Digital CxO

POWERED BY Techstrong | Group
a Futurum company

SPECIAL REPORT · THE LEADERSHIP ISSUE · AUGUST 2026

No Maintenance Window

What Technology Leaders Should Keep, Change and Wait On as Agentic AI Moves Into Production

The executive operating model for turning agentic experimentation into completed, accepted work.

By Alan Shimel
EDITOR IN CHIEF

PUBLISHED BY DIGITALCXO

DIGITALCXO

No Maintenance Window

The enterprise objective is not maximum autonomy or maximum AI adoption. It is durable production.

Written by **Alan Shimel**
Editor in Chief

Technology leaders must preserve the disciplines that make production dependable, change how those disciplines operate at agent speed, and withhold broader authority until a workflow earns it.

Audience: CIOs, CTOs, CISOs, CDOs, CDAOs, CAIOs and the platform, architecture, SRE, FinOps and risk leaders who support them.

Research note: Statistics are reproduced from the sources listed in the report. Futurum survey results are proprietary research and require publication permission. Forecasts and vendor-reported claims remain labeled as such.

THE EXECUTIVE BRIEF

Keep: reliability, least privilege, lineage, modular architecture, change discipline, human ownership and outcome measurement.

Change: move governance into runtime execution; give agents identities; measure completed-work economics; fund shared controls.

Wait: delay broader autonomy, custom foundation models and irreversible coupling until evidence justifies the commitment.

Act: inventory the estate, apply eight production gates and make explicit scale, hold or stop decisions within 90 days.

52.6%
Still researching or piloting agentic AI [1]

6.3%
Exception-governed autonomous ecosystems [1]

18.1%
Top-two-box maturity for AI-agent governance [2]

8
Gates to durable production

INSIDE THE REPORT

Leadership without a maintenance window

How to use this report

Read the executive summary for the central operating thesis. Use **Keep / Change / Wait** to test the current portfolio. Apply the eight gates before expanding autonomy. Finish with the 90-day agenda and assign dates and owners.

Reader promise: determine which workflows deserve investment, who must own them, which controls belong in runtime and what evidence is required before autonomy expands.

Completed work

Measure accepted outcomes, not attempts or generated output.

Bounded agency

Autonomy is an operating setting, not a permanent label.

Human ownership

Every consequential workflow needs a named stop authority.

Questions this report answers

- Which agentic workflows deserve investment?
- Who approves production and who can stop it?
- Which controls must operate in runtime?
- What evidence is required before autonomy expands?
- What should the leadership team do next Monday?

Executive tools inside

- The four stages of durable production
- The ten connected constraints
- Five responsibilities for one outcome
- The eight-gate production test
- A 90-day leadership agenda

Executive Summary: The Agentic Reality Check

> There is no maintenance window for this transition.

Technology leaders are being asked to rebuild the engine while the enterprise is still flying. Existing services must remain available, secure and compliant. Business units want immediate access to new AI capabilities. Boards and CEOs want measurable results. Employees are already using tools that central technology teams may not have selected, integrated or even discovered. Meanwhile, the technology itself is moving from systems that generate answers toward systems that can choose tools, retrieve data, invoke applications and take actions.

There is no maintenance window for this transition.

Agentic AI is not one product category or one level of autonomy. It is a spectrum of systems that can pursue a goal across multiple steps with some discretion over how the work gets done. At the low end, an agent might assemble information and ask for approval before every action. At the high end, it might coordinate other agents and act across several enterprise systems with humans reviewing exceptions rather than every step. The leadership challenge is deciding where on that spectrum each workflow belongs, then building the controls, ownership and economics required to keep it there.

The evidence shows why this cannot be treated as another round of tool adoption. Futurum's 1H 2026 AI Platforms survey found that 52.6% of respondents were still researching or piloting agentic AI, while only 6.3% described an autonomous ecosystem governed mainly by exception.[1] McKinsey's 2026 global survey found that 44% of respondents said AI was scaling across their enterprise, yet only 37% attributed any EBIT impact to AI and only 6% met McKinsey's definition of an AI high performer. At the same time, 80% reported personal productivity gains.[4] Individual progress is real, but it does not automatically become enterprise performance.

Technology leaders also face a control gap. IBM found that two-thirds of the 2,000 CIOs and CTOs it surveyed were accountable for AI systems they did not fully control, 70% said business teams were deploying technology faster than IT could track, and only 11% felt fully prepared for the expected scale of agent deployment.[5] Deloitte found that 71% of surveyed organizations already had five or more technology leaders. Its respondents were confident about their ability to deploy and govern AI, but three-quarters still said their operating model needed to change within 12 to 18 months.[6]

The practical issue is not whether companies can start AI projects. Most can. The issue is whether they can finish the work around those projects: integration, permissions, evaluation, observability, recovery, cost control, organizational adoption and accountable ownership. A model can be deployed while the surrounding production system remains unfinished.

This report proposes three decisions for technology leaders:

Keep the disciplines that make production dependable. Reliability engineering, least privilege, data lineage, change management, incident response, architecture standards, human accountability and outcome measurement matter more when software can act on its own initiative.

Change how those disciplines are implemented. Governance must move into the runtime and platform. Identity programs must include agents and other non-human actors. Funding must cover the operating lifecycle, not just the pilot. Success measures must move from tasks completed and content produced to accepted business outcomes, including the fully loaded cost of humans, agents and deterministic systems.

Wait before granting broad autonomy where evidence, controls or economics remain weak. Waiting does not mean ignoring the technology. It means testing within boundaries while withholding irreversible authority until a workflow earns it.

The objective is not maximum autonomy. It is durable production: work that can be completed repeatedly, observed, secured, recovered, economically defended, adapted and owned. The strongest technology organizations will not be those that deploy the most agents. They will be those that finish more valuable work safely and learn from every completed cycle.

PART I

Leading Without a Maintenance Window

1. The Technology Leader's Job Has Changed

For much of the modern enterprise technology era, the basic leadership bargain was understandable. The CIO kept the business running and delivered enterprise systems. The CTO shaped architecture and engineering. The CDO or CDAO governed data and analytics. The CISO managed cyber risk. Newer CAIO roles coordinated AI strategy, models and responsible-use programs. The boundaries were never perfect, but the systems themselves generally remained attached to a recognizable owner and a relatively stable operating process.

Agentic AI cuts across those boundaries by design. An agent may run on a platform managed by the CTO, use context governed by the CDO, act through credentials overseen by the CISO, alter a service owned by the CIO and belong to a portfolio directed by the CAIO. The business outcome may belong to someone outside the technology organization. When the workflow succeeds, the ownership ambiguity may be easy to ignore. When it exposes data, changes a production record, incurs an unexpected bill or interrupts a customer process, the seams become visible immediately.

The mandate has expanded at the same time. Deloitte's 2026 Global Technology Leadership Study found that 79% of surveyed technology leaders identified driving business outcomes as their top priority. Yet 42% reported low or no return on AI investments. The same study found 81% were confident their operating model could deploy and govern AI across the enterprise, while 75% said that model still needed to change within the next 12 to 18 months.[6] That is not necessarily a contradiction. Many organizations have assembled enough capability to deploy AI. Far fewer have settled the ownership, funding, workflow and measurement questions required to sustain value.

IBM's findings make the operating pressure even clearer. Eighty percent of surveyed technology leaders said their AI transformation mandate came directly from the CEO. Only 11% said they were fully prepared for the scale of agent deployment expected during the next year. Seventy-seven percent said adoption was already outpacing governance capability.[5] Leaders are being asked to accelerate and control the same transition, often with governance and architecture built for slower, more deterministic software.

The pressure is pointed directly at the technology estate. Among the 766 respondents to Futurum's question about likely agentic deployments during the next 18 months, 49.2% selected IT operations or cybersecurity, the largest response category. At the same time, 78.4% of the full AI Platforms sample expected AI budgets to increase during the next 12 months.[1] Technology leaders are not preparing for a distant possibility. They are deciding how to absorb funded, operational demand now.

Traditional applications execute code paths that engineers define in advance. Agentic systems introduce probabilistic decisions into the execution path. A team may define the goal, available tools and boundaries, but it cannot enumerate every route the system will take. This changes what operational control means. Testing a feature is no longer enough. Leaders must also understand what context can influence the agent, which actions it can take, whose authority it inherits, how its behavior is evaluated and what happens when it does something unexpected but technically permitted.

The technology leader's job therefore shifts in four ways.

First, leadership moves from approving tools to designing systems of work. The decision is not simply whether employees can use a model. It is how work should move among people, agents and conventional software, and where judgment, approval and responsibility remain human.

Second, control moves closer to execution. Policies written in a governance document matter, but a policy that cannot be enforced when an agent calls a tool is an aspiration. Runtime identity, permissions, budgets, action policies, logging and approval gates become part of the operating architecture.

Third, success moves from adoption to outcomes. Usage, prompts, generated code and hours saved can help diagnose progress, but they are not the result. The result is a resolved incident, completed onboarding, fulfilled order, approved claim, safer release or faster customer response at an acceptable cost and risk level.

Fourth, leadership becomes more collective without becoming less accountable. Deloitte found that 71% of organizations had five or more technology leaders.[6] Agentic AI gives those leaders more reason to operate as a team. It does not give them permission to leave an outcome without a named owner.

2. Production Is Not the Finish Line

> The objective is not maximum autonomy. It is durable production.

The word production has become too easy to claim. A model endpoint serving live traffic is in production. An assistant available to employees is in production. An agent used by a small team to update a low-risk system may also be in production. Those statements describe deployment status. They do not tell leaders whether the capability is dependable, economical or ready to scale.

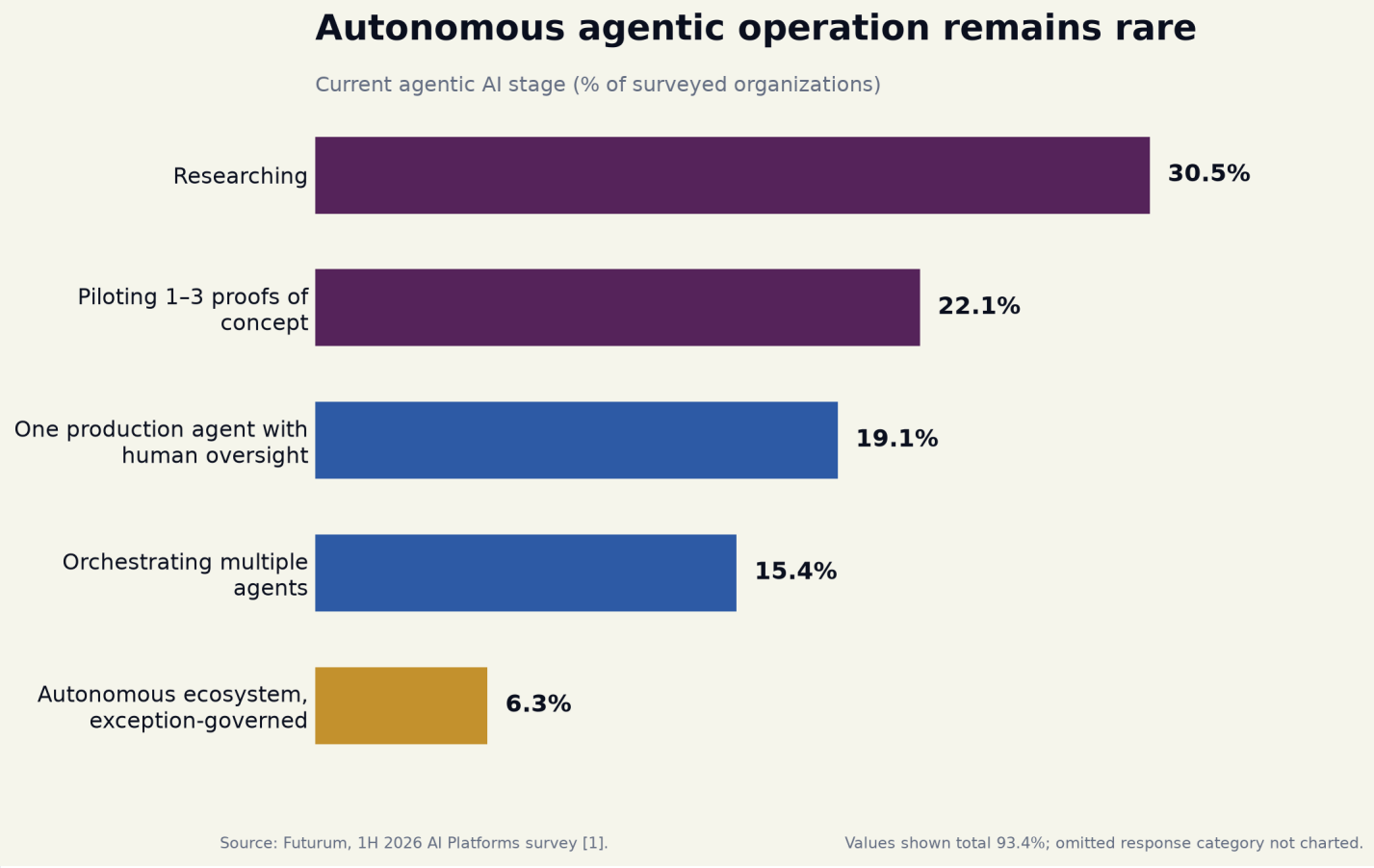


FIGURE 01 · AUTONOMOUS AGENTIC OPERATION REMAINS RARE

A more useful leadership model has four stages.

1. **Experiment:** The organization is testing whether the idea is technically feasible and potentially valuable. Data may be sampled, controls may be manual and the workflow may not yet have a production owner.
2. **Deployment:** The capability is available in a real environment for bounded use. It may have real users and integrations, but it is still proving its reliability, adoption and operating model.
3. **Production:** The capability participates in a live workflow and affects real work, customers, employees or systems.
4. **Durable production:** The workflow can be repeated, observed, secured, recovered, economically justified, adapted and retired. It has defined acceptance thresholds and a named human owner.

The fourth stage is where AI becomes an enterprise capability instead of a recurring demonstration.

Futurum's AI Platforms survey illustrates the distance still to travel. In 1H 2026, 30.5% of respondents said they were researching agentic AI and 22.1% were piloting one to three proofs of concept. Another 19.1% were deploying a single agent in production with human oversight, 15.4% were orchestrating multiple

agents, and 6.3% described an autonomous ecosystem governed mainly by exception.[1] These categories show movement, but they do not measure durable production. There is also no clean Futurum statistic for the percentage of pilots that become production systems, and the broader research market does not provide one universal conversion rate that can be applied responsibly across industries and use cases.

That absence matters. Leaders should reject the widely repeated but weakly sourced idea that one failure percentage explains the market. Gartner has forecast that more than 40% of agentic AI projects will be canceled by the end of 2027 because of escalating costs, unclear business value or inadequate risk controls.[16] That is a forecast, not an observed pilot-failure rate. Its value is as a warning about the conditions that may cause cancellation, not as proof that a particular initiative is destined to fail.

The practical response is to stop treating production deployment as the final approval. A live workflow should enter a managed operating lifecycle. Its quality, cost, exception rate, human review burden and failure modes should be evaluated continuously. Model changes, new tools, altered data sources and new permissions should be treated as material changes to the production system. The question after launch is not only, "Is it running?" It is, "Can we continue to trust, afford and own the result?"

PART II

Why More AI Does Not Automatically Mean More Progress

3. The Transition Tax

> The work has not disappeared. It has moved.

AI can reduce the effort required to produce an answer, a document, a block of code or a recommended action. That benefit is useful, but it is only one part of a workflow. When one stage accelerates faster than the stages around it, the constraint moves downstream.

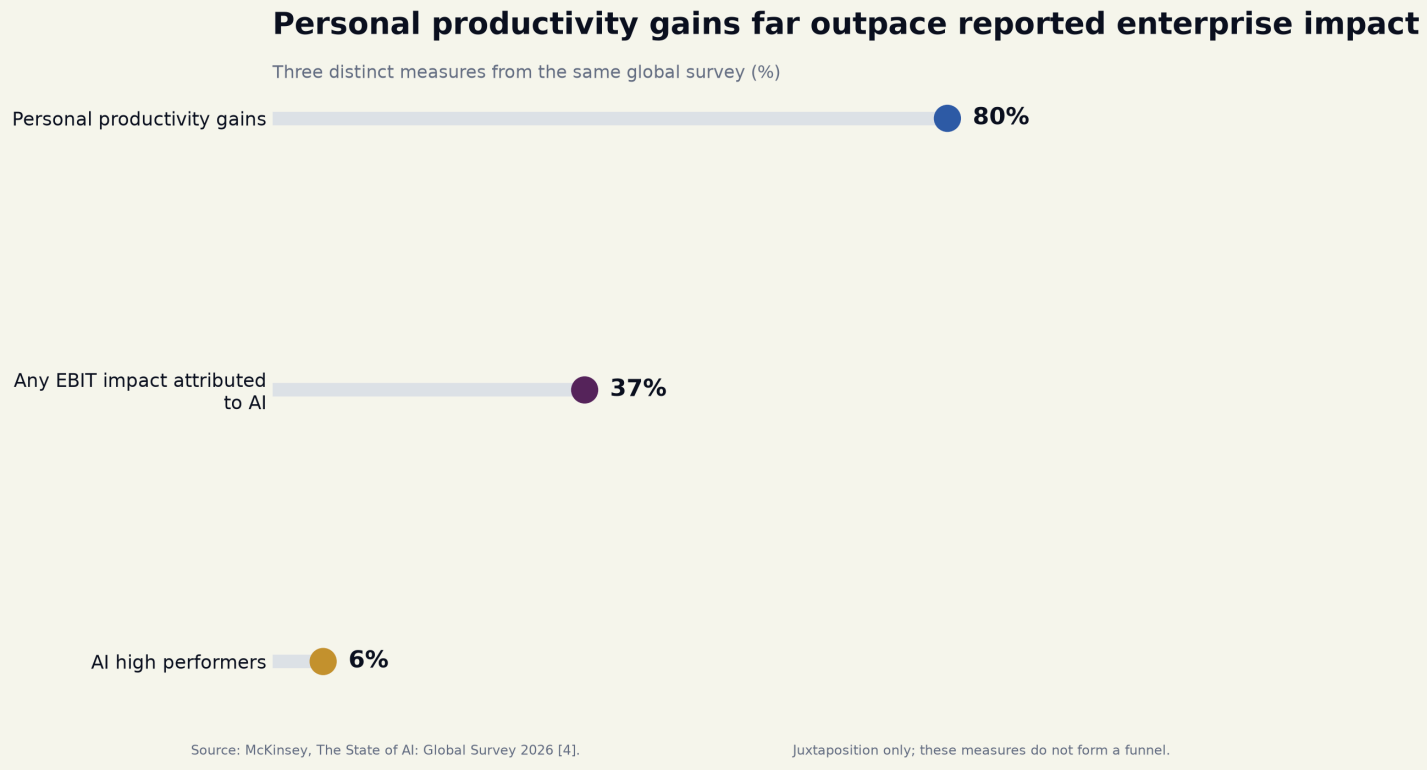


FIGURE 02 · PERSONAL PRODUCTIVITY GAINS FAR OUTPACE REPORTED ENTERPRISE IMPACT

This is the transition tax: the cost of redesigning workflows, controls, roles and systems around a new form of execution. Some of the tax is temporary. Teams need time to learn new tools, establish practices and integrate them into daily work. Some of it is structural. Probabilistic systems require ongoing evaluation, review and monitoring that deterministic software may not require in the same form.

DORA's 2025 research describes AI as an amplifier of the organization in which it operates. Based on nearly 5,000 technology professionals and more than 100 hours of qualitative research, DORA found a positive relationship between AI adoption and software delivery throughput and product performance, but a negative relationship with software delivery stability. Strong automated testing, version control, fast feedback and loosely coupled architecture helped determine whether acceleration became value or instability.[7]

DORA's 2026 work on AI return on investment gives the early period a useful J-curve shape. It identifies a learning curve, a verification tax and pipeline adaptation. When developers generate more code, testing and approval may become the bottleneck. When people spend less time creating the first version, they may spend more time reviewing, correcting and integrating it.[8] The work has not disappeared. It has moved.

Futurum's SLE survey shows why this deserves executive attention. Among 839 respondents, 53.7% reported that AI was involved in at least half of their software lifecycle work. Yet the dominant mode remained individual developer assistance for 47.2% of respondents. Supervised agents accounted for 18.4%, semi-autonomous agents for 13.6%, and autonomous end-to-end agents for 5.8%.[2] AI involvement is already substantial, but agentic autonomy remains limited.

The distribution of AI use across the lifecycle is also revealing. Futurum found 40.2% using AI for code generation and 37.7% for code review, compared with 28.0% for testing, 21.0% for observability, 16.2% for incident response, 13.2% for CI/CD operations, 12.4% for security scanning and 6.2% for deployment decisions.[2] These are independent multi-select categories, not a measured funnel. Still, the pattern is consistent with a broader operational concern: use is more common where AI creates or reviews output than where it assumes higher-consequence operational authority.

The same SLE survey found the leading factors limiting more AI use were almost tied: token or compute cost at 43.7%, code quality at 43.4%, lack of governance at 42.4%, and velocity exceeding oversight at 39.6%.[2] The constraint is not simply model capability. It is the organization's ability to absorb and govern what the model produces.

The maturity gap inside the same survey is equally important. The top-two-box maturity score was 58.3% for DevOps and 52.1% for observability, but only 32.9% for AI-assisted development and 18.1% for AI-agent governance.[2] Established operating disciplines are further along than the practices needed to govern agentic work. That gap supports extension, not abandonment, of what technology organizations already know how to do.

One SLE result deserves special editorial caution. A derived 75.2% of respondents reported one or multiple production incidents to which AI had contributed, with another 17.5% reporting a suspected but unconfirmed incident.[2] The values were returned directly by the survey tool, but the result is unusually high and based on self-reporting. It should be verified against the source report before publication, should not be used as a causal claim, and should not become the headline. Even with that caution, its presence reinforces the need to treat AI-generated and agent-executed work as production activity rather than informal experimentation.

This distinction explains why individual productivity can rise without a matching enterprise result. McKinsey found 80% of respondents reporting improved individual productivity, while 37% attributed any EBIT impact to AI and only 6% qualified as high performers.[4] Those figures come from one survey, but they should not be interpreted as a causal calculation. They show a gap between two different kinds of self-reported impact.

Microsoft's 2026 Work Trend Index reached a similar conclusion from a different population. Its research surveyed 20,000 AI users in 10 countries and analyzed Microsoft 365 activity. Fifty-eight percent said they were producing work they could not have produced a year earlier. Yet Microsoft reported that organizational factors such as culture, manager support and talent practices accounted for 67% of the variation in reported AI impact, compared with 32% for individual factors such as mindset and behavior.[9]

Gallup's 2026 workplace research reinforces the organizational point. Frequent AI use was much higher among employees who strongly agreed that AI integrated with existing systems and processes, 86% versus 52%. It was also higher when managers actively supported AI use, 79% versus 46%.[10] Access to a tool is not the same as integration into work.

Technology leaders should therefore measure the whole workflow. If an agent reduces preparation time by 50% but doubles review queues, increases exception handling or creates more production instability, the local productivity gain may have little enterprise value. Useful measures include end-to-end cycle time, accepted outcomes, rework, exceptions, human review minutes, incident rate, recovery time and fully loaded cost. This is how leaders determine whether AI throughput has become organizational throughput.

4. There Is No Single Bottleneck

The search for one reason AI projects stall is understandable. A single cause is easier to fund and easier to assign. Improve the data. Buy a better model. Add a governance committee. Hire more specialists. Each may help, but the available evidence does not support one universal bottleneck.

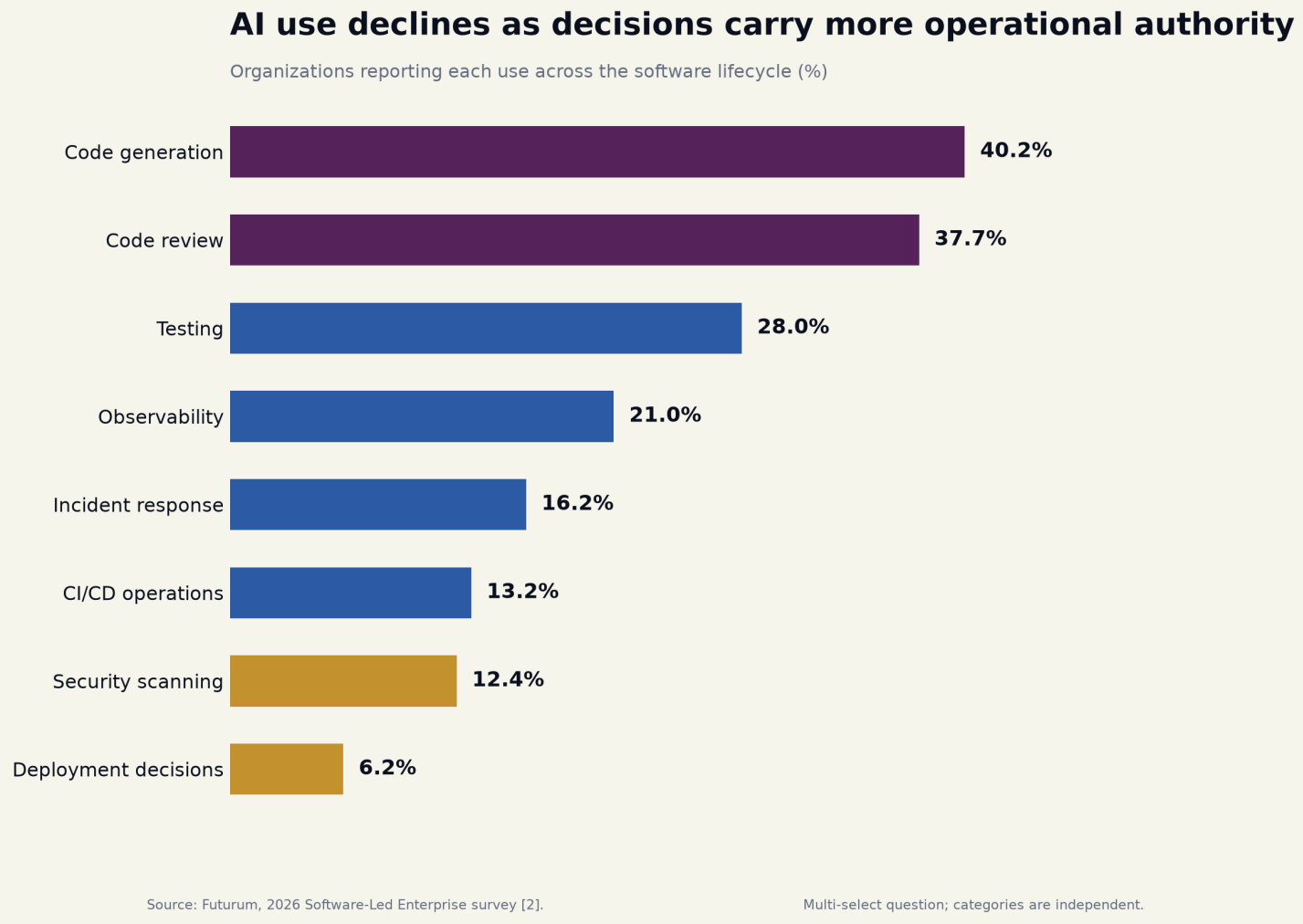
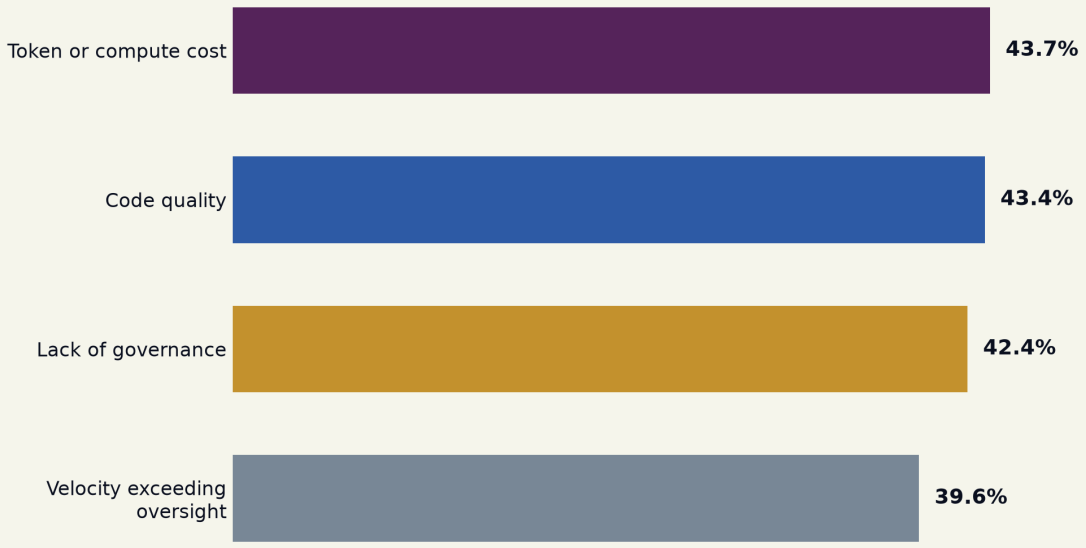


FIGURE 03 · AI USE DECLINES AS DECISIONS CARRY MORE OPERATIONAL AUTHORITY

Cost, quality and governance constrain AI use at nearly equal rates

Organizations selecting each limiting factor (%)

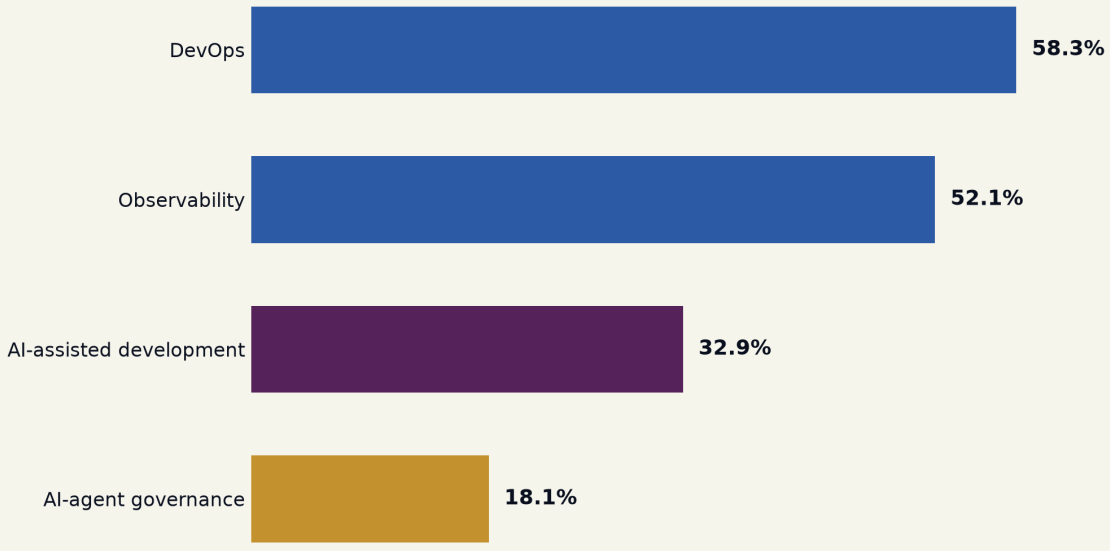


Source: Futurum, 2026 Software-Led Enterprise survey [2]. Multi-select question.

FIGURE 04 · COST, QUALITY AND GOVERNANCE CONSTRAIN AI USE AT NEARLY EQUAL RATES

Agent governance maturity trails DevOps by 40 percentage points

Top-two-box maturity score by operating discipline (%)



Source: Futurum, 2026 Software-Led Enterprise survey [2].

FIGURE 05 · AGENT GOVERNANCE MATURITY TRAILS DEVOPS BY 40 PERCENTAGE POINTS

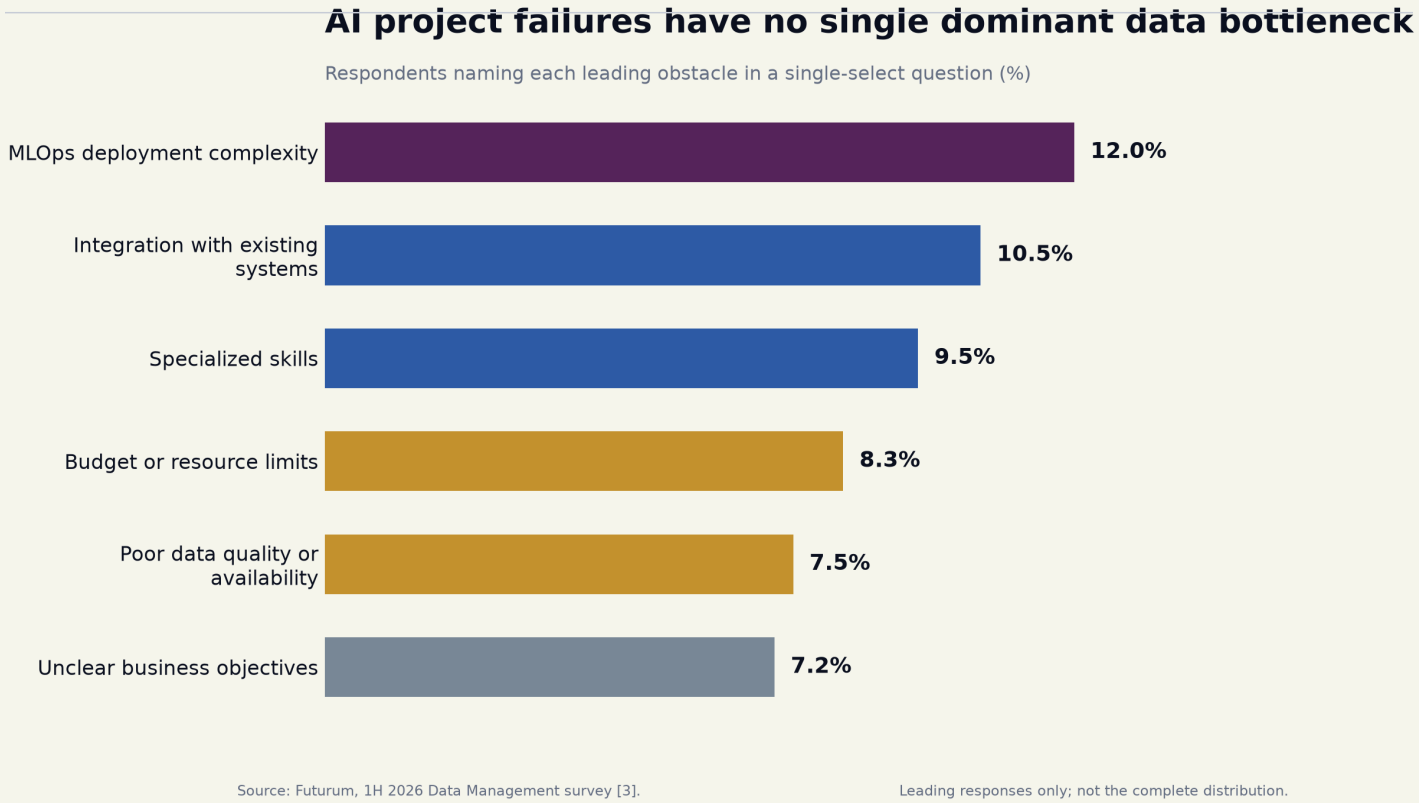


FIGURE 06 · AI PROJECT FAILURES HAVE NO SINGLE DOMINANT DATA BOTTLENECK

Futurum's 1H 2026 Data Management survey asked 818 respondents to identify the top data-specific factor contributing to AI project failure. The leading answer, MLOps deployment complexity, drew only 12.0%. Integration with existing systems and workflows followed at 10.5%, specialized skills at 9.5%, budget or resource limits at 8.3%, poor data quality or availability at 7.5%, and unclear business objectives at 7.2%.[3]

Poor data remains a serious issue, particularly when agents depend on current, authorized and well-described context. But in this single-select ranking, data quality was fifth, not first, and no answer came close to dominating. The stronger conclusion is that AI production is a distributed systems and operating-model problem.

The pattern also appears in Futurum's broader AI Platforms question about GenAI adoption challenges. Reliability or hallucinations was selected by 55.4%, privacy or security by 52.6%, business value or ROI by 43.3%, talent scarcity by 39.8%, and compute or infrastructure cost by 35.1%.[1] Because respondents could select multiple answers, these are not competing shares of one problem. They describe a production environment in which technical, economic and organizational constraints arrive together.

The dependencies reinforce one another. An agent can have excellent model performance and still fail because it cannot reach the required system. It can reach the system and still fail because its identity is overprivileged. It can use accurate data and still fail because the business outcome was not defined. It can complete the task and still fail economically because the workflow requires too much human review. It can be cheap and accurate most of the time and still be unacceptable because recovery from the remaining failures is slow or impossible.

Leaders should examine ten connected constraints:

- **Outcome definition:** Is the organization solving a valuable problem with a measurable baseline?
- **Workflow design:** Has the work been decomposed into bounded steps with explicit escalation points?
- **Data and context:** Are sources accurate, current, authorized, traceable and understandable to the system?
- **Integration:** Can the workflow interact with existing applications without brittle custom work?
- **Identity and permissions:** Does the agent have a distinct identity and only the authority required for the task?
- **Evaluation:** Are quality, safety and reliability thresholds defined before deployment?
- **Observability and recovery:** Can operators reconstruct what happened, stop it and recover cleanly?
- **Economics:** Is the cost of a completed and accepted outcome understood?
- **Skills and adoption:** Can employees and managers use the system inside the redesigned workflow?

- **Ownership:** Is one human accountable for the production result?

This list also changes the way leaders should think about platforms. A model gateway, vector database or agent framework can solve part of the problem. None solves the operating model. The enterprise needs a surrounding control plane or harness that binds the pieces together: identity, credentials, approved tools, data access, model routing, evaluation, policy, budgets, logging, human approval, fallback and rollback.

Connectivity standards can reduce integration effort, but connectivity is not accountability. Futurum's Data Management survey found that among a subpopulation of 416 respondents already interested in agentic AI, 20.0% said they were running Model Context Protocol in production, 26.9% were building pilots and 29.3% were strongly considering it.[3] That result should not be generalized to the full survey population. More importantly, a standard that helps an agent discover and use a tool does not decide whether that tool should be available, which credentials should be used or who owns the outcome.

PART III

The Accountability Problem

5. Agentic AI Lives in the Seams

> Connectivity is not accountability.

The fragmented structure of AI authority is visible in Futurum's survey. When respondents were asked who had primary authority for purchasing and implementing new AI software and infrastructure, the CTO was the largest category at 25.2%. The CAIO or AI lead followed at 13.9%, the CIO at 13.0%, the CEO at 11.8%, the CDO at 6.0% and the CISO at 5.5%, with several other roles accounting for the rest.[1] No role reached more than about one-quarter of respondents.

AI-GENERATED EDITORIAL ILLUSTRATION · ACCOUNTABILITY LIVES AT THE JUNCTIONS BETWEEN PLATFORMS, DATA, CONTROLS AND OUTCOMES.

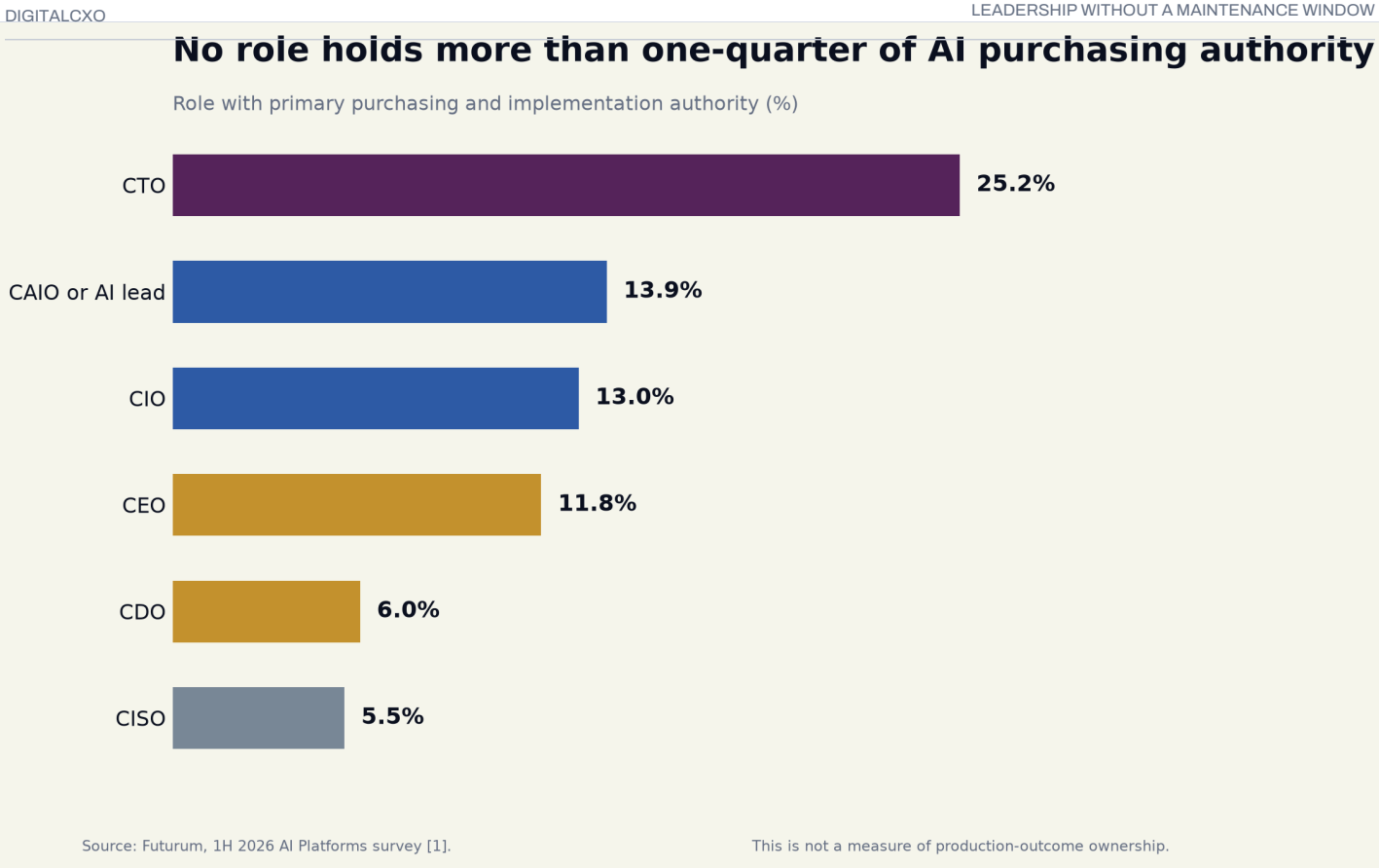


FIGURE 07 · NO ROLE HOLDS MORE THAN ONE-QUARTER OF AI PURCHASING AUTHORITY

That question measures purchasing and implementation authority, not production outcome ownership. It also does not prove that different executive roles hold different opinions about AI. What it does show is that the starting authority is distributed. Deloitte's finding that 71% of organizations have five or more technology leaders makes that distribution a structural reality rather than an edge case.[6]

The technology leadership team needs a shared operating model for the seams.

ROLE	PRIMARY CONTRIBUTION TO DURABLE PRODUCTION	THE SEAM THAT MUST BE RESOLVED
CIO	Service outcomes, enterprise applications, operating model and production accountability	Agents introduced outside central IT can alter services the CIO is still expected to run
CTO	Architecture, engineering platforms, agent runtime and technical portability	The platform may execute work whose data, permissions and business outcome belong elsewhere
CDO/CDAO	Data quality, context, metadata, lineage, retrieval and data-use policy	Governed data can be consumed or changed by agents owned by another function
CISO	Agent identity, least privilege, runtime authorization, security controls and incident response	Security may be accountable for agents it cannot fully discover or constrain
CAIO	AI portfolio, model strategy, evaluation standards and responsible-AI coordination	The AI portfolio may not control the platform, data, credentials or affected service
Digital or Transformation leader	Cross-functional workflow redesign and adoption where the role exists	Transformation goals may be separated from production reliability and ownership

The answer is not to put every decision under one executive. The work is too broad, and centralizing all of it can slow teams without resolving the underlying expertise. The answer is to separate shared execution from final accountability.

For every agentic production workflow, leaders should name at least five responsibilities:

1. **Outcome owner:** The person accountable for the business or service result. This owner approves the baseline, target, acceptance threshold and continuation decision.
2. **Platform owner:** The team responsible for the runtime, model access, tool interfaces, evaluation services, logging and recovery mechanisms.
3. **Data authority:** The person or function that approves the data and context the workflow can read, retain and change.
4. **Control authority:** The security, risk or compliance owner who defines identity, permissions, action policies and evidence requirements.
5. **Runtime operator:** The team that monitors the workflow, handles incidents and can pause or roll back execution.

One person may hold more than one responsibility in a smaller organization. What matters is that the responsibilities are explicit before the workflow becomes consequential.

The leadership team should also settle six decisions in advance:

- Who can approve a new agent for production?
- Who can grant or expand its access?
- Who can raise its level of autonomy?
- Who accepts the quality and safety thresholds?
- Who responds when the agent causes or contributes to an incident?
- Who has unquestioned authority to stop it?

These questions sound operational because they are. They also reveal why committee governance alone is insufficient. A committee can approve a policy or use case, but it will not be present for every tool call. The policy must become enforceable through the platform, and an accountable operator must remain able to intervene.

PART
IV

Keep, Change and Wait

6. Keep: The Disciplines That Make Production Possible

Agentic AI does not make the core disciplines of enterprise technology obsolete. It increases the cost of neglecting them.

Keep reliability engineering and service ownership. Agents can increase the volume and speed of change. DORA found that AI adoption was positively associated with delivery throughput but negatively associated with stability, especially where testing, feedback and architecture were weak.[7] Existing SRE and DevOps practices such as small batches, service-level objectives, automated testing, error budgets, rollback and blameless learning remain essential. They should be extended to agent behavior and workflow outcomes.

Keep least privilege and separation of duties. An agent should not inherit every permission of the person who invoked it. It should have its own managed identity, narrowly scoped authority and short-lived credentials wherever possible. NIST's 2026 agent identity concept paper focuses on identification, authorization, auditing and non-repudiation because conventional user-centric identity models do not fully address software agents acting across tools and applications.[12]

Keep data governance, lineage and provenance. Agents need context, but more context is not automatically better. Leaders need to know where retrieved information came from, when it was updated, which policy permits its use and whether the agent is allowed to write back. Data lineage becomes part of incident reconstruction as well as analytics governance.

Keep architecture discipline. Hybrid build-and-buy is already the dominant approach. Futurum found 51.0% of AI Platforms respondents using a hybrid model, compared with 24.9% favoring vendor or off-the-shelf technology, 20.1% working primarily in-house and 4.0% outsourcing to consultants.[1] A mixed estate makes modularity, standard interfaces, replaceability and dependency management more important.

Keep change and incident management. The processes may need to become faster and more automated, but agents still alter production state. A material change to a prompt, model, tool, permission, retrieval source or memory policy can change behavior without a traditional code release. Those changes need versioning, testing, evidence and recovery plans.

Keep a human owner. Human oversight does not require a person to approve every step. It requires someone to remain accountable for the design, thresholds, operation and result. Governance by exception still needs a named person who owns the exceptions.

Keep business-outcome measurement. Productivity and usage can be leading indicators. They cannot be the final score. Futurum found productivity was the most common AI success metric at 55.1%, followed by cost savings at 50.6%. Customer satisfaction, revenue, process automation, time to market and error reduction were used less often.[1] Leaders should keep measuring productivity, but connect it to an accepted workflow result.

Keeping these disciplines does not mean preserving every manual gate or legacy process. A weekly committee that reviews static documents may not match the speed of agentic execution. The principle should remain while the mechanism changes.

7. Change: The Operating Model Around AI

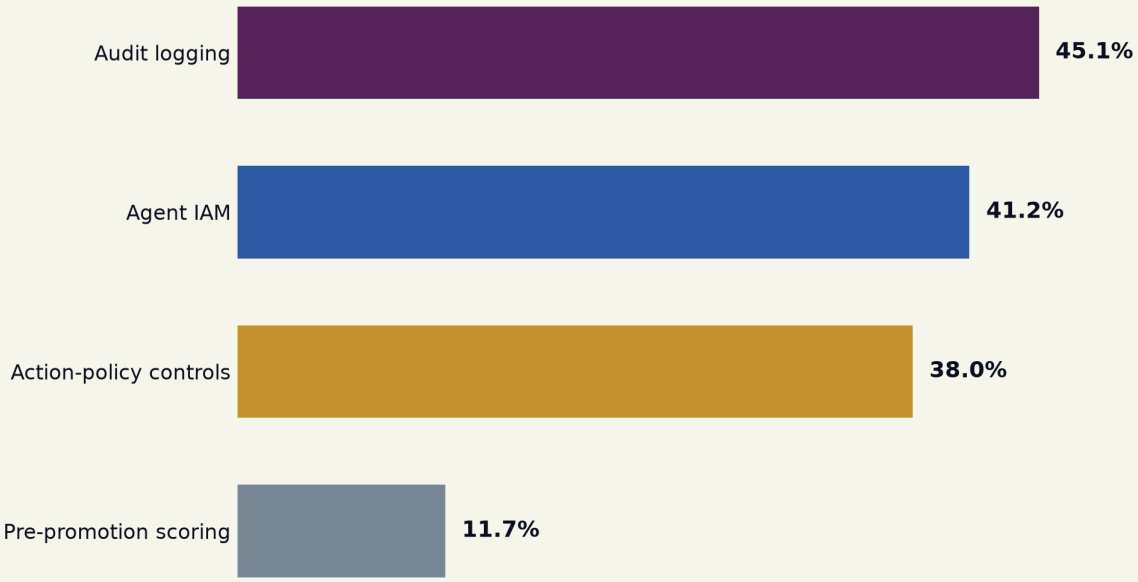
> **The platform must enforce what the governance program decides.**

The first change is to replace the pilot portfolio with a workflow portfolio. A pilot asks whether a technology can perform a task. A workflow portfolio asks which business outcomes deserve ongoing investment, what stage each workflow has reached and what must be true for it to advance. Every item should have a baseline, an owner, a next decision date and a defined path to scale, hold or stop.

AI-GENERATED EDITORIAL ILLUSTRATION · AUTONOMY SHOULD EXPAND THROUGH VISIBLE CHECKPOINTS, NOT DISAPPEAR INTO AN OPAQUE SYSTEM.

Fewer than half report even the most common runtime control

Organizations reporting each control (%)



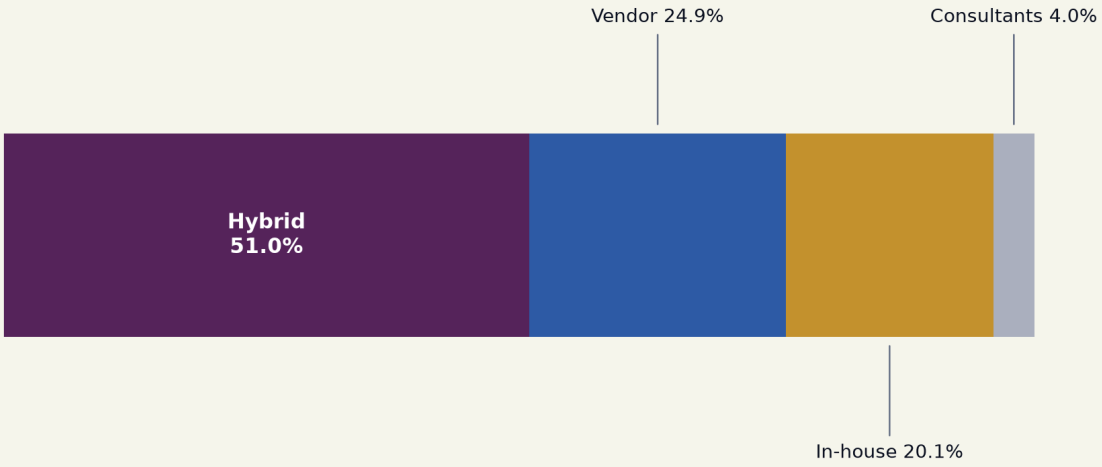
Source: Futurum, 2026 Software-Led Enterprise survey [2].

Control categories are reported independently.

FIGURE 08 · FEWER THAN HALF REPORT EVEN THE MOST COMMON RUNTIME CONTROL

Hybrid build-and-buy is the majority approach to enterprise AI

Primary approach to building AI capabilities (% of organizations)



Source: Futurum, 1H 2026 AI Platforms survey [1].

FIGURE 09 · HYBRID BUILD-AND-BUY IS THE MAJORITY APPROACH TO ENTERPRISE AI

The second change is to move governance into the runtime. NIST's AI Risk Management Framework organizes work around governing, mapping, measuring and managing risk across the lifecycle.[11] Agentic systems add an execution requirement. The platform must enforce what the governance program decides.

That enforcement layer should control:

- Which models and versions may be used
- Which tools are available to each workflow
- Which data sources can supply context
- Which agent identity invokes each action
- Which credentials are issued and for how long
- Which actions require approval
- Which quality and safety evaluations must pass
- Which cost, token, rate and time limits apply
- Which events are logged and retained
- Which fallback, pause and rollback mechanisms are available

Current control adoption remains uneven. Futurum's SLE respondents most often reported audit logging at 45.1%, agent IAM at 41.2% and action-policy controls at 38.0%. Per-agent budgets and human approval gates were each present in about one-third, while pre-promotion scoring was reported by 11.7%.[2] These findings do not show that controls are absent everywhere. They show that no common control has yet become universal practice.

OWASP's Top 10 for Agentic Applications makes the need concrete. Its risk categories include agent goal hijacking, tool misuse, identity and privilege abuse, supply-chain vulnerabilities, unexpected code execution, memory and context poisoning, insecure inter-agent communication, cascading failures, exploitation of human trust and rogue agents.[13] OWASP recommends least agency as well as least privilege. If a use case does not need autonomous behavior, adding it increases attack surface without creating value.

The third change is to extend identity programs to agents and other non-human actors. An agent using a person's broad session credentials creates an attribution gap. Leaders need to know which agent acted, on whose behalf, under which policy, with which temporary permissions and in response to which approved goal. Identity should be bound to the workflow and action, not merely to the application hosting the model.

The World Economic Forum's 2026 cybersecurity outlook warns that multiplying agent identities and connections makes credentials, permissions and interactions at least as important as human identity management. It recommends continuous verification, audit trails and accountability grounded in zero-trust principles.[14] This does not require waiting for one final agent-identity standard. Organizations can start with distinct service identities, short-lived tokens, explicit delegation, action-level authorization and immutable logs.

The fourth change is measurement. Leaders need to move from cost per token or task to cost per accepted outcome. McKinsey's 2026 analysis of agent economics argues that the meaningful unit is the fully loaded cost to finish a job, combining humans, agents and deterministic systems.[18] An agent that appears cheap may be expensive after review, exception handling and rework. An agent that uses a more expensive model may still produce better economics if it finishes more work correctly.

A completed-work scorecard should include:

- Accepted outcomes divided by attempted workflows
- End-to-end cycle time
- Human review and exception-handling time
- Rework and correction rate
- Reliability and incident rate
- Cost of models, infrastructure and connected services
- Cost of human participation
- Cost of failures and recovery
- Value created or cost avoided

FinOps needs to become part of the agent operating model. The FinOps Foundation's 2026 survey of 1,192 respondents representing more than $83 billion in annual cloud spend found that 98% of FinOps practitioners were managing AI spend, up from 31% two years earlier.[15] IBM found that 84% of surveyed technology leaders had not fully operationalized AI financial management and 85% lacked full real-time visibility into AI spending.[5] The mandate has reached the FinOps team faster than the operating capability has matured.

The fifth change is architecture. The enterprise should assume models, frameworks, tools and standards will continue to change. Portability does not mean abstracting every feature to the lowest common denominator. It means isolating high-change dependencies, maintaining evaluation suites that allow alternatives to be compared, and keeping state, policy and business logic from becoming unnecessarily trapped inside one vendor's runtime.

The sixth change is funding. A proof of concept can be funded as a project. Durable production needs product-style lifecycle funding for evaluation, monitoring, model changes, data maintenance, security testing, incident response and retirement. Leaders should stop approving experiments whose sponsors have no plan to fund the surrounding production work.

The seventh change is decision cadence. An experiment should not survive indefinitely because no one wants to declare it unsuccessful. Each workflow needs continuation criteria, a decision

date and four possible outcomes: scale, continue within bounds, hold pending a named condition, or stop. Stopping is evidence of portfolio discipline, not a failure of innovation.

8. Wait: Where Caution Creates Options

Waiting is an active decision.

Waiting is appropriate when the authority being granted is greater than the organization's ability to observe, constrain or recover from it. It should be an active decision with tests and exit conditions, not a substitute for learning.

WAIT ON	WHAT TO TEST NOW	EXIT CONDITION
Broad unsupervised autonomy	Bounded workflows with approval checkpoints and exception logging	Stable performance across a defined period, low exception rates, tested recovery and an accountable owner
Irreversible write access to systems of record	Read-only access, sandbox writes, staged changes and dry-run diffs	Action-level authorization, rollback or compensating controls, and demonstrated accuracy above the agreed threshold
Multi-agent orchestration in high-consequence work	Small numbers of specialized agents in a constrained environment	Traceable inter-agent messages, bounded delegation, failure containment and reproducible evaluation
Autonomous production or security changes	Recommendations and pre-approved low-risk actions	Reliable policy enforcement, blast-radius limits, real-time observability and successful incident exercises
Premature dependence on one interoperability standard	Adapters and controlled pilots	Sufficient ecosystem stability, security controls and a clear migration path
Architecture locked to one model or framework	Use the best current capability behind isolated interfaces	Evidence that lock-in benefits exceed switching risk, with a funded exit plan

Gartner forecasts that 40% of enterprises will demote or decommission autonomous agents by 2027 because governance gaps are discovered after production incidents.[17] The forecast may not materialize exactly as stated, but the underlying distinction is useful: governance should reflect an agent's level of autonomy and access. Treating every agent as either fully trusted or completely prohibited creates the wrong choice.

Leaders should make autonomy conditional. A workflow earns more authority when it demonstrates stable outcomes, bounded behavior, economic value and recoverability. It loses authority when conditions change or performance drifts. The autonomy level is an operating setting, not a permanent label attached to the product.

PART V

From Experiment to Durable Production

9. The Durable Production Gate

The most effective way to end endless experimentation is to define what must be finished. A durable production gate turns that definition into an executive decision framework. It is not one review meeting at the end of development. Teams should use it from the beginning so that missing production work becomes visible before the pilot generates expectations it cannot meet.

Gate 1: Measurable outcome

The workflow needs a baseline, target and owner. "Improve productivity" is not enough. A better objective might be to reduce mean time to resolve a defined class of incidents from 90 minutes to 45 minutes without increasing recurrence or security exceptions. The baseline should include the current human process, because the existing workflow is the relevant competitor.

Key questions:

- What completed result has value?
- How is that result accepted or rejected?
- What is today's cycle time, cost, quality and failure rate?
- Who owns the outcome?

Gate 2: Bounded workflow

The team must define what the agent may do, may not do and must escalate. Open-ended goals encourage unpredictable paths. Durable workflows decompose work into bounded stages and place explicit limits around tools, time, cost and state changes.

Key questions:

- What starts and ends the workflow?
- Which actions are read-only, reversible or irreversible?
- Where is human judgment required?
- What conditions force a stop or escalation?

Gate 3: Production-grade data and context

The agent needs the right context, not every available document. Sources should be authorized, current, traceable and appropriate for the task. Retrieval quality should be evaluated separately from model quality. Write-back should have its own policy and evidence.

Key questions:

- Which sources are authoritative?
- How are freshness, lineage and access enforced?
- Can retrieved content contain untrusted instructions?
- What information may be retained in memory and for how long?

Gate 4: Identity and authority

Every production agent should have a distinct identity and an explicit delegation path. Permissions should be scoped to the workflow and issued for the shortest practical period. High-impact actions should require additional authorization.

Key questions:

- Can every action be attributed to a specific agent and initiating user or service?
- Does the agent have more access than the task requires?
- Are credentials short-lived and revocable?
- Can delegation to another agent expand authority without a new decision?

Gate 5: Evaluation and acceptance

Teams need repeatable evaluation before and after release. The evaluation set should include normal work, difficult edge cases, adversarial inputs and known failure patterns. Thresholds must be agreed before stakeholders see a favorable demonstration.

Key questions:

- What quality, safety and reliability thresholds must be met?
- Which failures are tolerable, and which are automatic blockers?
- How will model, prompt, tool and data changes be retested?
- Does evaluation measure the completed workflow or only the model response?

Gate 6: Observability and recovery

Operators must be able to reconstruct the chain from goal to context, model decision, tool call, system change and final outcome. Logs without correlation are not enough. The workflow also needs a tested way to pause, fall back and recover.

Key questions:

- Can operators see which goal, model, data and tools produced an action?
- Are policy decisions and human approvals recorded?
- Can execution be stopped in real time?
- Has rollback or compensating recovery been tested?

Gate 7: Completed-work economics

The financial model should cover the whole workflow. Include model and infrastructure consumption, platform services, data movement, human review, exceptions, failed runs, integration maintenance and recovery. Compare the cost with the value of accepted outcomes, not the number of calls or tasks attempted.

Key questions:

- What is the fully loaded cost per accepted outcome?
- How does cost change with volume, complexity and model choice?
- What portion of cost comes from human review and exceptions?
- Which thresholds trigger a cheaper model, human handoff or stop?

Gate 8: Ownership and retirement

Every workflow needs one accountable owner, an operating team and a retirement plan. Models and dependencies will change. A workflow that no longer meets its thresholds should not remain in production because ownership has become diffuse.

Key questions:

- Who decides whether to scale, hold or stop?
- Who responds to incidents and degraded performance?
- When is the next formal continuation decision?
- How can data, credentials, integrations and retained memory be retired safely?

The gate produces four decisions:

1. **Scale:** The workflow meets its thresholds and can receive more volume, users or authority.
2. **Continue within bounds:** It creates value at the present scope but has not earned broader authority.
3. **Hold:** A named condition, owner and deadline must be resolved before further expansion.
4. **Stop:** The workflow lacks sufficient value, control or feasibility, and its resources should be redirected.

There is no permanent pilot category. A pilot without a next decision is simply unfinished work.

10. Empower Teams Without Drowning Them

> Give teams a paved road, not an open field and not a locked gate.

Centralizing every AI decision will create a queue. Allowing every team to assemble its own stack will create fragmentation, hidden risk and repeated production work. Technology leaders need a paved road with controlled exits.

The paved road should give teams self-service access to approved models, agent identities, tool connectors, data and retrieval services, evaluation suites, policy templates, logs, traces, approval patterns, budget controls and recovery mechanisms. Teams remain responsible for their outcome and workflow. The platform removes the need to rebuild common controls.

DORA's AI Capabilities Model identifies quality internal platforms as one of seven capabilities that amplify AI's positive effects. It describes the platform as a source of automated, secure paved roads that keep individual productivity gains from becoming downstream bottlenecks.[7] That is the right ambition for an agentic platform: accelerate responsible delivery by making the safe path easier than the improvised path.

Controlled exits matter because not every use case will fit the standard path. A team may need a different model, an unusual data source or a specialized tool. The exception process should be explicit, fast and evidence-based. The team explains the reason, additional risk, compensating controls, owner and expiration date. Exceptions should not become permanent architecture through neglect.

Managers are also part of the platform, in the organizational sense. Gallup found that manager support and integration with existing work systems were the two strongest drivers of frequent AI use in its Q1 2026 U.S. workforce data.[10] Teams need leaders who can translate enterprise policy into the work itself: where AI should be used, where human judgment remains essential, what good performance looks like and how employees can raise concerns without being treated as obstacles.

Empowerment therefore has two dimensions. Technical empowerment gives teams reusable capabilities and clear boundaries. Managerial empowerment gives them permission, coaching and outcome clarity. One without the other produces either unused infrastructure or unmanaged experimentation.

PART VI

What Leaders Should Do Monday Morning

11. A 90-Day Leadership Agenda

The next 90 days should not be another enterprise AI strategy exercise. The goal is to prove that the technology leadership team can see the current estate, make shared decisions and move a small number of workflows into durable production.

Days 1 through 30: Establish reality

Create one inventory of active AI experiments, deployed agents and production workflows. Do not limit it to projects funded or approved by central IT. Ask business units, engineering teams, data teams, security teams and major SaaS owners what is already operating.

For each item, record:

- The business outcome and current stage
- The human outcome owner
- The models, tools, data and systems involved
- The agent identity and permissions, if they exist
- Whether the workflow reads, recommends or writes
- Current volume, cost, review burden and known incidents
- The next scale, hold or stop decision date

Classify every item as experiment, deployment, production or durable production. The classification will be uncomfortable, which is useful. It separates a live demo from an owned service.

Select a small number of workflows for focused work. Choose valuable but bounded processes with measurable baselines. Include at least one workflow that crosses two technology leadership domains so the team is forced to settle a seam rather than avoiding it.

Days 31 through 60: Build the paved road

Adopt the eight-gate durable production framework and apply it to the selected workflows. Do not wait for a perfect enterprise standard. Establish a minimum viable set of shared services:

- Managed agent identity and short-lived credentials
- Approved model access and basic routing
- Tool registration and action policies
- Evaluation and release evidence
- Correlated logs and traces
- Human approval and escalation patterns
- Budget, rate and time limits
- Pause, fallback and rollback mechanisms

Clarify decision rights across the CIO, CTO, CDO/CDAO, CISO and CAIO organizations. Name the outcome owner, platform owner, data authority, control authority and runtime operator for each selected workflow.

Define the first completed-work scorecard. It should show value, cycle time, accepted outcomes, human effort, exceptions, reliability and fully loaded cost. Avoid a dashboard dominated by prompts, tokens or generated artifacts.

Days 61 through 90: Prove the organization can finish

Move the selected workflows through all eight gates. Run failure exercises that include an unavailable model, poisoned context, an over-budget loop, a rejected action, expired credentials and a rollback. If the team cannot recover during a controlled exercise, the workflow has not earned broader autonomy.

Publish the results to the technology leadership team and business sponsors. For each workflow, make one of four decisions: scale, continue within bounds, hold pending a named condition, or stop.

Stopping at least one weak initiative may be a healthy result. It demonstrates that the portfolio has standards and frees capacity for work that can be finished. Scaling one well-owned workflow through reusable controls is more valuable than announcing another dozen proofs of concept.

At day 90, the organization should have more than a report. It should have an inventory, a common production gate, explicit decision rights, a working paved road, completed-work economics and evidence that it can recover from failure. Those are the beginnings of an operating model.

Conclusion: The Leaders Who Finish

> **The organizations that lead will become better at ending experimentation.**

Agentic AI will continue to improve. Models will become more capable, tools will become easier to connect and vendors will keep pushing autonomy deeper into enterprise products. None of that removes the leadership work. Greater capability makes the questions of authority, accountability and production discipline more important.

Technology leaders should keep the practices that make systems dependable, change the mechanisms that cannot operate at agent speed and wait before granting authority that the organization cannot observe or recover. They should give teams a paved road, not an open field and not a locked gate.

The central measure is completed, accepted work. An agent that produces more output but leaves more review, risk or recovery for others has not necessarily improved the enterprise. A workflow that finishes reliably, learns from exceptions and remains economical has.

The organizations that lead this transition will not eliminate experimentation. They will become better at ending it. They will know which workflows to scale, which to contain and which to stop. Their advantage will come from finishing more valuable work safely, then using what they learned to finish the next piece faster.

Sources and Research Notes

1. Futurum, *AI Platforms Decision Maker, 1H 2026*, n=820. Proprietary research supplied for this report. Selected subquestions used different bases, which are identified in the text. Publication permission required.

2. Futurum, *Software Lifecycle Engineering Decision Maker, 2H 2026*, n=839, fielded June 2026. Proprietary research supplied for this report. Publication permission required. The reported 75.2% confirmed AI-contributed incident figure requires source-report verification before publication.

3. Futurum, *Data Management Decision Maker, 1H 2026*, n=818. MCP question base n=416. Proprietary research supplied for this report. Publication permission required.

4. McKinsey & Company, *The State of AI in 2026: On the Road to ROI*, August 25, 2026. Online survey of 1,719 participants in 97 countries, fielded May 4 through June 8, 2026. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

5. IBM Institute for Business Value and Oxford Economics, *2026 Tech Leader Study: Building the IT Foundation for Agentic AI at Scale*, June 2026. Survey of 2,000 senior technology executives in 33 geographies and 19 industries, fielded January through April 2026. https://newsroom.ibm.com/2026-06-08-new-ibm-study-finds-cios-and-ctos-face-growing-ai-control-gap-as-enterprise-deployment-scales

6. Deloitte, *2026 Global Technology Leadership Study* and *Rewiring the Enterprise Operating Model for AI Scale*, April and June 2026. Survey of 662 senior technology leaders, 87% of them C-suite, fielded December 22, 2025 through February 23, 2026. https://www.deloitte.com/us/en/insights/topics/technology-management/rewiring-ai-operating-model.html

7. DORA and Google Cloud, *2025 State of AI-Assisted Software Development*, September 2025. Nearly 5,000 technology professionals plus more than 100 hours of qualitative research. https://dora.dev/research/2025/dora-report/

8. Google Cloud and DORA, *How to Unlock True ROI in Software Development: A Deep Dive Into the Latest DORA Research*, June 9, 2026. https://cloud.google.com/blog/products/ai-machine-learning/how-to-measure-the-business-value-of-generative-ai/

9. Microsoft, *Microsoft 365 Copilot, Human Agency, and the Opportunity for Every Organization*, May 5, 2026, reporting findings from the 2026 Work Trend Index. Survey of 20,000 AI users in 10 countries plus anonymized Microsoft 365 signals. https://www.microsoft.com/en-us/microsoft-365/blog/2026/05/05/microsoft-365-copilot-human-agency-and-the-opportunity-for-every-organization/

10. Gallup, *State of the Global Workplace: 2026 Report*, including Q1 2026 U.S. workforce analysis of AI integration and manager support. https://www.gallup.com/workplace/349484/state-of-the-global-workplace.aspx

11. National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, NIST AI 600-1, July 26, 2024. https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence

12. NIST National Cybersecurity Center of Excellence, *Accelerating the Adoption of Software and Artificial Intelligence Agent Identity and Authorization*, initial public draft, February 5, 2026. https://csrc.nist.gov/pubs/other/2026/02/05/accelerating-the-adoption-of-software-and-ai-agent/ipd

13. OWASP GenAI Security Project, *OWASP Top 10 for Agentic Applications 2026*, December 2025. Developed with more than 100 industry experts, researchers and practitioners. https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/

14. World Economic Forum in collaboration with Accenture, *Global Cybersecurity Outlook 2026*, January 12, 2026. https://www.weforum.org/publications/global-cybersecurity-outlook-2026/

15. FinOps Foundation, *State of FinOps 2026*, sixth annual survey, 1,192 respondents representing more than $83 billion in annual cloud spend. https://data.finops.org/

16. Gartner, *Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027*, June 25, 2025. Forecast, not an observed failure rate. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027

17. Gartner, *Gartner Says Applying Uniform Governance Across AI Agents Will Lead to Enterprise AI Agent Failure*, May 26, 2026. Forecast that 40% of enterprises will demote or decommission autonomous agents by 2027. https://www.gartner.com/en/newsroom/press-releases/2026-05-26-gartner-says-applying-uniform-governance-across-ai-agents-will-lead-to-enterprise-ai-agent-failure

18. McKinsey & Company, *Where AI Agents Pay Off: A Practical Guide to the Economics of Agentic Workflows*, August 24, 2026. https://www.mckinsey.com/capabilities/quantumblack/our-insights/where-ai-agents-pay-off-a-practical-guide-to-the-economics-of-agentic-workflows

Digital CxO

POWERED BY Techstrong | Group
a Futurum company

The leaders who finish will turn experimentation into an operating capability.

Not because they moved fastest at the beginning, but because they knew what to keep, what to change, what to wait on and when to stop.