Open Future Forum

OPEN FUTURE FORUM

AI Leaders AI Leverage Report

NOVEMBER 2026

November 2026 | Edition 2 | Research report | Murray Newlands | Open Future Forum

Prepared 2 November 2026. Evidence cutoff: 31 October 2026, 23:59 America/Los_Angeles. Direct operator measures are carried forward and reconciled to October. This edition develops evidence requirements for changing the scope of a production workflow.

About this publication: Open Future Forum publishes research for executive and investor audiences. Murray Newlands is its founder. This report separates OFF survey observations, external research, author analysis and illustrative scenarios. Vendor-authored guidance is identified as such and does not constitute an independent product assessment. The collection and supporting methods are available through the OFF research index: https://openfutureforum.com/research/

4,326 substantive body words | 5,662 total words
Tables, FAQs, methods and references are excluded from the body count.

In this report

01 The decision in brief
02 Executive summary
03 What November adds
04 The source before the findings
05 1. Define the service before counting agents
06 2. Ask which evidence a change invalidates
07 3. Account for unsuccessful work in the economics
08 4. Separate platform guarantees from application work
09 5. Keep authority tied to the actual action
10 6. Plan service recovery before widening the commitment
11 External guidance: testing the system that will actually run
12 Practical application: a change that deserves a new gate
13 The next production review
14 Reader questions
15 Citable facts
16 Methods and limitations
17 References
18 Related research and citation
19 Editable working tool

The decision in brief

Expand a production agent only when the evidence still applies to the proposed users, workload, data and actions. OFF's operator sample already contains substantial production use, but it does not verify reliability or economics. November's contribution is a change-sensitive release record that separates platform capabilities from application responsibilities and makes unresolved operating limits visible before expansion.

Executive summary

The direct operator instrument has small but useful question-level bases. Seventy-five respondents answered the fleet-size question: nine reported no production agents, 27 reported one to five, 11 reported six to twenty, 22 more than twenty and six uncertainty. Sixty of 75 therefore reported at least one production agent. These are self-reported counts under the instrument's wording, not an independently verified inventory of autonomous systems. [FLEET]

Cost visibility has a different base. Of 77 respondents, 34 answered yes to real-time visibility into what AI costs to run, 30 partly and 13 no. The question does not establish complete cost attribution to workflows. In the genuinely matched group of 60 respondents reporting at least one production agent, 30 answered yes, 23 partly and seven no. This matched result was already part of October analysis; it is retained, not presented as a new discovery. [COST; OPERATOR_COST]

Completion is a central limitation. The source export contains 314 eligible non-invited application rows, while individual operator question bases range from 71 to 101. People answering a detailed production question may differ from those who leave it blank. The report therefore keeps each denominator visible and does not generalize the answers to all applicants or all enterprises.

The operational issue is what happens after an initially successful release. A new user population, longer task, additional tool or broader data source can change reliability, cost and authority without changing the product name. The AI leader needs to know which part of the original evidence remains applicable and which part must be renewed. A deployment record that lists only a model and an owner cannot answer that question.

This edition supplies a production control register, a cross-functional expansion gate and an illustrative review of a customer-service agent. Repeated testing can slow iteration when teams cannot reuse relevant evidence. The recommended response is to tie review to the changed boundary and consequence, preserve reusable evidence and keep routine changes delegated. The aim is to make the next release decision faster and better informed, while preventing an old test result from becoming a permanent assurance for a materially different system.

What November adds

The November export reproduces the October operator distributions. There are no new comparable answers for maturity, fleet size, running-cost visibility, bottlenecks, access methods or planned spending. The edition retains those observations with their dates and question bases. It does not describe a change in deployment, spending or readiness simply because the reports are published in a new month.

The first addition is a classification of expansion changes. Increasing volume is different from adding a new user group, granting another action or connecting a new data source. Each can invalidate a different part of the evidence. The report turns that distinction into a release review that asks what changed, which assumption is affected and who must accept the resulting uncertainty.

The second addition is an explicit separation of platform and application responsibility. A platform may provide identity, logging and model access without defining the application's accepted business result or guaranteeing that every downstream action is correctly authorized. The application owner must know which guarantees are supplied, which depend on configuration and which remain their responsibility. This reduces the risk of a control being assumed by both teams and operated by neither.

The third addition is a failure-accounting view of production economics. The report connects retries, escalations, rework and rejected outcomes to the unit of useful service. It does not repeat October's workflow-cost-accounting method or claim the matched cost result as new. Instead, it asks how the economics change when the workflow attempts harder work or when the operating boundary expands.

The production register is also distinct from the personal Enterprise Data Agent Governance resource maintained by Murray Newlands. That resource provides synthetic answer, clarification and refusal cases for data agents. This report addresses the decision to release or expand a live workflow, including service capacity, evidence freshness, economics and ownership. It does not reuse the synthetic benchmark as a new OFF survey or claim a vendor has passed it.

The source before the findings

The Enterprise AI at Microsoft export contains 314 eligible non-invited application rows. The self-described seat question has 101 answers, maturity 91, planned spending 76, running-cost visibility 77, fleet size 75, bottlenecks 75 and access method 71. These are different completion bases. The report does not replace them with the total roster or treat every response as a separate company.

Eligibility and deduplication are applied separately to each question. Invited-only rows are excluded. The latest eligible nonblank answer is selected using a normalized respondent identifier and registration timestamp within the cutoff, with a stable tie rule. Registration time is not a verified answer-edit time. The instrument is selective, self-reported and cross-sectional; no probability-sample margin of error is claimed.

Production agent is the respondent's count under the question as asked. The research did not inspect uptime, autonomy, business value or control effectiveness. Similarly, a yes answer to cost visibility is not an audit of cost completeness. Planned spending covers the next twelve months and is not actual expenditure, committed procurement or a market-size estimate.

The matched fleet/cost result uses respondents who answered both questions and reported at least one agent. The applicable access base excludes not applicable yet, with the full base also shown. Small categories remain visible as counts without unsupported subgroup conclusions. These boundaries matter because the recommendations are intended to generate stronger deployment-specific evidence than the survey itself can provide.

1. Define the service before counting agents

An agent count is useful inventory context, but it is a weak unit for judging production capability. One agent may support a consequential workflow with several tools; several agents may divide a narrow internal task. The OFF fleet distribution does not standardize architecture or verify how respondents counted their systems. The result should prompt an inventory review, not a league table in which a larger fleet implies a more mature organization. [FLEET]

The AI leader should define the service the business expects. What request enters the workflow, what outcome is accepted and what happens when the system cannot complete it? The definition should include the relevant user population, data, actions and operating period. A service boundary makes it possible to evaluate reliability and economics without assuming that every component called an agent is comparable.

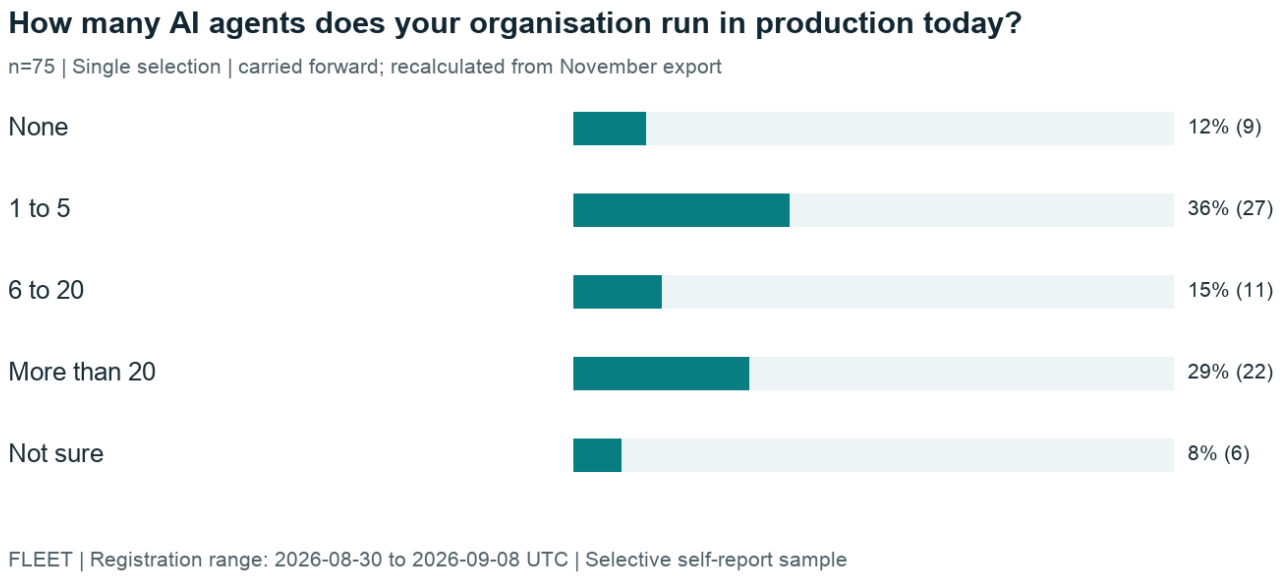
The accepted outcome should be observable outside the system's own claim of success. If the workflow prepares a customer response, acceptance may require an accurate answer and a recorded disposition. If it changes a record, the team should verify the resulting state and authorization. A plausible final message can be useful interface behavior without proving that the intended action occurred correctly.

Include exclusions in the service definition. A system may intentionally route complex or sensitive cases to a person. That can be an effective design, provided the excluded work, handoff and human capacity are part of the service model. It becomes misleading when the automation rate omits difficult cases while the business case assumes that their cost has disappeared.

The definition also helps separate a model change from a service change. A new model may improve one component while leaving the service boundary intact. A new tool can change the available actions even if the model remains the same. A new user group can alter the task distribution without any code change. The release record should identify the change in the service, not just the name of the technical component updated.

The practical outcome is a stable description that the business owner, platform team and application team can all use. It should be short enough to maintain and specific enough to test. Once that boundary exists, fleet counts can help locate owners and dependencies, while quality, cost and authority are assessed at the level of the work the organization actually relies on.

Exhibit 1. Reported production fleet size



Exact selection	Count	Base	Share
None	9	75	12%
1 to 5	27	75	36%
6 to 20	11	75	15%
More than 20	22	75	29%
Not sure	6	75	8%

Exhibit note: FLEET; carried-forward single-selection operator instrument. At least one agent: 60/75, 80%. Agent counts are respondent-defined and do not establish reliability, value or autonomy.

2. Ask which evidence a change invalidates

Production evidence has a scope. A test result applies to a particular system, task set, data condition and operating design. It may remain useful after a small change, but it should not be assumed to cover every later expansion. The AI leader needs a repeatable way to identify what changed and which part of the assurance must be renewed.

Separate changes in volume, population, information and authority. Higher volume can expose capacity, latency and cost limits. A new user group can introduce different tasks or expectations. A new data source can change quality, access and interpretation. A new action can change consequence and reversibility. These are recommended review categories, not survey variables. Their value is that they direct the team toward the evidence most likely to become stale.

The evaluation should distinguish routine tasks from consequential failures. An average success rate can improve while a rare but important action becomes less reliable. The business owner should identify which failures are tolerable, which require human intervention and which prevent release. The threshold should reflect the service and its consequence, rather than a universal percentage chosen because it looks rigorous.

Repeated trials can help reveal variability, but the team should report what the repetition means. Several attempts on the same task do not create several independent business cases. A system that eventually succeeds after repeated retries may be acceptable for one use and unsuitable for a time-sensitive service. The evaluation should preserve first-attempt behavior, retries and final accepted outcome where those distinctions affect the user or cost.

The test set itself requires maintenance. If the workflow expands to a new customer population, an unchanged test set may no longer represent the work. Add relevant cases while retaining a stable subset for regression comparisons. Record why cases are included, how they are graded and whether a qualified reviewer agrees with ambiguous judgments. A larger test set is not automatically a better one if its coverage is poorly matched to the decision.

The release recommendation should state which evidence remains valid, which has been refreshed and which uncertainty remains. That is more useful than a binary statement that all tests passed without a defined scope. The leader can then authorize a bounded rollout, require additional evidence or narrow the proposed change. This makes evaluation part of the operating decision rather than an isolated engineering score.

3. Account for unsuccessful work in the economics

The cost-visibility result should be read narrowly. Thirty-four of 77 respondents answered yes to real-time visibility into what AI costs to run; 30 answered partly and 13 no. The survey does not establish whether those respondents include every cost category or can allocate cost to a particular workflow. The AI leader should therefore distinguish a metering capability from a complete economic view. [COST]

The matched result improves the comparison but does not remove that limit. Among 60 respondents who reported at least one production agent and answered the cost question, 30 answered yes, 23 partly and seven no. This is a valid matched base, unlike subtracting the overall production share from the separate overall cost-visibility share. It remains a self-report about visibility rather than a cost audit. [OPERATOR_COST_YES; OPERATOR_COST_PARTLY; OPERATOR_COST_NO]

For an actual service, measure unsuccessful work as part of cost-to-serve. Retries consume resources, escalations consume human time and rejected outputs may require rework. A low cost per model call can coexist with a high cost per accepted outcome if the workflow makes many calls or frequently fails. The economic denominator should follow the service definition established before the release decision.

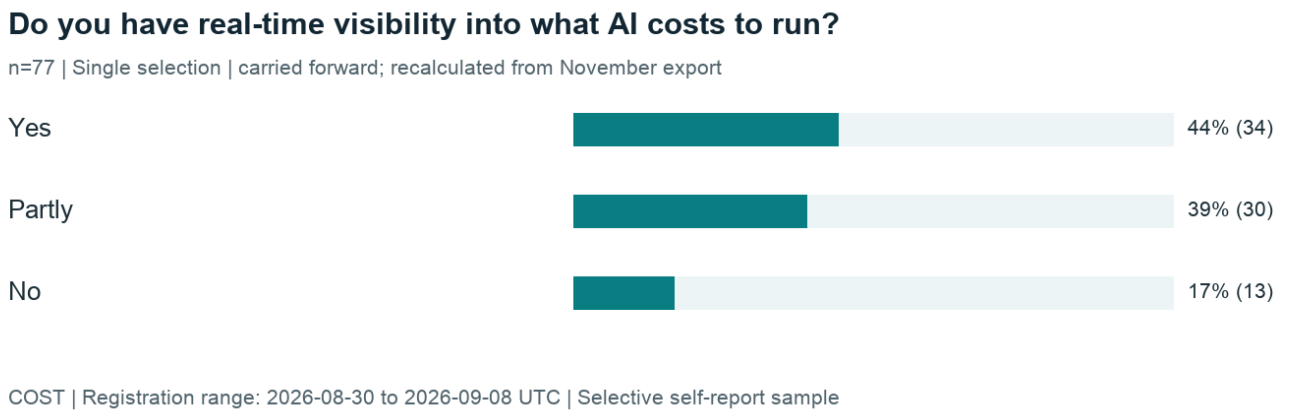
Expansion can change that denominator. Early users may bring simpler tasks, while later users require longer context, more tools or more specialist review. A larger workload can also change caching, concurrency or support requirements. The team should compare the proposed task mix with the tested mix

and identify which cost assumptions are sensitive to the difference. An attractive pilot average should not be applied to every future case without examination.

Finance and engineering need complementary views. Engineering should explain technical consumption and its drivers. Finance should reconcile supplier charges, shared costs and any claimed benefit. The business owner should confirm accepted outcomes and the use of released capacity. These records can share a workflow identifier without forcing each function to maintain the same level of detail.

The release decision should include the action available when cost exceeds the intended boundary. A team may cap attempts, route difficult work to a person, narrow a task or pause expansion. Each action can affect service quality, so it should be evaluated with the business owner. A spending alert without an effective operating response is incomplete cost control, even if the dashboard updates in real time.

Exhibit 2. Real-time AI running-cost visibility



Response	Full question count / base	Share	Matched production operators	Share
Yes	34 / 77	44%	30 / 60	50%
Partly	30 / 77	39%	23 / 60	38%
No	13 / 77	17%	7 / 60	12%

Exhibit note: COST and OPERATOR_COST; carried forward. Matched respondents answered fleet and cost questions and reported at least one production agent. A yes response does not establish complete workflow attribution. The matched analysis already appeared in October.

4. Separate platform guarantees from application work

The leading reported bottleneck is data access and quality, selected by 29 of 75 respondents. Compute, integration and inference cost also appear frequently. The instrument does not diagnose the root cause inside any respondent's organization, but it warns against assuming that model capability alone determines production progress. The application depends on an operating environment that several teams may own. [BOTTLENECK]

A platform responsibility should be stated as a capability with a boundary. It might provide an authenticated connection, a logging mechanism, an execution environment or a spending control. The application owner needs to know how that capability is configured, what it covers and what it does not. A platform's ability to collect logs does not prove that the application records the business context needed for investigation.

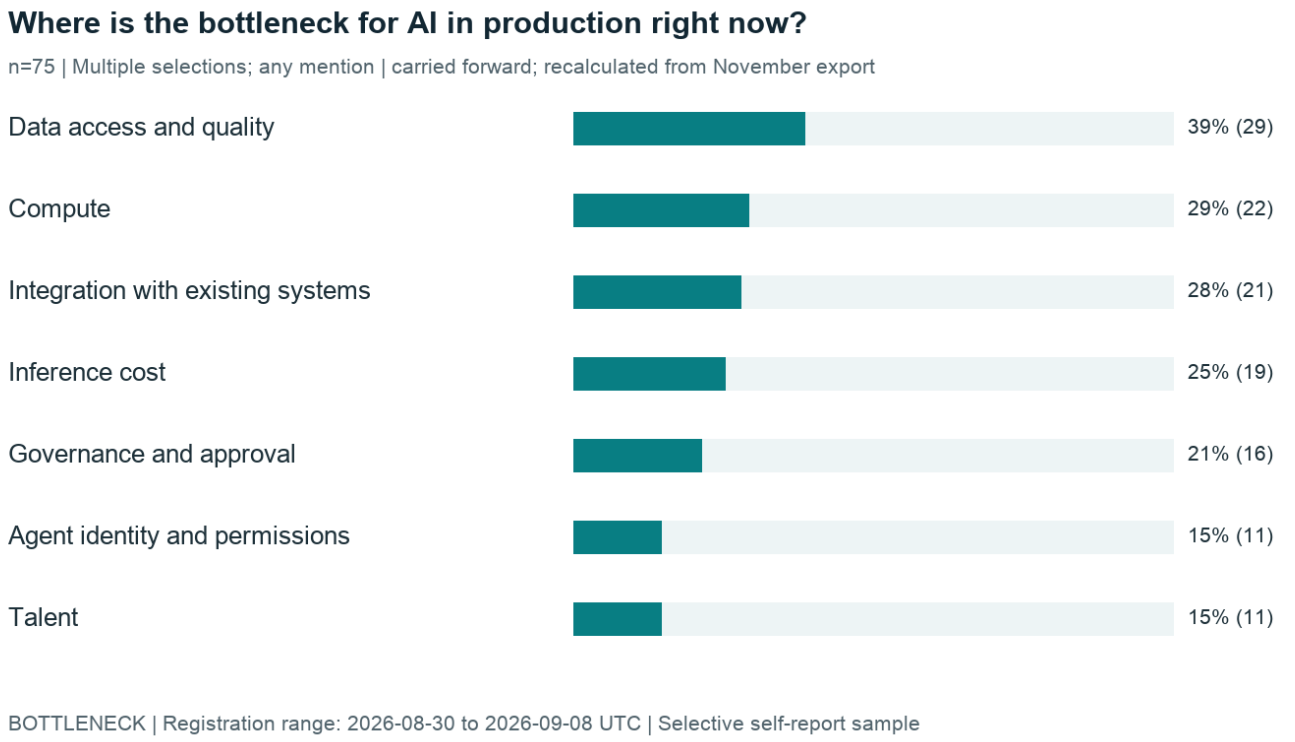
Data responsibility is similarly divided. A connector can make information available without establishing that the information is current, complete or interpreted correctly. The source owner should maintain the relevant meaning and quality expectations. The application team should handle missing or uncertain data appropriately. The business owner should decide which uncertainty is acceptable for the intended use. Leaving all three responsibilities under a label such as data readiness makes the problem difficult to resolve.

The handoff should include change notification. If a source schema, permission model, data definition or service limit changes, the application may need retesting even when its own code is untouched. The platform and source teams should identify which changes they will communicate and how the application owner can determine whether the release evidence remains valid. The arrangement should be operationally feasible, not a promise to notify every team about every minor update.

Where several applications depend on one service, the AI leader should make the shared constraint visible. A local workaround may help one team while increasing complexity or exposure across the platform. Conversely, insisting on a central solution for every narrow need can delay useful work. The decision should compare the scope, reuse, control and maintenance cost of the alternatives rather than assume that one ownership model is always superior.

The practical output is a responsibility boundary that can be tested. Ask each team to demonstrate the evidence it supplies and the action it can take when its component fails. The application owner remains accountable for the service outcome, while relying on explicit contributions from the platform and source owners. This creates a release decision based on actual operating responsibilities rather than a diagram in which every important control appears somewhere but belongs to nobody.

Exhibit 3. Reported production constraints



BOTTLENECK | Registration range: 2026-08-30 to 2026-09-08 UTC | Selective self-report sample

Exact selection	Count	Base	Any mention
Data access and quality	29	75	39%
Compute	22	75	29%
Integration with existing systems	21	75	28%
Inference cost	19	75	25%
Governance and approval	16	75	21%
Agent identity and permissions	11	75	15%
Talent	11	75	15%

Exhibit note: BOTTLENECK; carried-forward operator instrument, multi-select. Categories can coexist. The distribution does not measure the cause of a failure or prove that removing one constraint would produce a particular financial result.

5. Keep authority tied to the actual action

The access-method question has 71 answers: 24 per-agent credentials, 18 shared service accounts, four delegated user identity, three credential-free or brokered access and 22 not applicable yet. The applicable base is 49. These counts describe the methods respondents selected; they do not establish how narrowly permissions are scoped or how effectively actions are controlled. [ACCESS]

For the AI leader, the release issue is what the application can actually do. A read workflow and a write workflow may use the same identity infrastructure while creating different consequences. The owner should identify the permitted resources, actions and initiating authority, then demonstrate the enforcement that keeps execution within that scope. The CISO report develops the control review; the application team must supply evidence from the deployed path.

Human intervention should be designed around a decision the person can meaningfully make. An approval step is weak if the reviewer cannot see the relevant action, parameters or consequence, or if the system changes those details afterward. The release review should examine what information is presented, which actions are blocked pending approval and what happens when approval is denied or delayed. The existence of a button is not sufficient evidence of oversight.

The application also needs a defined response to uncertainty. Missing data, ambiguous instructions or conflicting evidence may require clarification, a narrower result or escalation. The appropriate behavior depends on the service. The team should test that behavior without treating a refusal rate as inherently good or bad. A system that refuses useful work unnecessarily and a system that proceeds without authority create different problems.

Changes in tools and integrations can alter authority indirectly. A previously read-only connector may gain a write function; a downstream service may expose a broader scope; a workflow may begin sending information to a new destination. The application owner should assess those changes before relying on the old release approval. This is why a register of model names alone is inadequate for governing production behavior.

The release record should end with a clear statement of the actions approved and the person who can change that approval. It should also identify how the workflow can be constrained during an incident without unnecessarily disabling unrelated services. The technical design and the management decision need to meet at that boundary, so that the team can act quickly when the observed behavior differs from the approved purpose.

6. Plan service recovery before widening the commitment

Reliability includes what the service does when a dependency fails or a result cannot be accepted. A workflow may be useful even when it sometimes requires human intervention, provided that the handoff is timely, clear and supported by adequate capacity. The AI leader should evaluate the whole service, including fallback and backlog, rather than report only the success of completed automated cases.

Define the intervention path before expansion. Who receives the unresolved case, what context accompanies it and how is duplicate work prevented? If the agent has already taken a partial action, the receiving person needs to know what occurred and what remains possible. A generic error message may be insufficient for restoring the intended business outcome safely.

The team should distinguish an application pause from a service outage. It may be possible to narrow the workflow, disable a consequential action or route a subset of cases to people while preserving useful functions. Those options should be designed and tested where they matter. During an incident, the organization should not have to choose only between unrestricted operation and a complete shutdown because no intermediate mode was considered.

Recovery capacity can become a scaling constraint. A fallback that works for a small pilot may overwhelm the human team at production volume. The expansion review should estimate the plausible exception load and examine what happens when failures are correlated, such as a shared source becoming unavailable. Average escalation rates alone may not capture that concentration of demand.

Restoration needs an owner and evidence. The team should know what must be fixed, which pending actions require review and whether the original scope can be restored safely. The business owner should understand any temporary limitation. A rapid restart that recreates the same unresolved failure does not establish resilience, even if it reduces the apparent outage duration.

The final production commitment should therefore include accepted outcomes, failure handling and the ability to change operating mode. Planned spending can indicate ambition, but it cannot establish that these capabilities exist. In OFF's separate spending question, the answers are forward-looking bands, not observed investment or capacity. The release decision must rely on the service evidence the team can demonstrate now, with additional commitments made only as that evidence supports them. [SPEND]

External guidance: testing the system that will actually run

Anthropic's January 2026 engineering guidance distinguishes a trial's transcript from the resulting state of the environment. [E-ANTHROPIC-EVAL] That distinction is especially useful for action-taking workflows: an agent's statement that it completed a task is not the same as an independently verified outcome. The report's release record asks the team to identify how success is checked in the actual service, while recognizing that a vendor's engineering guidance is not an independent evaluation of the company's application.

FinOps unit-economics guidance separates technical resource measures from business units. [E-FINOPS] The application here is to connect consumption, retries and human intervention to accepted service outcomes. A team may need both technical and business views to explain a change. The guidance does not establish a target cost or prove that a particular architecture is economical.

The MCP authorization specification supplies concrete requirements within its protocol scope, including audience validation and restrictions on token passthrough. [E-MCP] For a production owner, the broader lesson is to identify which security property is actually guaranteed by the protocol implementation and which remains a deployment responsibility. Use of a recognized protocol does not, by itself, establish that every tool action is appropriately authorized.

NIST's voluntary AI risk framework places governance and ongoing management alongside mapping and measurement. [E-NIST] This supports treating release evidence as part of a continuing operating process. The November register is an OFF recommendation that connects those responsibilities; it is not a NIST certification checklist or a substitute for applicable requirements in the organization's sector.

The sources are deliberately not presented as four independent studies proving the same result. One is vendor engineering guidance, one a cost-management framework, one a technical specification and one an institutional risk framework. They help define what to inspect. The OFF survey identifies reported practices and constraints, while the deployment team must supply the evidence of actual implementation and performance.

The difficult tradeoff is speed. Teams cannot rerun every possible test after every minor change, and a test set can never cover all future work. The recommended approach is to preserve a stable regression basis, add evidence for the changed boundary and monitor the live service for failures the evaluation missed. The organization should state the remaining uncertainty and retain an effective intervention path. That is a more credible release argument than either claiming exhaustive safety or relying entirely on post-launch complaints.

Source	Contribution to the release review	Limit
Anthropic, Demystifying evals for AI agents	Distinguishes observed outcome from the agent's account	Vendor engineering guidance, not application certification
FinOps unit economics	Connects resource use to useful service	No universal cost target
MCP authorization specification	Defines protocol-specific requirements	Deployment and business authorization remain separate
NIST AI RMF	Connects measurement with ownership and ongoing management	Voluntary framework, not proof of control effectiveness

Practical application: a change that deserves a new gate

Consider an illustrative customer-service agent initially approved to draft answers for human review. The proposed expansion would let it issue a limited class of account credits directly. The model may be unchanged, and the team may have strong evidence that its drafts are useful. The expansion nevertheless changes the action boundary, financial consequence and recovery requirements. This is an invented scenario, not an OFF respondent case.

The application owner should identify what existing evidence still applies. The knowledge-retrieval evaluation may remain relevant, while the authorization and action tests require additional work. Finance must define the permitted credit and its accounting treatment within the company's policies. The business owner must define eligible cases and customer commitments. Security and legal teams must assess the relevant access and obligations. The AI leader coordinates the release evidence rather than claiming that the previous drafting score resolves all these questions.

Working tool: production control register

Field	Required content	Owner
Service boundary	Users, tasks, data, permitted actions and accepted outcome	Application / business owner
Version and change	Model, tools, source definitions and material configuration	Application owner
Evaluation	Test coverage, grading, repeated-trial treatment and unresolved failures	Evaluation owner
Economics	Accepted-outcome denominator, retries, review and cost drivers	Engineering and finance
Authority	Effective permissions, approvals and intervention path	Application and identity teams
Service recovery	Escalation capacity, constrained modes and restoration criteria	Service owner
Evidence validity	Review date and change that invalidates the result	Release approver

The test should verify the account state after an approved action and confirm that disallowed cases remain blocked or routed appropriately. It should include repeated requests and interrupted execution where those can create duplicate or partial actions. The team should record what the system does when required information is missing. These are defensive release tests, not claims that the survey measured a particular failure rate.

The economics should include the changed operating mix. Direct action may reduce review for straightforward cases while increasing investigation of exceptions. A lower average human touch rate is not enough if the remaining cases require more specialist time or if erroneous credits create costly recovery. The team should define the accepted outcome and measure the associated work before extrapolating savings to all customers.

Working tool: cross-functional release / expansion gate

Question	Evidence	Decision / failure condition
What changed?	Explicit comparison of old and proposed service boundary	Hold an expansion whose consequence is not defined
Is the result acceptable?	Outcome tests and qualified review for relevant cases	Narrow or retest unsupported work
Is authority constrained?	Effective permissions, approval and negative tests	Do not release consequential actions without a demonstrated boundary
Are economics understood?	Cost drivers and accepted-output comparison	Bound exposure while unresolved costs are measured
Can the service recover?	Escalation and constrained-mode demonstration	Limit volume to supported recovery capacity
Who can change course?	Named pause, restore and expansion owners	Resolve missing authority before release

The release decision may authorize a limited set of cases and users first, with defined monitoring and a review point. The organization should choose the scope and thresholds based on its actual service and risk tolerance. A generic percentage or fixed number of days would create false precision here. The important requirement is that the team can explain why the chosen boundary fits the evidence and what would cause it to tighten or expand.

After release, compare observed cases with the evaluated population. Record new failure types, unexpected cost drivers and work that people handle outside the intended process. Feed those observations into the next evaluation and operating decision. The register should remain a current account of the service, not an archive created once to obtain approval. If the boundary changes again, the same structure helps the team determine which evidence can be reused and which must be renewed.

The next production review

Start with the material workflows proposed for expansion. Ask their owners to state the change in users, volume, data or actions and identify which existing evidence covers it. Require a clear accepted outcome and a record of unresolved failures. This should make the decision more specific than whether the team is comfortable with the model.

Next, reconcile the application and platform responsibilities. Confirm who supplies each control, who maintains the source definitions and who can intervene. The business owner should know the service limits, while finance should understand the cost drivers and the denominator used in the benefit case. Missing ownership should be resolved before it becomes a production incident.

Finally, test the available response when the service falls outside its intended boundary. A spending alert, a failed evaluation or a degraded source must lead to an action that someone can take. The organization

should know whether it will constrain a function, route work to people, pause expansion or restore an alternative process.

The November operator figures remain a selective historical snapshot. Their value is to focus attention on production questions that deserve direct evidence. The AI leader's deliverable is a service whose outcome, economics, authority and recovery can be explained under the scope actually being approved, with a record that makes the next change easier to assess.

Reader questions

Did the production-agent share rise in October?

No comparable new operator answers establish a change. The 60/75, or 80%, result is carried forward. [FLEET_AT_LEAST_ONE]

Does a yes answer to cost visibility prove fully loaded workflow costs?

No. The exact question asks about real-time visibility into what AI costs to run. Workflow attribution and cost completeness require separate evidence. [COST]

Why use the matched cost result?

It describes the same 60 respondents who both reported production agents and answered the cost question. It avoids comparing unrelated denominators. The result is historical and self-reported. [OPERATOR_COST]

Does the access-method question prove least privilege?

No. The selected architecture label does not establish the effective permissions, controls or revocation behavior. [ACCESS]

Is planned spending actual investment?

No. The question records next-twelve-month spending bands. It cannot establish current spend, procurement commitments or a market total. [SPEND]

Must every small change trigger a complete release review?

The report recommends review proportionate to the changed boundary and consequence, with reusable evidence and routine changes delegated.

Is a successful final message enough to pass an action test?

No. Where a task changes an external state, verify the intended result and authorization in that state. [E-ANTHROPIC-EVAL; recommendation]

Citable facts

- At least one production agent: 60/75, 80%; more than twenty: 22/75, 29%, carried-forward operator sample. [FLEET]

- Real-time AI running-cost visibility: yes 34, partly 30, no 13, base 77. [COST]

- Matched production operators: yes 30, partly 23, no seven, base 60; already reported in October. [OPERATOR_COST]

- Applicable access-method base: 49 of 71, after excluding 22 not applicable yet. [ACCESS_APPLICABLE]

Methods and limitations

Calculation-v1.1 uses authorized 1 November exports and the stated cutoff. Byte-identical exports and invited-only records are excluded. Latest eligible nonblank answers are selected per normalized respondent identifier per question. Registration timestamps do not establish answer-edit dates. Each distribution retains its own denominator, unknowns and not-applicable responses.

The source export has 314 eligible application rows. Item bases are SEAT 101, MATURITY 91, SPEND 76, COST 77, FLEET 75, BOTTLENECK 75 and ACCESS 71. The matched fleet/cost set has 75 respondents, of whom 60 report at least one production agent. The applicable access base is 49. No unmatched percentage difference is presented as a measured gap.

The observations are selective, self-reported and cross-sectional. No deployment audit, benchmark ranking, causal inference, transition rate, company-level ROI or average spending estimate is claimed. Proposed release gates and the service-agent scenario are recommendations and illustrations. The report does not claim that a vendor or deployment passed Murray Newlands' separate synthetic governance benchmark.

References

[OFF-DATA] Open Future Forum. Executive AI Statistics, November 2026; source-specific distributions, matched results and aggregate data.

[E-ANTHROPIC-EVAL] Anthropic Engineering. Demystifying evals for AI agents. 9 January 2026. https://www.anthropic.com/research/demystifying-evals-for-ai-agents

[E-FINOPS] FinOps Foundation. Unit Economics capability, current guidance consulted 2 November 2026. https://www.finops.org/framework/capabilities/unit-economics/

[E-MCP] Model Context Protocol. Authorization Security Considerations, specification 2026-07-28. https://github.com/modelcontextprotocol/modelcontextprotocol/blob/main/docs/specification/2026-07-28/basic/authorization/security-considerations.mdx

[E-NIST] NIST. AI Risk Management Framework 1.0, current official resource page. https://airc.nist.gov/airmf-resources/airmf/

[OFF-FLEET] Open Future Forum. Operating More Than 20 Production AI Agents, October 2026. https://openfutureforum.com/blog/operating-more-than-20-production-ai-agents

[MN-GOVERNANCE] Newlands, M. Enterprise Data Agent Governance, version 0.1.0, 15 September 2026. Personal practitioner resource, distinct from OFF research. https://huggingface.co/datasets/murraylovecode/enterprise-data-agent-governance

Related research and citation

The November CISO report tests the ending of authority, CFO qualifies financial benefits, Transformation examines operating dependence and CEO assigns cross-functional decision rights.

Suggested citation: Newlands, Murray. AI Leaders AI Leverage Report, November 2026, Edition 2. Open Future Forum. Evidence cutoff 31 October 2026.

Open Future Forum research and methods: https://openfutureforum.com/research/

Editable working tool

Change-sensitive release gate: download the editable CSV worksheet. The worksheet implements the report's recommendation; it is not a validated score, mandatory standard or completed assessment.