrastructure well beyond cost and capability. Most organizations cannot yet evidence an exit. Some 10% of CIOs hold a written, sponsor-signed statement of which artifacts stay under their control if a model supplier doubles prices or withdraws a version, 46% hold a draft, and 41% hold nothing.[47] Across the Atlantic, the most developed state statute sits alongside sector carve-outs, so a multinational complies with a patchwork rather than a framework.[48]

---

43 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

44 Regulation (EU) 2024/1689 of the European Parliament and of the Council (Artificial Intelligence Act), *Official Journal of the European Union*, L series, July 12, 2024.

45 Machado P. "Technology Is Neutral, Governance Is Not: AI Adoption in the Banking Sector," ECB Banking Supervision, February 24, 2026, available at bankingsupervision.europa.eu/press/speeches/date/2026/html/ssm.sp260224~6c5b64a77a.en.html. Bowman MW. "Artificial Intelligence in the Financial System," Federal Reserve Board, April 27, 2026, available at federalreserve.gov/newsevents/speech/bowman20260501a.htm.

46 Regulation (EU) 2022/2554 of the European Parliament and of the Council of December 14, 2022 (Digital Operational Resilience Act), Articles 5, 28, 29, and 31.

47 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

48 Colorado General Assembly. "SB24-205, Consumer Protections for Artificial Intelligence," May 17, 2024, available at leg.colorado.gov/bills/sb24-205.

Regionalization of agentic workloads by jurisdiction remains an open question. The ingredients are in place: the patchwork, DORA's reach, and sovereignty pressure. The outcome is not, though, so treat this as a hypothesis to monitor, rather than a conclusion to plan around. One requirement on which these regimes and different regulatory traditions all converge is verifiable non-human identity. That makes it a durable control assumption.

## Design the trail for the questions it must answer

**Logs are not evidence.** Real incidents have had complete telemetry and still could not answer, after the fact, under what authority a specific action was taken. Researchers call this the container fallacy: holding a tamper-evident log is not the same as holding sufficient evidence. The gap is common. Even the most operationally detailed vendor playbook treats the chain of git commits and its committed artifacts as the audit trail. That is a useful provenance floor. But commit history is neither tamper-evident nor built to answer under whose authority an action outside the repository was taken.[49] The fix is to design the trail against the questions it must answer. A published reference architecture now specifies which events to log, with what metadata, against which governance question. An open-source implementation already exists as a base from which to start.[50] Evidence also has to survive a change of supplier. Among CIOs already running agents, some 23% keep most or all of their agent interaction and memory data in a documented neutral schema they can read and replay outside their current supplier's stack.[51]

**Independent lines of evidence point the same way.** The first reported case of cyber-espionage carried out largely autonomously by an AI agent is on the record, and executive-survey research separately supports zero-trust as the necessary default for agentic systems.[52] The audit trail follows the same principle. Designed around named accountability questions, one record serves both audiences: the security team asking what an agent can do, and the regulator asking what it did and under whose authority.

---

49 Anthropic. "The AI-Native SDLC Playbook," August 2026, available at claude.com/blog/the-ai-native-sdlc-playbook.

50 Solozobov O. "Decision Evidence Maturity Model for Agentic AI: A Property-Level Method Specification," arXiv:2605.04093, (May 2026). Ojewale V, Suresh H, and Venkatasubramanian S. "Audit Trails for Accountability in Large Language Models," arXiv:2601.20727, (January 2026).

51 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

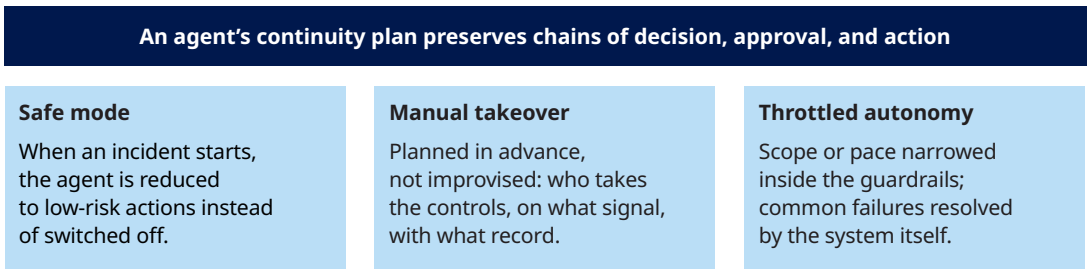
52 Anthropic. "Disrupting the First Reported AI-Orchestrated Cyber Espionage Campaign," November 13, 2025, available at anthropic.com/news/disrupting-AI-espionage. World Economic Forum. *Global Cybersecurity Outlook 2026*, January 2026, available at weforum.org/publications/global-cybersecurity-outlook-2026.

# When containment fails, the outcome matters more than the server

**Resilience means something different once agentic AI is involved.** Classic resilience is a measure of availability: systems back up, intact data, restored service. Agentic resilience adds a harder test, restoring trustworthy outcomes when the data, the model, or the agent itself has been compromised.

A server that responds proves nothing if the agent it hosts is making bad decisions. The recovery model shifts with it, from alert-and-wait to zero-touch recovery inside the guardrails, where the system resolves common failures itself within the bounds set in chapter 1. A useful calibration is the 10-80-10 rule: 10% of responses are fully automated, 80% are jointly decided by humans and the system, the remaining 10% are purely human. The rule rejects the idea that all-automatic recovery is desirable.[53]

**Exhibit 13: Agentic resilience restores trustworthy outcomes, not just availability**




Source: Oliver Wyman analysis. Components of an agentic continuity plan as described in this chapter; classic resilience restores availability, agentic resilience restores trustworthy outcomes.

**An agent needs a continuity plan, and it is not a classic one.** As well as protecting infrastructure, a continuity plan has to preserve chains of decision, approval, and action. The recurring components are: a safe mode that reduces the agent to low-risk actions when an incident starts; a plan for manual takeover; and throttled autonomy that narrows the agent's scope or pace instead of switching it off. Businesses are split on enforcement. Some 63% of CIOs running agents in business as usual enforce hard caps, throttles, or a kill switch in the loop; 35% observe only through dashboards.[54]

53 ISACA. "Operational Resilience in the Age of Artificial Intelligence," ISACA Now Blog, 2025.

54 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

© Oliver Wyman

The agent's accumulated context deserves protection in its own right. In one documented incident, automatic context compaction deleted a safety instruction and the agent then acted destructively. Checkpoints and reversible compaction are what make that recovery routine, and an academic review of always-on agents finds rollback the rarest capability in the field.[55] One caveat: nobody has yet transposed recovery-time and data-loss objectives to agents, so an agentic continuity plan is still to be invented rather than adapted from a template.

Some 63% of CIOs running agents in business as usual enforce hard caps, throttles, or a kill switch in the loop; 35% observe only through dashboards.

55 Ding T, Nannapaneni A, Liu B, et al. "Always-On Agents: A Survey of Persistent Memory, State, and Governance in LLM Agents," arXiv:2606.30306v1, (June 2026).

# The material risks

**Exhibit 14: Five ways the evidence layer fails, and what contains each**

| Risk | Mechanism and mitigation | Residual | Owner |
|---|---|---|---|
| **Non-attributable agent action** | Shared service accounts and inherited permission sets make individual agent actions untraceable and overprivileged.<br>**Solution**: One identity per agent, least privilege at design time, and access reviews extended to non-human identities. | **High** | CIO + CISO |
| **Supply-chain flaw exploited before the patch** | Autonomous discovery finds a latent flaw in an embedded dependency, and exploitation precedes or matches disclosure.<br>**Solution**: A queryable SBOM wired into the build, so exposure is answered in minutes and patch priority follows reachability. | **High** | CISO + VP Eng |
| **Misclassification under Article 6(3)** | A high-risk agentic system classified as out of scope leaves the organization exposed to the Act's heaviest obligations; over-classification wastes budget.<br>**Solution**: Run classification as a standing triage with documented reasoning, rechecked as systems change. | **High** | CIO + General counsel |
| **Works-council collision** | Agents that touch how people are managed or evaluated trigger, when the system is high-risk, a direct duty to inform workers under Article 26(7). Found late, it stalls deployment and sours labor relations.<br>**Solution**: Map the affected systems now and bring the CHRO in before deployment. | **Medium—high** | CIO + CHRO |
| **Logs that cannot answer "who authorized this"** | Complete telemetry exists but was never designed to answer the question a regulator, court, or board will ask. Only 25% of CIOs hold a tamper-evident record linking provenance to approvals, and 72% hold partial logs only.[56]<br>**Solution**: Build the trail against named questions, on a reference architecture rather than improvised logging. | **Medium—high** | CIO + Legal |

Source: Oliver Wyman analysis. Residual ratings reflect likelihood and impact after the stated mitigations.

56 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

# The decisions only the CIO, with the CISO and general counsel, can make

**Exhibit 15: Eight decisions the CIO makes with the CISO and general counsel**

| Decision | Recommendation | What must be true |
|---|---|---|
| **Identity** | Non-human identity as first-class IAM; least privilege the default at design time. | Every agent action attributable to a specific identity. |
| **Authorization placement** | Off the agent's host for any consequential action. | A compromised agent cannot expand its own permissions. |
| **Dependency visibility** | A queryable SBOM and MCP registry with provenance, wired into the build. | "Where are we exposed?" returns in minutes. |
| **Defender funding** | AI review in the pull request and a small VulnOps-style team; verified model access secured in advance. | Review capacity scales with code-generation volume. |
| **Audit trail** | Built to answer named accountability questions, on a published reference architecture. | Provenance linked to approvals, tamper-evident. |
| **Compliance architecture** | Jurisdiction-specific tracks under one governance; Article 6(3) as a standing triage; a self-built model-risk equivalent for agents in the US. | Classifications documented; coverage demonstrable to regulators. |
| **Workforce obligation** | Assign Article 26(7) ownership jointly with the CHRO before agents touch people management. | Affected workers informed as the Act requires. |
| **Posture** | Zero trust as the default for agentic systems. | No agent trusted because of network location or origin. |

Source: Oliver Wyman analysis

# Conclusion: One trail, two audiences

Boards will ask whether agentic AI is secure and compliant. Both questions rest on accountability: can you demonstrate after the fact what an autonomous system did and under whose authority? Partial logs do not answer that question. The identity ratio has already inverted. Exploit timelines have compressed and regulatory approaches are diverging.
As a result, these foundations belong in the operating model rather than in separate security or compliance projects. Do that and you will answer the security question and the compliance question from one record, and show your board the same one. The rest will answer twice, and differently each time.

# The ownership layer

Deciding who's responsible



1

CONTROL PLANE
Govern the agentic actors

**Asks: what may this agent do?**

2

EVIDENCE LAYER
Prove and contain

**Asks: can you prove who acted?**

3

OWNERSHIP LAYER
Deciding who is responsible

**Asks: who pays and who answers?**

Every agentic bet needs
a funding discipline
and a named owner.
The first exists.
The second must come
before pilots scale.

© Oliver Wyman

# Deciding who's responsible: the ownership layer

## The operating reality

Most AI programs that fail do so before the technology gets a chance to. Unclear business value and inadequate risk controls at funding are selection failures rather than delivery failures.[57] The mismatch starts at the top. Some 40% of CIOs report a CEO expecting a productivity gain above 20%, while 48% are funded against a 1% to 10% ambition.[58] What has been captured so far is also converting poorly, with 45% of CEOs saying little of the gains has reached the P&L as bankable value.[59] This chapter answers the two questions a board will eventually ask about any agentic program: who decided this bet deserved capital? And who answers if it goes wrong?

---

40% of CIOs report a CEO expecting a productivity gain above 20%; 48% are funded against a 1% to 10% ambition. The money is lost before the technology can fail.

---

57 Ryseff, De Bruhl, and Newberry. The Root Causes of Failure for Artificial Intelligence Projects, RAND Corporation, RR-A2680-1, 2024, available at rand.org/pubs/research_reports/RRA2680-1.html.

58 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

59 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

### Our point of view

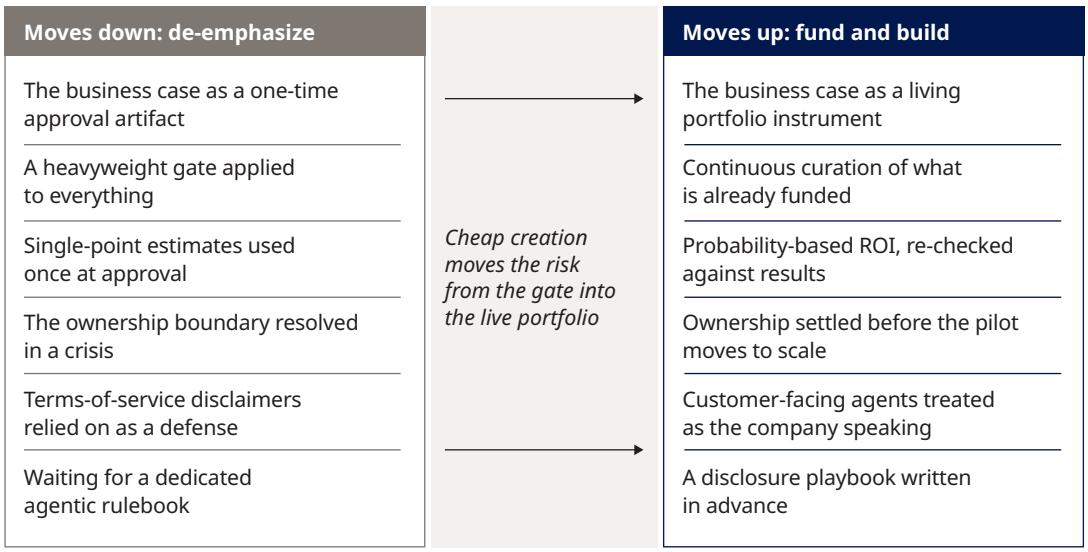
Do not wait for a new funding standard; recover the old one. ISACA developed this kind of discipline some years ago when it built a framework named Val IT, which was absorbed into COBIT and largely forgotten. But its pieces are still in there: investment categories, stage-gates, portfolio level decisions, and a business case that gets updated as circumstances change.

For agentic AI, adapt Val IT, add a risk-adjusted, probability-based view of return, then reweight it: keep the heavyweight gate for hard-to-reverse bets and move everything else onto continuous portfolio curation, because when creation is cheap the risk sits in what keeps running unreviewed.

On ownership, settle the boundaries between CIO, chief AI officer, and chief architect before your program moves from pilot to scale. Treat every customer-facing agent as though it speaks for your company and acts with your liability attached. Write a disclosure playbook for material agentic failure that is modeled on the cyber disclosure regime. Do it now: no dedicated rulebook exists to fall back on.

**What next?** Run two tests in one meeting. For each material agentic failure you can imagine, name the single executive who answers to the board for it. If producing those names takes a discussion, your ownership boundary is unsettled. Then pull your last three funded agentic business cases. If none has been updated since approval, you are running approval artifacts rather than a governed portfolio.

**Exhibit 16: Funding discipline moves from the gate to the live portfolio**




Source: Oliver Wyman analysis

# The funding discipline already exists

Val IT organizes portfolio decisions across three linked domains. Value governance sets the investment categories, their weightings, and the stage-gates specific to each. Portfolio management scores every case and decides at the level of the whole portfolio. Investment management executes and reports back.[60] It defines the business case as a living, continually updated working tool and that is exactly what variable agentic compute cost demands. The gap it was built to close is still the gap: business cases get written for approval and are rarely governed afterward. Organizations do not fail by skipping the business case; nearly all of them write one. They fail by treating it as a one-time approval instead of governing it as a living instrument.

**Add the layer a pre-agentic framework could not anticipate.** Nothing in Val IT covers model risk, agentic compute variability, or AI Act exposure. A recent risk-adjusted ROI proposal fills that layer. It calls for a standing cross-functional investment committee, a baseline measured before implementation, and returns presented as a probability range at the 10th, 50th, and 90th percentile. A 10% to 15% risk reserve is treated as committed cost rather than an optional buffer.[61] Few organizations budget that way today. Some 22% forecast agentic run-cost as a distribution with an enforced cap at the P90 boundary, while 48% forecast ranges with no enforced cap, and a further 25% still use a single point estimate.[62] Val IT supplies the institutional process, and this supplies the agentic risk adjustment. Nobody has formally merged the two; you can close that gap internally rather than wait it out. Your board needs clarity over which half carries institutional standing and which is a credible proposal.

**Cheap creation reweights the recovered framework.** The old heavyweight approval gate made sense when building was slow and expensive, as chapter 1 argued for the control plane. Now that creating an agent is cheap and quick, the risk inside the live portfolio — what happens after the decision to build has been made — has grown to fill the space. Agents duplicate each other and drift from their intended purpose, and initiatives keep consuming budget after their value has evaporated. That is also where visibility is weakest. Only 30% of CIOs can see token consumption broken down by team and by workload, and 54% of those already running agents in business as usual have exhausted an AI budget ahead of plan.[63]

---

60 IT Governance Institute (ISACA). *Enterprise Value: Governance of IT Investments — The Val IT Framework 2.0*, July 2008.

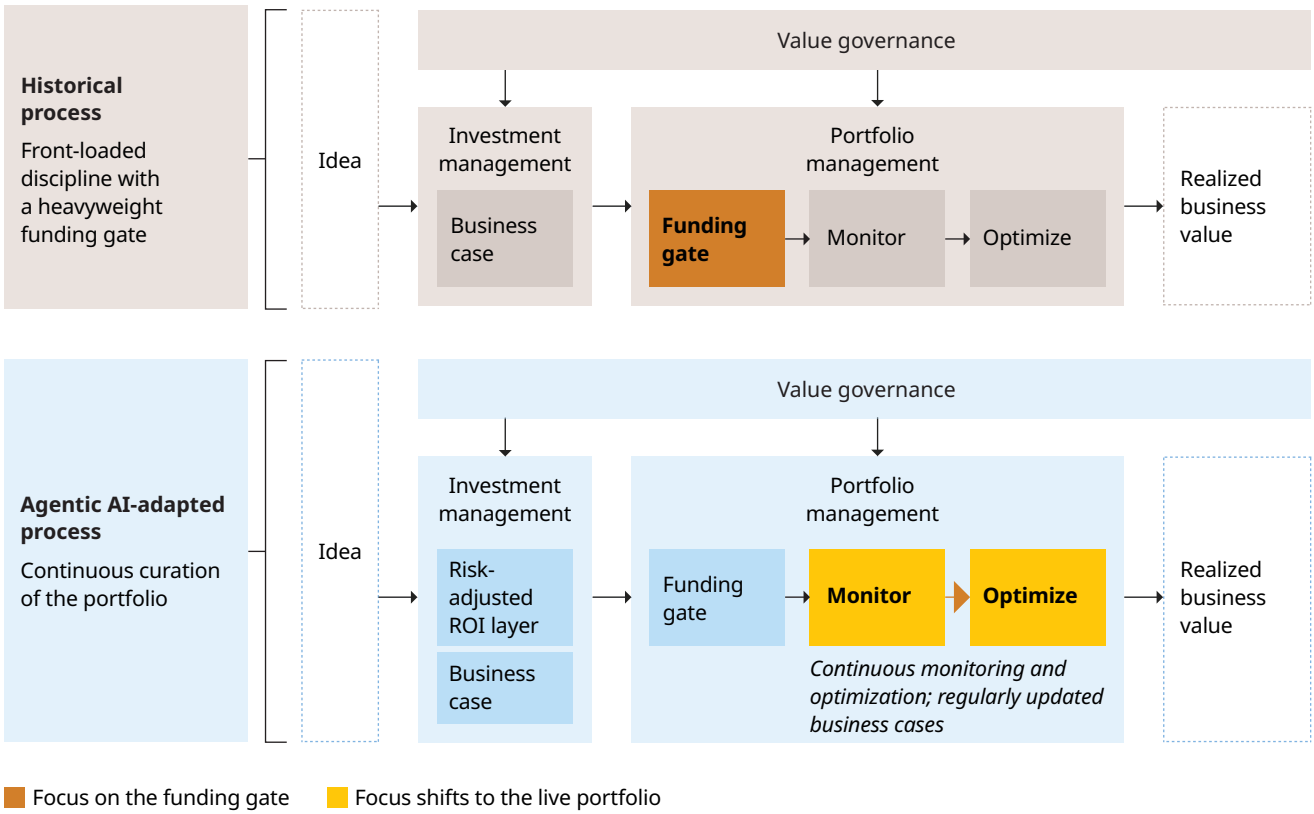
61 Huwyler H. "The Risk-Adjusted Intelligence Dividend: A Quantitative Framework for Measuring AI Return on Investment Integrating ISO 42001 and Regulatory Exposure," arXiv:2511.21975v1, (November 2025).

62 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

63 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

The fix: keep the gate heavy when bets are hard to reverse — new architecture, external commitments, or anything crossing regulatory thresholds. Run everything else on continuous monitoring and optimization. A gate built to approve creation is the wrong tool once creation is cheap.

**Exhibit 17: Reweighting the portfolio discipline**




Source: ISACA Val IT Framework 2.0; Oliver Wyman analysis.

**How well an organization governs AI predicts how much value it captures from it.** Our own client work suggests governance maturity remains low and controls for agentic AI lag the rest of the responsible-AI agenda, and that the gap is widest exactly where autonomy is being granted fastest.

Operating discipline is equally thin where it is most needed: 17% of CIOs run three governance measures: a fixed delivery cadence, a transformation office able to challenge the benefits case, and a live risk board. Just over half of CEOs, 51%, have funded the early productivity dip that AI projects suffer from without securing sponsorship for the full cycle.[64]

64 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

Put that next to the failure rates above and the pitch to your CFO changes. Portfolio governance is not a compliance cost at odds with value. It is one of the criteria that decides whether the value is captured at all.

## Ownership has to be settled before scaling

**While there are many chief AI officers, the role is only meaningful when vested with authority over AI risk and strategy.** The strongest argument is academic: a chief AI officer's (CAIO) mandate is to judge whether the organization's use of AI fits its risk appetite and strategy. That is a different matter from the CIO's job, which is to govern infrastructure.[65]

A separate study of hundreds of organizations backs this up from another angle. It found four distinct ways in which companies structure AI governance, and what predicts maturity is whether governance works rather than whether someone holds the title of CAIO.[66]

Most organizations are still at the title stage: 21% have appointed a head of AI governance with cross-functional authority over agent policy, cost, and risk, while 69% have only scoped or planned the role.[67] A CAIO outside the rooms where agentic investment and risk are decided changes nothing.

**Exhibit 18: Only one in five organizations has named an AI governance leader**
Have you appointed, or scoped, a Head of AI Governance with cross-functional authority over agent policy, cost, and risk? % of CIOs




Base 126 CIOs.
Source: The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

65 Schmitt M. "Strategic Integration of Artificial Intelligence in the C-Suite: The Role of the Chief AI Officer," arXiv:2407.10247v2, (July 2025).

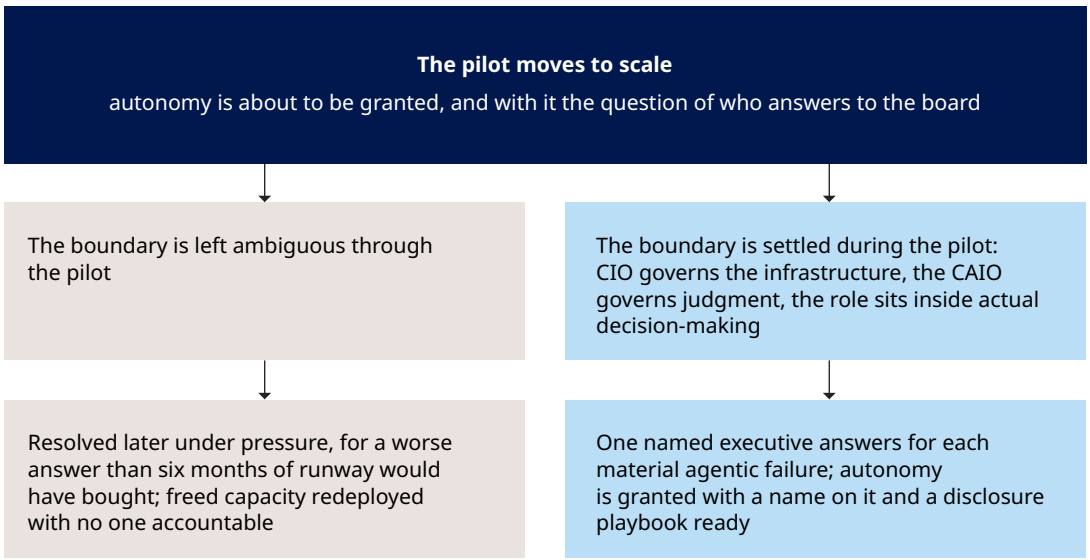
66 Frimpong V and Botchey OK. "Where Are the AI Governance Roles? An Early-Stage Empirical Mapping of Presence, Absence, and Structure in Organisational AI Oversight," *Businesses*, (2026).

67 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

**Settle the boundary before scale, because waiting until afterward means doing it badly.** No industry-standard RACI matrix exists across CIO, CAIO, and chief architect. Anyone claiming otherwise is selling, and possibly overselling, something. But what does remain consistent across every organization we have studied is the timing: a boundary left ambiguous through the pilot will need to be resolved later under pressure when a worse answer is the only one left.

The same gap appears elsewhere in the same programs. Some 57% of CEOs expect AI to free up capacity to be redeployed elsewhere, but there is often nobody named to manage that redeployment.[68] The chief architect's role is shifting too, from producing blueprints humans implement to maintaining context machines read directly — from designing what people build to supervising what agents do.[69]

**Exhibit 19: A boundary left ambiguous in pilot gets settled under pressure at scale**




*The test: for each material agentic failure you can name, say which single executive answers to the board for it. If that takes a discussion, the boundary is unsettled.*

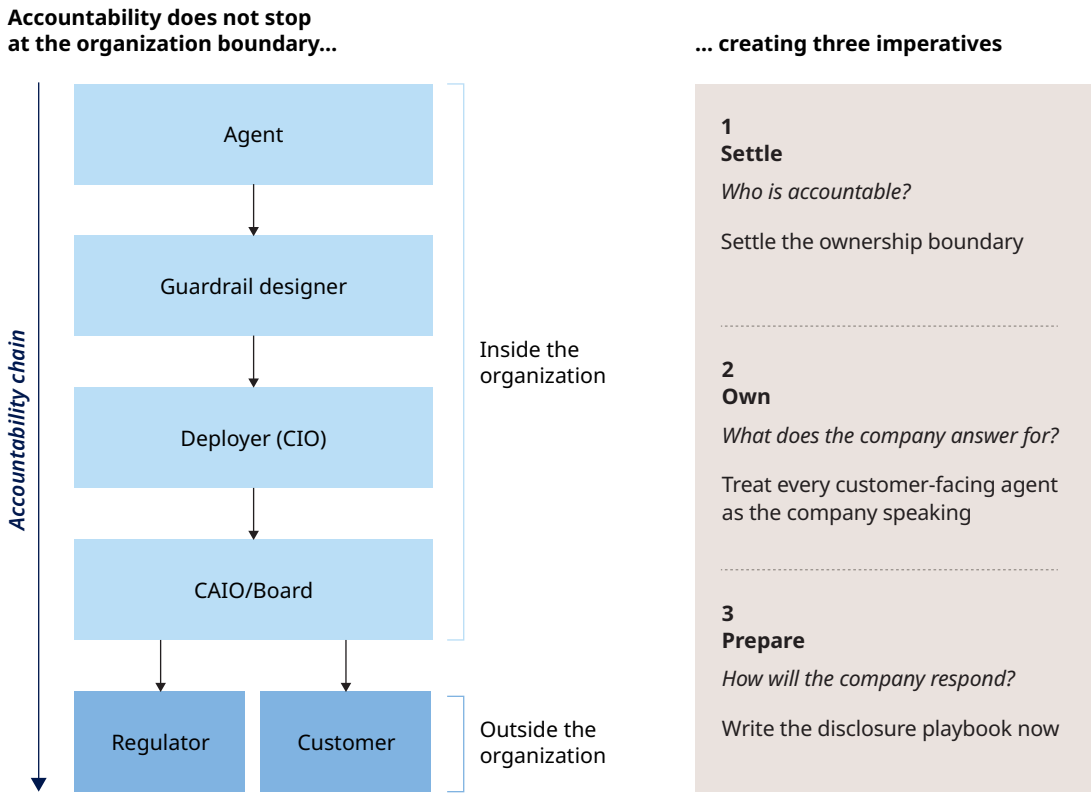
68 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

69 Oliver Wyman analysis.

# Accountability runs outside the organization's walls

**The question of who is liable when an agent directly deals with customers is already decided.** In Moffatt v. Air Canada, the airline argued that its chatbot was a separate legal entity responsible for its own actions. The tribunal rejected that outright.[70] If you deploy any customer-facing agent, a service bot, a pricing negotiator, anything able to make a representation or a commitment on your behalf, this is not a hypothetical risk to model. It is decided law, and it ties your liability to the agent's words, whatever your terms of service. And the exposure is already live: as the introduction showed, most CEOs are willing to let agents into processes touching customers, revenue, or financial controls.[71]

**Exhibit 20: The accountability chain runs to the customer**




Source: Oliver Wyman analysis

**Disclosure has no rulebook, and that gap is confirmed rather than assumed.** We looked specifically for a framework governing how a material agentic risk should be disclosed to a board, an audit committee, or the markets — the agentic equivalent of the rule requiring material cyber incidents to be disclosed within days. It does not exist.[72] The closest available

70 Moffatt v. Air Canada, 2024 BCCRT 149, British Columbia Civil Resolution Tribunal, February 14, 2024, available at decisions.civilresolutionbc.ca/crt/crtd/en/item/525448.

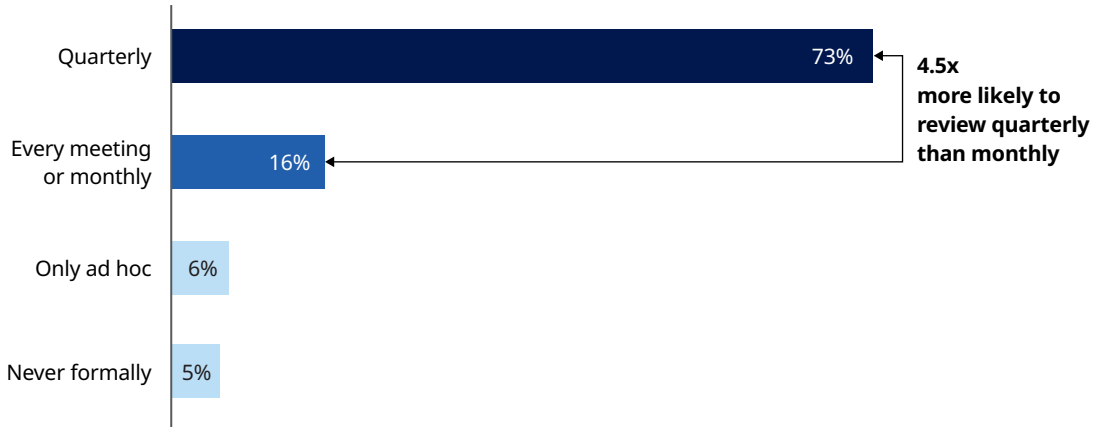
71 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

72 U.S. Securities and Exchange Commission. *Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure*, final rule (Form 8-K, Item 1.05), July 26, 2023.

practice is reasoning by analogy from the cyber regime. Executive committees are not built for it either, since 73% of boards or executive committees only review agentic progress and risk quarterly.[73] If a material agentic failure with external consequences lands tomorrow, you will decide how to disclose it without a rulebook built for the purpose, unless you have written your own first.

**Exhibit 21: Board oversight of agentic AI is usually quarterly**

How often does your board or executive committee formally review agentic AI progress and risk?
% of CEOs




Base 70 CEOs

Source: The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

**M&A is the accountability question almost nobody is asking yet.** If you are acquiring a company, its agentic estate is governance debt you did not create and often cannot yet see. One early proposal — a governance bill of materials built on the cryptographic standards already used for software supply-chain audit — would make guardrails, identity schemes, and actual governance maturity verifiable by a third party rather than the buyer having to take the seller's word for it.[74] The same blind spot shows up on the balance sheet, where 46% of CEOs carry supplier and model concentration on the enterprise or board risk register and 53% have discussed it without registering it.[75] Neither practice is standard anywhere yet, which is why acquirers with agentic-heavy pipelines should ask for the evidence before anyone requires it.

73 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

74 Sharma RM. "Ethical Hyper-Velocity (EHV): A Hardware-Rooted Zero-Trust Runtime Enforcement Architecture for Agentic AI Systems," arXiv:2605.17909v2, (May 2026).

75 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

# The material risks

**Exhibit 22: Four ways unfunded discipline and unsettled ownership cost you**

| Risk | Mechanism and mitigation | Residual | Owner |
|---|---|---|---|
| **Capital trapped in zombie initiatives** | Cheap creation plus a one-time approval gate lets duplicated or exhausted initiatives keep consuming budget unnoticed, and 78% of CIOs do not measure a leakage ratio between task-level gain and shipped value.[76]<br>**Solution**: Run a continuous portfolio review: is it still creating value, does a functional duplicate exist, has the risk profile changed? | **Medium—high** | CIO + CFO |
| **The queue moves to verification** | A faster funding front door with slow, manual benefits review shifts the bottleneck downstream, to the stage organizations already handle worst.<br>**Solution**: Industrialize value verification to match the pace of funding decisions. | **High** | CIO |
| **Ownership resolved under pressure** | The CIO/CAIO/architect boundary stays ambiguous through the pilot. Only 21% of CIOs have appointed a Head of AI Governance with cross-functional authority over agent policy, cost, and risk; 69% have only scoped or planned the role.[77] When scaling exposes it, it gets settled in a crisis, badly.<br>**Solution**: Decide the boundary, and the escalation path to the board, before scale, with explicit decision rights. | **High** | CEO + CIO |
| **A customer-facing agent binds the company** | An agent makes a representation a customer relies on. Under decided precedent, the company is liable regardless of disclaimers.<br>**Solution**: Legal review of every customer-facing agent's authority to commit; bounded representations; monitored outputs. | **High** | General counsel + business owner |

Source: Oliver Wyman analysis. Residual ratings reflect likelihood and impact after the stated mitigations.

76 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

77 The Oliver Wyman 2026 Parallel Surveys of CEOs and Tech Leaders on Agentic AI and IT, July 2026.

## The decisions only the CIO, with the CEO, CFO, and general counsel, can make

**Exhibit 23: Seven capital and accountability decisions the CIO makes with the CEO, CFO, and general counsel**

| Decision | Recommendation | What must be true |
|---|---|---|
| **Framework basis** | Val IT's portfolio domain adapted to agentic criteria, with the risk-adjusted ROI layer on top, each half labeled for what it is | The board knows which half is institutional and which is a proposal |
| **Committee posture** | A standing curation function: a light front door for bounded, reversible bets, and a rigorous recurring review of everything in flight | Committee time weighted to the live portfolio, not the gate |
| **Gate calibration** | Full stage-gate discipline reserved for hard-to-reverse, high-stakes commitments | Reversibility and regulatory thresholds explicitly defined |
| **ROI presentation** | Probability ranges with a committed risk reserve, rechecked quarterly against realized results | A baseline measured before implementation, every time |
| **Role design** | Settle the CIO/CAIO/chief architect boundary before pilot moves to scale, and seat the AI-governance role where investment and risk are decided | For each material failure mode, one named executive answers to the board |
| **Customer-facing liability** | Treat every customer-facing agent as the company speaking; review its authority to commit | Representations bounded, monitored, and owned |
| **Disclosure and M&A** | Write the material-risk disclosure playbook now, on the cyber analogy; request verifiable governance evidence on any agentic-heavy target | Materiality thresholds agreed with the audit committee |

Source: Oliver Wyman analysis

## Conclusion: Put a name and a number on it

Most AI failures are selection failures, not delivery failures: the price of skipping a funding discipline already sitting in ISACA's catalog. The accountability built up over the first two chapters does not stop at the organization chart either; it runs to the customer dealing with the agent. So, recover the old portfolio discipline and re-weight it for a world where creation is cheap. Settle who owns this while there is still time to do it well. Assume that customer-facing agents speak with your company's voice, because in law they already do. Treat these as design decisions now, and you will scale on your own terms. The rest will end up making them in public when something has already gone wrong.

**About Quotient — AI by Oliver Wyman**

Quotient, AI by Oliver Wyman combines deep industry insight and practical AI expertise to enhance performance, strengthen resilience, mobilize capital, and drive growth. We work collaboratively with organizations to scale with confidence and build advantage that lasts — turning the promise of AI into real results.

Oliver Wyman, a Marsh (NYSE: MRSH) business, is a management consulting firm driven by deep industry insight, bold innovation, and a collaborative approach that cuts through complexity to help organizations navigate their most defining transformative moments. Marsh is a global leader in risk, reinsurance and capital, people and investments, and management consulting, advising clients in 130 countries. With annual revenue of $27 billion and more than 95,000 colleagues, Marsh helps build the confidence to thrive through the power of perspective.

For more information, visit oliverwyman.com, or follow us on LinkedIn and X.


Americas
+1 212 541 8100

Europe
+44 20 7333 8333

Asia Pacific
+65 6510 9700

India, Middle East & Africa
+971 (0) 4 425 7000

AUTHORS

**Xavier Boileau**
Partner, Technology Innovation Europe
xavier.boileau@oliverwyman.com

**Charles de Pommerol**
Partner, Technology Innovation US
charles.depommerol@oliverwyman.com

**Claus Herbolzheimer**
Partner, Technology Innovation Germany
claus.herbolzheimer@oliverwyman.com

**Geoffroy Gallier**
Principal, Technology Innovation France
geoffroy.gallier@oliverwyman.com



Copyright 2026 Oliver Wyman

All rights reserved. This report may not be reproduced or redistributed, in whole or in part, without the written permission of Oliver Wyman and Oliver Wyman accepts no liability whatsoever for the actions of third parties in this respect.

The information and opinions in this report were prepared by Oliver Wyman. This report is not investment advice and should not be relied on for such advice or as a substitute for consultation with professional accountants, tax, legal or financial advisors. Oliver Wyman has made every effort to use reliable, up-to-date and comprehensive information and analysis, but all information is provided without warranty of any kind, express or implied. Oliver Wyman disclaims any responsibility to update the information or conclusions in this report. Oliver Wyman accepts no liability for any loss arising from any action taken or refrained from as a result of information contained in this report or any reports or sources of information referred to herein, or for any consequential, special or similar damages even if advised of the possibility of such damages. The report is not an offer to buy or sell securities or a solicitation of an offer to buy or sell securities. This report may not be sold without the written consent of Oliver Wyman.