David Soden

MARKET ASSESSMENT | AUGUST 2026

The Evidence Gap in Agent Governance

What supervisors are about to require autonomous agents to prove, and what shipped control planes can actually emit

David Soden

Independent analyst and practitioner, enterprise automation and applied AI

davidsoden.com/market-assessment

REPORT DS-2026-616A · 2026 · VERSION 1.0 · PUBLIC

About this report

Every substantive claim in this report carries an evidence classification and, where applicable, a numbered source. Facts are separated from vendor claims, third-party estimates, the author's assessments, and untested hypotheses. Forward-looking sections use scenarios with observable tripwires rather than point forecasts. Stated assumptions are listed before the analysis begins.

Market Assessment reports are published and commissioned at **davidsoden.com/market-assessment**. This document is analysis, not investment, legal, or procurement advice.

Scope and evidentiary standard

Four governance regimes published material on autonomous agents between May and August 2026, and three vendor mechanisms shipped into general availability over roughly the same window. This report reads the regimes against the products, artefact by artefact, and asks a narrow question: when a supervisor, an auditor, or a procurement officer asks an operator to prove how an agent behaved, what exactly is the operator expected to hand over, and can anything currently on sale produce it.

The regulatory side covers the Monetary Authority of Singapore's proposed Guidelines on Artificial Intelligence Risk Management and the industry SAFR framework MAS convened alongside them; the UK National Cyber Security Centre's interim advice of 20 August 2026 and its May 2026 predecessor; the United Arab Emirates federal programme deciding which government tasks may be handed to an agent; and the United States Marine Corps agent registry effort. The vendor side covers Amazon Bedrock AgentCore Policy, the Dogwood temporal policy language AWS open-sourced on 6 August 2026, the zero standing privilege access model as its advocates describe it, and the state of automated agent red-teaming.

This category has almost no published sizing worth citing, and that is deliberate on my part rather than an omission. Agent-count forecasts circulate widely, trace to analyst firms through trade coverage rather than to a readable methodology, and would not change a single conclusion here if they were doubled or halved. Where I mention one I say where it came from and decline to rest anything on it. What the analysis does rest on is the text of the published documents, which anyone can read, and the documented behaviour of shipped software, which anyone can test.

Three questions organise the report. What specific evidence do the published regimes demand. Can the shipped tooling emit that evidence. On what criteria is the automate-or-not boundary being drawn, and does a published taxonomy exist. Each is settled from primary documents where primary documents exist, and each answer says plainly when they do not.

Label	Definition	How to treat it
FACT	Verifiable in the cited primary source: a filing, press release, official documentation, or standards body announcement.	Reliable as stated. Check the source if the exact wording matters.
VENDOR CLAIM	Reported by a vendor about its own business or product. Self-measured and not audited by any third party.	The vendor said it. Directionally useful, not a measurement. Definitions of the metric are usually undisclosed.
THIRD-PARTY ESTIMATE	A research firm forecast or survey result. Methodology is proprietary and generally not reproducible.	Use for direction and order of magnitude only. Do not build a model on it.
ASSESSMENT	The author's reasoned judgment from the cited evidence. Falsifiable, and the reasoning is shown so it can be argued with.	This is analysis, not reporting. Disagree with it where your evidence differs.
HYPOTHESIS	A proposition offered for testing. Not currently supported by sufficient evidence to assert.	Treat as a research question, not a conclusion. Each one names what would confirm or refute it.
SCENARIO	A structured future state with a stated subjective probability. The probability is the author's judgment, not a calculated figure.	Use the leading indicators attached to each, not the probability. The indicators are observable; the probability is not.

Assumptions this analysis rests on

Five premises are assumed rather than shown. Each carries a note on what fails if it turns out to be wrong.

1. MAS finalises its Guidelines substantially as consulted. The consultation closed on 31 January 2026 and the Chairman told Parliament on 5 August 2026 that they will be finalised soon. If the lifecycle control sections move materially, the Singapore column of the tabulation moves with them and the finding that MAS does not require a per-action record could be overturned in a single publication.

2. The NCSC's forthcoming formal guidance keeps the substance of the interim advice. The interim post says formal guidance will supersede it. If the logging and immutability language is softened on the way through, the UK becomes the weakest of the four regimes rather than the most demanding on evidence, and the control checklist loses its strongest citation.

3. AWS's documented limits are current and apply as written. The 24-hour trajectory window and the single concurrent authorization request per session are the two numbers doing the most work in the assessment of what temporal policy can evidence. Both are the kind of limit a cloud provider raises quietly. If they move, the second half of the analysis needs rerunning, though the integrity argument survives either way.

4. Reading the published texts is the right way to answer the first question. Supervisors routinely ask for things that appear nowhere in a guideline, and examination practice diverges from published expectation within about two cycles. If that happens here, this report understates demand rather than overstating it, which is the safer direction to be wrong in.

5. No UAE task classification taxonomy has been published. I searched for one in English and found programme descriptions, workshop readouts and ministerial statements that describe classification work in progress. If a framework exists in Arabic, in a Cabinet circular, or behind a government portal I cannot reach, the central claim of the third section is wrong and the section should be discarded rather than discounted.

The call

ASSESSMENT Four regimes now ask agent operators for the same four things: a registered identity per agent, a record of the authority under which each action was taken, a record of every escalation and the human decision that closed it, and enough of the tool-call sequence to reconstruct what happened without trusting the agent's own account. No shipped product emits all four. The gap is not in policy expression, which has run ahead of the guidance, but in the retention and integrity of the decision record. I would change position if a major control plane shipped an authenticated action trace with a stated retention period. Until then buyers are purchasing enforcement and filing it as evidence.

My judgment on the evidence assembled here, nothing more. There is data I can't see and data I may have missed. New evidence can move me, including to the opposite position, and I reserve the right to move.

Executive summary

The published regimes are more consistent with each other than the products built to satisfy them are with any of the regimes. Read side by side, four documents from three jurisdictions converge on a short list of artefacts. Read against the release notes, the shipped control planes are strong on stopping an action and weak on proving what happened, which is the opposite of what the guidance emphasises.

FINDING 1 The regimes converge on a short artefact list, and diverge on one item

> ASSESSMENT Agent registration, distinct per-agent identity, pre-deployment validation, a trace of the tool-call sequence, an escalation record, and a halt capability appear in more than one regime each. Not one of the four documents states a retention period for any of it. A record with no stated retention is an operational log, not an evidentiary artefact, and the distinction will surface the first time one is requested after the fact. [1] [2] [4] [5]

FINDING 2 Singapore's binding-ish instrument does not require a per-action record; the document that does is explicitly not regulation

FACT The proposed MAS Guidelines require documentation sufficient for an independent party to replicate the development process, and independent pre-deployment review. On runtime, they ask that "information flow and decision-making paths across workflows that use AI, such as reasoning processes, actions taken, tools used" be monitored, hedged with "where relevant". SAFR specifies the per-action record in detail and states it "does not constitute regulatory guidance or supervisory expectations". [1] [2]

FINDING 3 The NCSC asks for the most and can compel the least

FACT The 20 August interim advice asks for chain of thought traces and transcripts, logs "protected from modification or deletion" and immutable "where possible", a distinct identity per agent, and the ability to pull the plug. It is a blog post that the NCSC says formal guidance will supersede. On evidence demanded per unit of legal force, it is the most demanding document in the set. [4]

FINDING 4 AgentCore Policy is an enforcement point that produces telemetry, not an action trace

FACT Generally available since 3 March 2026 across thirteen regions, it intercepts every tools/list and tools/call at the Gateway, denies by default, and logs each decision to CloudWatch. What it retains is what was requested, by which principal, and what was decided. That is a strong access log and it is not the reasoning path the NCSC and SAFR both ask for. [9] [11] [4] [2]

FINDING 5 Dogwood's own authors document a concurrency defeat and the loss of formal analysis

FACT AWS's own write-up shows a rate limit written over response events being beaten by three concurrent transfers slipping past a $5,000 cap, because a policy summing responses sees nothing in flight. It also states that temporal conditions do not support Cedar's automated reasoning, so a policy set using them cannot be formally analysed. Publishing both is to AWS's credit and neither is fixed by publishing it. [6] [18]

FINDING 6 A 24-hour trajectory window cannot carry a reporting period

FACT Agent trajectories in AgentCore carry a maximum look-back of 24 hours and older events are automatically deleted. There can be no more than one concurrent authorization request per session, and AWS recommends keeping session scope narrow, which fragments the history further. Any control expressed over a week, a quarter, or a customer relationship is outside what the mechanism can see. [7]

FINDING 7 Automated red-teaming cannot yet stand as pre-deployment validation

FACT A meta-audit of three agent safety benchmarks across twelve environments and eight backbone models found that existing benchmarks leave more than 20 percent of unsafe interaction patterns uncovered, and surfaced around 11 percent novel patterns using the same tools and environments. An agent that passes the suite has demonstrated that it passes the suite. [12]

FINDING 8 The automate-or-not line is being drawn on operational criteria, and no public taxonomy exists

ASSESSMENT The UAE assessment sorts candidate work by transaction volume, procedure clarity, documentation quality, data currency and existing automation level. Those are readiness tests, close to a classical process automation screen. Reversibility, blast radius and contestability of the decision do not appear in the published criteria. The classification framework itself remains unpublished. [13] [14]

FINDING 9 The weakest link is the agent's account of itself, and SAFR says so

FACT SAFR notes that the action trace and the action details are both agent-declared contents of the same envelope and "can be fabricated together by a sophisticated adversarial injection that maintains internal consistency while departing from the original task". Its answer is to treat the envelope as a document authenticated against its origin. No shipped product I could find does that authentication. [2]

Four regimes, read side by side

The four documents were written by different institutions for different audiences with no evident coordination, which makes the overlap between them informative. Where three of four independently ask for the same artefact, that artefact is likely to appear in whatever consolidates later.

FACT The MAS Guidelines apply to all AI use by financial institutions including agentic AI, per the Chairman's written reply to Parliament on 5 August 2026, and "will be finalised soon". MAS proposes a transition period of 12 months after the Guidelines are issued. [3] [1]

FACT The proposed Guidelines are organised into oversight of AI risk management, risk management systems and procedures, and lifecycle controls, capabilities and capacity. Paragraph numbers cited throughout this report are to the text annexed to the November 2025 consultation, read directly and cross-checked against law firm summaries of the same document. If the final version renumbers, the references need remapping even where the substance survives. [1] [22]

FACT MAS's own classification of its regulatory instruments places Guidelines below Notices. Guidelines "set out principles or 'best practice standards'", contravening them "is not a criminal offence and does not attract civil penalties", but "how well an institution or person observes the guidelines may have an impact on MAS' overall risk assessment". Notices, by contrast, "impose legally binding requirements on a specified class of financial institutions". [21]

ASSESSMENT That distinction is the single most useful thing a bank risk officer can hold on to this year. The AI Guidelines are not law, and a supervisor's view of how well you observe them still shapes your risk rating, your examination intensity, and in practice your capital and governance conversations. Treating them as optional because they are not Notices misreads how Singapore supervises. [21]

FACT The NCSC published its interim advice on 20 August 2026 as a blog post by a Principal Security Architect, describing it as interim practical advice pending formal guidance. Its May 2026 predecessor set the bar for deployment readiness in one sentence: "If you cannot understand, monitor or contain an agent's actions, it is not ready for deployment." [4] [5] [16]

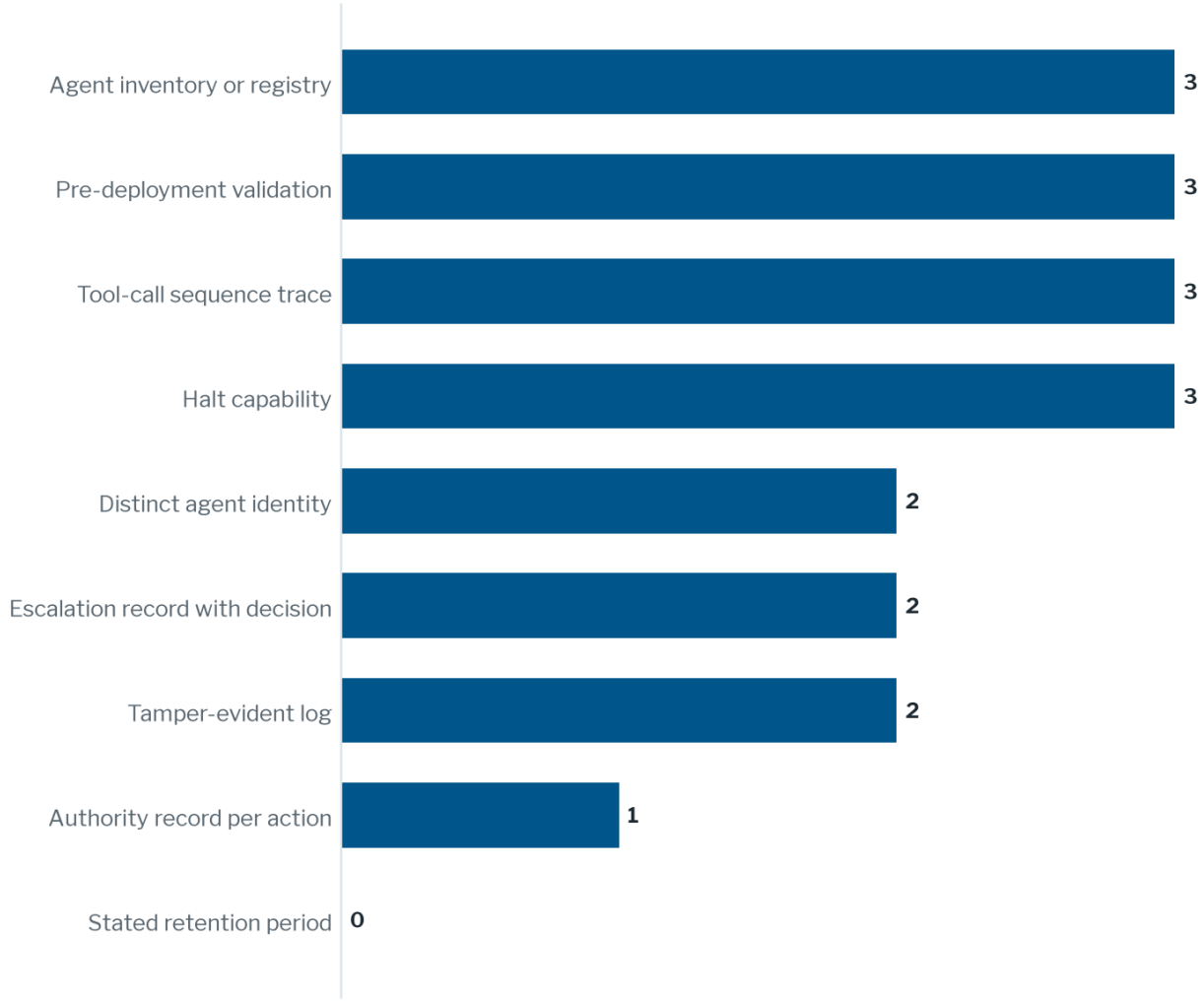
THIRD-PARTY ESTIMATE Trade coverage of the Singapore position, published two days before the NCSC post, summarises the supervisory expectation as four demonstrables: that agents respect their assigned mandates, that transactions are validated before execution, that unexpected behaviour triggers escalation, and that every decision remains auditable. That is a fair reading of where MAS and SAFR point together, and it is a journalist's synthesis rather than a quotation from either document. [17]

FACT SAFR was published as a MAS information paper at version 1.0 in July 2026, developed with financial industry members under the BuildFin.ai initiative. It defines four runtime components, a Governance Envelope carrying the proposed action, the action trace and context metadata, four dispositions of Deny, Escalate, Auto-Execute and Observe, and an append-only tamper-evident Audit Log. [2]

Evidence artefact	MAS Guidelines	MAS SAFR	NCSC, Aug 2026	UAE and USMC
Agent inventory or registry	Required, para 3.4, with attributes at 3.5	Registry entry verified before any other check	Not named as a registry	USMC registry in build; UAE not stated
Distinct identity per agent	Not stated	Agent Identity is the first component	"own unique identity in a class" distinct from humans	USMC says identity work still needed
Pre-deployment validation	Required, paras 4.14 and 4.18, review by parties not involved in development	Explicitly out of scope, runtime only	Threat model and bounded pilots first	UAE 90-day service assessment
Authority record per action	Not stated	Machine-readable mandate travels in every envelope	Implied by least privilege, not specified	Not stated
Escalation record with the human decision	Document and review oversight decisions and interventions, para 4.10(d)	Escalate disposition logged with basis and elapsed time	Named responsibility and ability to intervene	Not stated
Tool-call sequence trace	"Where relevant", monitor reasoning processes, actions taken, tools used, para 4.23(a)	Action trace preserved at the point of proposal	Chain of thought traces and transcripts	Not stated

Evidence artefact	MAS Guidelines	MAS SAFR	NCSC, Aug 2026	UAE and USMC
Tamper-evident or immutable log	Not stated	Append-only, immutable once written	Protected from modification or deletion, immutable "where possible"	Not stated
Stated retention period	Not stated	Not stated	Not stated	Not stated
Halt or kill switch	"kill switches" to consider for high risk materiality, para 4.23(b)	Deny disposition rejects before execution	Emergency shutdown, including network access	Not stated
Legal force	Guidelines: no criminal or civil penalty, affects supervisory risk assessment	"does not constitute regulatory guidance or supervisory expectations"	Interim blog post, to be superseded	Programme mandate and internal service policy

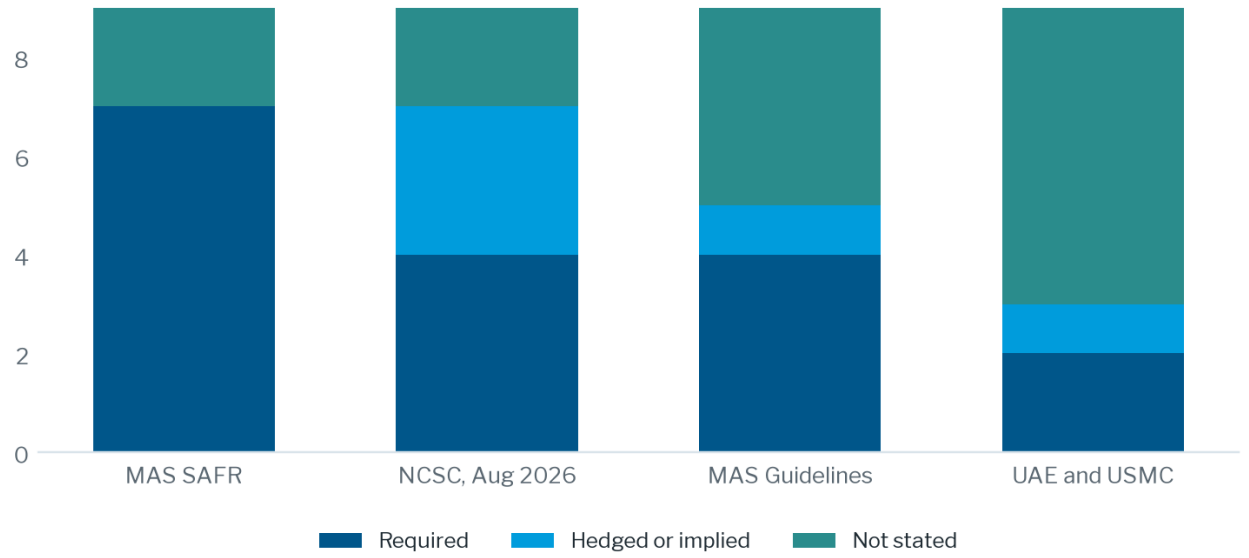
FIGURE 1 **Evidence artefacts, by how many of the four regimes ask for them explicitly**



ASSESSMENT My coding of the four texts, counting only artefacts named or clearly described. Hedged language such as "where relevant" is counted as present. Another reader coding the same texts would move one or two bars by a point. Nobody would move the last one. [1] [2] [4] [13] [15]

ASSESSMENT Retention is the interesting zero. Every one of these documents describes a record that exists to be consulted later, and not one says how much later. That is normal for a first pass at a new technology and it will not survive contact with a records management function, because every other financial and government record already has a schedule attached to it. Expect retention to arrive in the second version of each of these documents, and expect it to arrive as a number of years rather than a number of hours. [1] [2] [4]

FIGURE 2 How each regime covers the nine artefacts



ASSESSMENT Derived from the table above, so it inherits the same coding judgment. The shape of the chart is the argument: the most complete regime is the one with no legal force at all, and the one that will actually be examined is the third bar. [1] [2] [4] [13] [15]

ASSESSMENT SAFR reads like a specification because it is one. It was written by practitioners solving a design problem rather than by a regulator setting a floor, and it shows in the detail: the Controls Repository requires five elements per control including a validity period and the principal authority behind it, and dispositions are calibrated on reversibility, financial materiality, customer impact, regulatory sensitivity and novelty. That is a better articulated risk model than anything in the binding documents. [2]

FACT SAFR also states the principle that breaks most current implementations: in multi-step workflows the control flow applies to each action independently, and "an Auto-Execute or Observe outcome at one step carries no authority into the next". Prior authorisation does not carry forward as the agent adapts to intermediate results. [2]

ASSESSMENT Read that against how agent frameworks are typically built and the implication is uncomfortable. A great many deployments authorise a task, not an action, and then let the agent run the plan. Under SAFR's rule every step in that plan needs its own decision and its own log entry, which multiplies the volume of governance events by the length of the plan. That is a cost problem before it is a compliance problem, and it is the reason I expect per-action governance to be adopted last by the teams with the longest workflows. [2]

What the tooling can actually emit

The second question is answerable more precisely than the first, because policy language semantics and documented operator limits are matters of record rather than interpretation. AWS has published unusually candid documentation here and most of what follows is drawn from it.

AgentCore Policy: the enforcement point

FACT Policy in Amazon Bedrock AgentCore went generally available on 3 March 2026 in thirteen regions. It stores Cedar policies in a policy engine attached to an AgentCore Gateway, and the Gateway "intercepts agent-tool traffic and evaluates each request against the policies" before the tool is invoked. Teams can author in natural language, which the service converts to Cedar. [9]

FACT Evaluation is default deny and forbid always wins: if no forbid matches and at least one permit matches the decision is allow, and if neither matches the decision is deny. An empty policy engine denies everything. Denied calls return a structured MCP error with an AuthorizeActionException naming the reason. [11]

FACT Every enforcement decision is logged through CloudWatch metrics and logs, described as "a continuous audit trail of policy evaluations, what was requested, by which principal, and what the decision was". X-Ray traces can localise where in the request path an authorization check failed. [11]

ASSESSMENT That is a good access log and it answers a different question from the one the guidance asks. What was requested, by whom, and what was decided establishes that a control fired. It does not establish why the agent proposed the action, what it had read beforehand, or which retrieved data the proposal rested on. The NCSC asks for chain of thought traces and transcripts; SAFR asks for the steps taken, tools used, data retrieved and checks performed. Neither is in the policy decision record. [4] [2] [11]

FACT Three documented boundaries matter more than the quotas. Policy governs agent-to-tool interactions through a Gateway and does not apply to tool calls that bypass it: an agent invoking an SDK directly from its own code never crosses the policy boundary. Tool list filtering is "a hardening and UX feature, not the enforcement point". And if the Gateway execution role lacks the GetPolicyEngine permission, LOG_ONLY mode can fail silently, so everything appears healthy while the engine was never evaluating. [11]

ASSESSMENT The silent LOG_ONLY failure is the one to put in a test plan. A control that looks configured, reports nothing wrong, and is not running is worse than no control, because it produces a clean record of an evaluation that did not happen. Any assurance exercise over this stack should include an intentionally forbidden call in LOG_ONLY and a check that it was actually evaluated. [11]

Dogwood: policy language ahead of the guidance

FACT AWS open-sourced Dogwood on 6 August 2026 under Apache 2.0, authored by Marc Brooker, Joseph Tassarotti and Jean-Baptiste Tristan. It embeds Cedar and adds a when temporal clause that reads the agent's event history. Four operators cover the common shapes, defined as

standard-library macros over Metric First-Order Temporal Logic rather than as language primitives: formerly for whether something happened in a window, count_within for how many times, count_distinct_within for how many different values, and sum_within for a running total, with bind naming aggregates for comparison. [6] [8]

ASSESSMENT This is genuinely ahead of every regime in the tabulation. Nothing in MAS, NCSC or the public sector documents asks for a constraint expressed over a sequence, and Dogwood lets you write one that says an approval must precede a trade, market data must be under thirty seconds old, and a portfolio load must follow a client profile fetch. The guidance asks for a record of what happened. Dogwood attempts to make certain sequences impossible. Those are complementary, and only one of them is what an auditor will ask for. [7] [6]

FACT AWS documents the operational limits plainly. Agent trajectories carry a maximum look-back of 24 hours and older events are automatically deleted. There can be no more than one concurrent authorization request per session, and AWS recommends keeping session scope as narrow as possible. Any change to the policies in an engine invalidates existing sessions. The first 100 temporal policies per engine are included in the existing per-authorization-request price. [7]

Documented limit	Value	What it does to the evidence
Trajectory look-back window	24 hours	No temporal control can be expressed over a week, a quarter or a customer relationship. Events outside the window are deleted, not archived, so the sequence cannot be reconstructed later even manually.
Concurrent authorization requests per session	1	Serialises decisions within a session and pushes designers toward narrow sessions, which fragments a single business process across several unlinked trajectories.
Effect of a policy change	Existing sessions invalidated	Updating a control mid-flight discards the history the next decision would have read. Tightening a limit during an incident is exactly when this bites.
Temporal policies included in price	First 100 per engine	Sequence constraints carry a marginal cost past a hundred, which will shape how many get written.
Formal analysis of temporal policies	Not supported	A policy set using temporal conditions loses Cedar's automated reasoning, so it cannot be proven free of contradictions or over-permission. Review becomes manual.
Reference interpreter	Not for production	The open-source implementation validates semantics. Enforcement infrastructure, event authentication, durable storage, tenant isolation and retention are the adopter's problem.

FACT The reference implementation carries its own warnings. The interpreter "accepts timestamps as provided and does not validate them", the built-in in-memory temporal engine "has no eviction or size cap", event traces are lost on restart because storage is purely in memory, and there is no event authentication mechanism. [8]

ASSESSMENT Publishing that list is the right call and buyers should read it as what it is: a statement that the trustworthy event history, which is the entire foundation of temporal authorization, does not ship. An unvalidated timestamp in an authorization decision is a forgeable input. Whoever adopts this is building the evidentiary layer themselves, and most will discover that after the proof of concept. [8] [6]

The vendor's own claims, separated

VENDOR CLAIM AWS says decisions made by temporal policies are "deterministic, deny by default, and logged with the full context behind them", so reviewers can see "not only that a call was blocked but why". Self-reported, unaudited, and describing the managed service rather than the open-source interpreter. [10]

ASSESSMENT The claim is narrower than it reads. Full context behind a decision means the policy that matched and the trajectory events it read, which is real and useful. It is not the agent's reasoning, and the trajectory it read expires in a day. A reviewer asking why a call was blocked is well served. A reviewer asking why the agent wanted to make the call in the first place, three weeks later, is not. [10] [7]

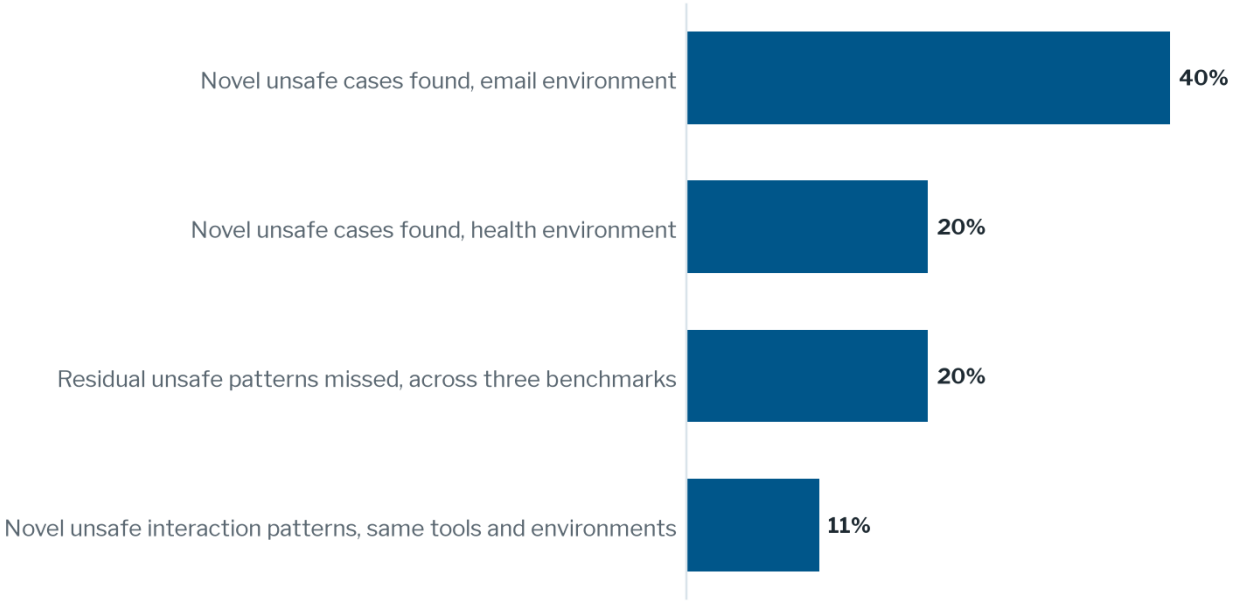
VENDOR CLAIM The zero standing privilege case, as argued by Britive's chief executive in August 2026, is that no identity holds always-on access by default, access is "minted when needed, scoped to specific tasks, and destroyed upon completion", and "every action is its own logged, time-bounded event, traceable from the person or process that initiated it". The piece cites no data and names no measurement. [19]

ASSESSMENT The diagnosis is sound and widely shared; the evidence offered for it is zero. Credentials outliving tasks because an agent has no concept of done is a real failure mode that anyone who has run agents recognises. But an access model tells you what an identity was permitted to do, not what it did or why, and the last clause of that vendor claim quietly imports an action record that the access model does not itself produce. Buyers should treat zero standing privilege as necessary and nowhere near sufficient for the evidence question. [19] [2]

Automated red-teaming and the pre-deployment claim

FACT A meta-audit published in August 2026 by researchers at Purdue introduced SafeAudit, which enumerates valid tool-call workflows and distils compact safety rules from existing benchmarks to find unsafe patterns those rules miss. Across three benchmarks, twelve environments and eight backbone models it found existing benchmarks "systematically leave more than 20 percent of unsafe interaction patterns uncovered", and identified around 11 percent novel patterns under the same tools and environments. Coverage grew monotonically with testing budget. [12]

FIGURE 3 What benchmark-based agent safety testing misses

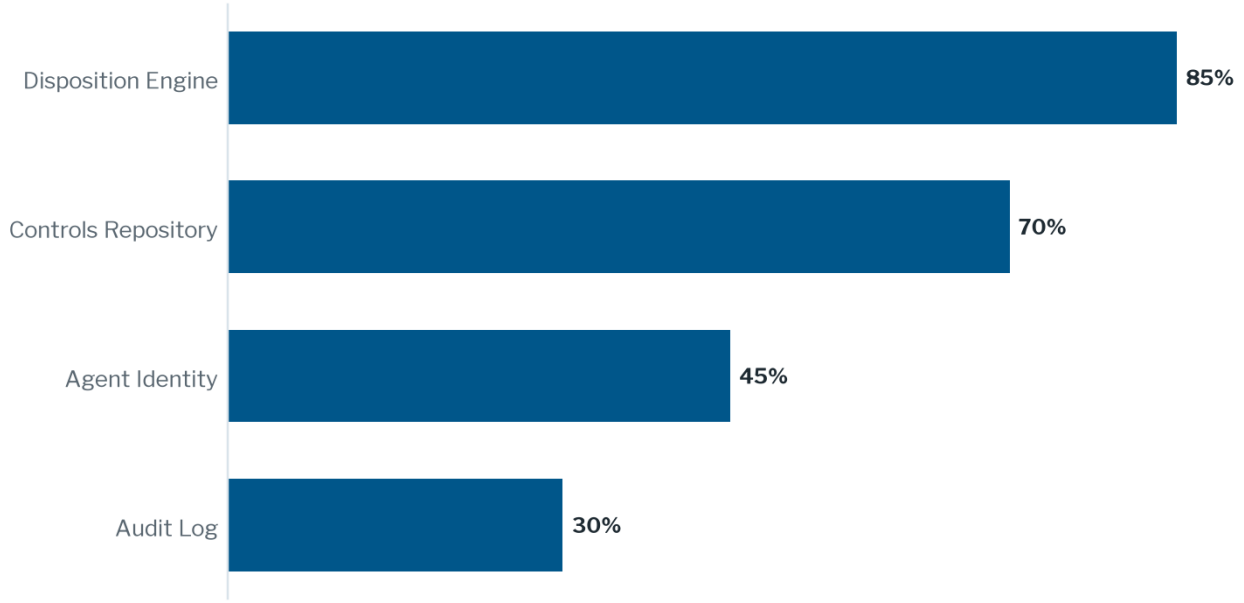


FACT Four related but non-identical measures from one paper, shown together because they point the same way. The third is a floor: the authors report more than 20 percent. Unsafe behaviour is scored by model judges, one validated at 91.5 percent accuracy, so the figures carry classifier error. Single study, not replicated. [12]

ASSESSMENT The practical consequence is that no red-teaming vendor can honestly tell a buyer what share of the tool-call risk space their suite covers, because the enumeration needed to compute that share is itself a research result from this month. That is not a reason to skip automated red-teaming, which finds real problems cheaply and is the only way to test at the rate agents change. It is a reason to reject the phrase "validated" in a procurement document and replace it with the number of cases run and the method used to generate them. [12]

ASSESSMENT MAS's pre-deployment expectation is compatible with this, which is worth noticing. Paragraph 4.18 asks for review "by parties not involved in its development", proportionate to risk materiality, and paragraph 4.15 says evaluation of AI agents should cover their key failure modes. Neither asks for a coverage percentage. An institution that runs an enumerated adversarial suite, documents the method, and states the residual uncertainty is closer to the intent than one that presents a passed benchmark as assurance. [1] [12]

FIGURE 4 How completely the shipped AWS stack covers SAFR's four runtime components



ASSESSMENT My scoring, not a vendor's and not a benchmark. The Disposition Engine maps almost exactly onto default-deny Cedar evaluation. The Audit Log scores low because what is retained is a policy decision record with a 24-hour trajectory, not an authenticated action trace with a retention schedule. Treat the numbers as ordering, not measurement. [2] [7] [9] [11]

ASSESSMENT Taken together, the answer to the second question is no, with one qualification that matters. Shipped tooling can emit a strong record of authorization decisions and can prevent some classes of bad sequence outright. It cannot currently emit an authenticated trace of what the agent did and why, retained for a period anyone has specified, spanning more than a day. The qualification is that the gap is closable with engineering rather than research: the pieces are components, not inventions, and any of the large platforms could ship this within a release cycle if a buyer with signature authority asked for it. [7] [8] [11] [2]

Drawing the automate-or-not line

The third question has the thinnest evidence base of the three, and the honest answer is mostly negative. What exists is a live programme with published criteria, a military registry effort with a stated purpose, and no published taxonomy anywhere.

FACT The UAE is running the largest publicly described exercise of this kind. A national agentic AI project aims to convert half of federal government operations, services and tasks to agentic models within two years. Cabinet approved a first service package in May 2026 alongside a training programme covering 80,000 federal employees. A June 2026 workshop drew more than 300 participants from 50 federal entities with a 90-day assessment goal, and an August workshop of more than 100 federal officials opened the strategic track under the National Committee for the Agentic AI Project. [13] [14]

FACT The published assessment criteria are operational. Participants sorted candidate work by volume and frequency of operations, annual transaction numbers, number and type of beneficiaries, procedure clarity and documentation, data currency and availability, existing automation levels, and expected impact on service quality, cost efficiency and customer satisfaction. Ten areas were named, including human resources, procurement, financial affairs, legislative affairs and internal audit. The governing principle is stated as "human leads, AI enables". [13]

ASSESSMENT Those are readiness criteria, and they are the criteria a competent process automation programme has used for twenty years: high volume, well documented, clean data, clear rules. They tell you which tasks an agent could do. They do not tell you which tasks an agent should be allowed to do, because not one of them asks whether the action is reversible, how wide the blast radius is if it goes wrong at scale, or whether the affected person can contest the outcome. [13]

ASSESSMENT Legislative affairs and internal audit appearing on a list selected by volume and procedure clarity is the clearest illustration. Both are highly documented and rule-governed, which scores well on every published criterion. Both also produce outputs whose legitimacy depends on who produced them, which no published criterion captures. This is not a prediction of failure; it is an observation that the screen is measuring the wrong axis for the hardest cases. [13]

FACT The classification framework itself has not been published. Trade coverage on 20 August 2026 identified the task classification frameworks as the document to watch and noted that no defensible published methodology yet separates autonomous from advisory tasks. [14]

FACT The United States Marine Corps is approaching the same problem from the inventory end. Maj. Christopher Clark, the Marine Corps AI lead, said participants in an announced hackathon will help create a registry of AI agents to monitor and audit their use: "Ideally with this registry we'll reduce the number of agents that are out there and track and understand who has what and who's using it." He described Marines building tools that can scrape their email, and aims to start using the registry by the end of 2026, noting more work is needed on identity management. [15]

ASSESSMENT The Marine Corps framing is the more realistic of the two, because it starts from the fact that agents are already deployed by people who did not ask. A registry that reduces the agent count and names owners is a prerequisite for any classification scheme, and classification without inventory is a policy about a population you cannot enumerate. The order of operations here is right even though the ambition is smaller. [15]

ASSESSMENT So the answer to the third question is that the boundary is being drawn on process readiness, not on decision risk, and that no published taxonomy exists in any of the jurisdictions examined. The nearest thing to one is SAFR's disposition calibration, which sorts actions by reversibility, financial materiality, customer impact, regulatory sensitivity and novelty. That is a risk taxonomy in everything but name, it comes from a banking consortium rather than a government, and it applies to actions rather than to jobs. [2] [13]

HYPOTHESIS The first published automate-or-not taxonomy will come out of the UAE programme rather than from a financial supervisor, and will be organised around service

categories rather than action risk. Confirmed if the National Committee publishes a classification framework before mid-2027 whose top-level split is by government function. Refuted if the first published framework is issued by a financial or safety regulator, or if its top-level split is by reversibility or human impact. [13] [14]

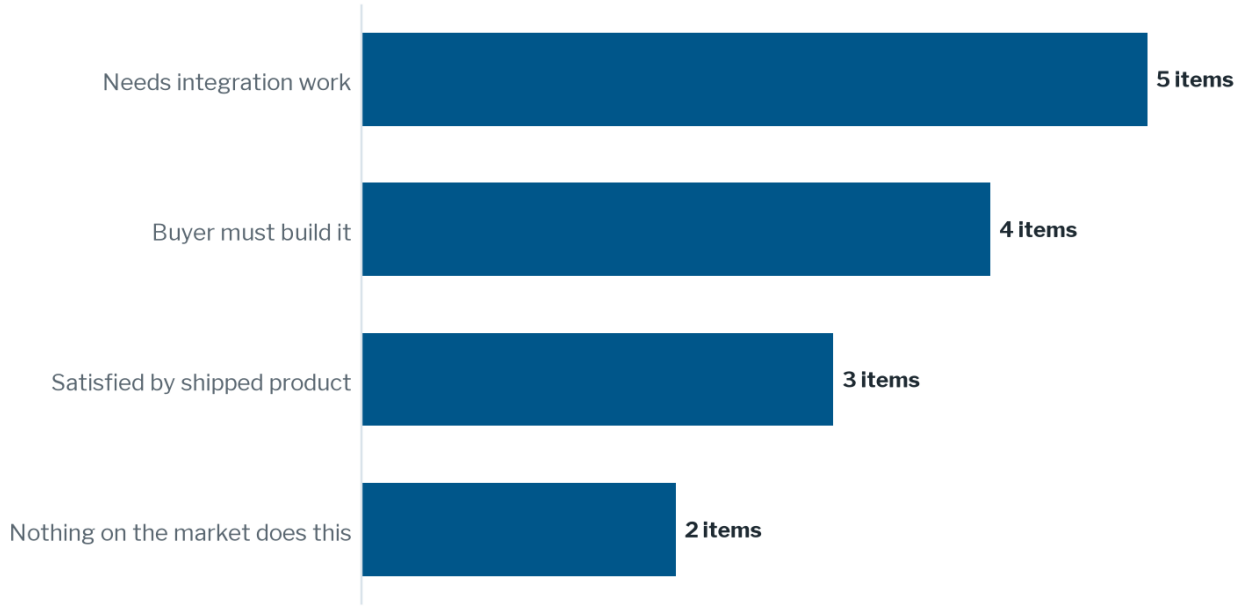
A control checklist for buyers

Fourteen things to make a vendor demonstrate rather than describe. Each maps to something a published regime asks for, and each is phrased as an observation you can make in a room rather than a capability you can be told about. The order is roughly by how often the demonstration fails.

#	Control	What to make them demonstrate, live
1	Authenticated action trace	Show that the record of what the agent did is signed or anchored outside the agent process. SAFR's own text says the trace and the action details can be fabricated together by a consistent injection. Ask who signs, and expect no good answer.
2	Stated retention period	Get a number of years in the contract, not in the documentation. Then ask what happens to records past the platform's own window and who pays to export them.
3	Reconstruction beyond one session	Ask for a full reconstruction of a business process that ran for three days across several sessions. The 24-hour trajectory window makes this a build, not a feature.
4	Authority record per action	For one executed action, show the machine-readable mandate that permitted it, in the form it had at the time of the action rather than the form it has now.
5	Bypass coverage	Have them call a tool via an SDK directly from agent code rather than through the gateway, and show you what the control plane recorded. AWS documents that this never crosses the policy boundary.
6	Silent-failure proof	Strip the policy engine read permission from the gateway role in LOG_ONLY mode and show that the system reports a fault rather than looking healthy.
7	Concurrency under load	Fire enough concurrent requests to breach an aggregate limit and show it holds. Then read the policy and confirm it sums requests, not responses.
8	Policy change during flight	Update a control while a workflow is running and show what happened to the in-flight history and to the next authorization decision.
9	Escalation with a real decision record	Show a held action, the named reviewer, the decision, the timestamp, and the timeout that would have fired had nobody answered.

#	Control	What to make them demonstrate, live
10	Reviewer capacity arithmetic	Divide expected escalations per day by reviewers available. SAFR warns that an escalation function generating more reviews than capacity produces rubber-stamping, which is worse than no review.
11	Registry completeness	Export the agent inventory with owner, purpose, approved scope, tool grants and decommission date. Then ask how an unregistered agent is detected.
12	Distinct agent identity	Show agent credentials sitting in a class your directory can distinguish from human accounts, with the shortest lifetime the workflow tolerates.
13	Test coverage, stated honestly	Ask what share of the tool-call risk space the pre-deployment suite covers. Accept "we do not know" and reject any number offered without a stated enumeration method.
14	Halt, timed	Pull the plug on a running agent including network access, with a stopwatch. The NCSC asks for immediate; find out what immediate means here.

FIGURE 5 **What satisfying each checklist item currently takes**



ASSESSMENT My classification of the fourteen items against the tooling examined here, which is AWS-heavy and therefore not a market-wide statement. The two in the last bar are the authenticated trace and an honest coverage figure for pre-deployment testing. [7] [8] [11] [12]

ASSESSMENT One recommendation, since the audience is mostly buying rather than building. Do not buy an agent control plane on its policy expressiveness this year. Policy expressiveness is the part vendors have solved and the part the guidance asks least about. Buy on what the platform retains, for how long, in what form, and with what integrity guarantee, and treat everything else

as table stakes. The vendor who answers the retention question with a number is ahead of the ones showing you a policy DSL. [7] [8] [2] [4]

Tripwires: where guidance becomes enforceable

Ten observable events, none requiring private information. They are ordered by how much each would change the picture, not by how likely each is. The point of a tripwire is that it can be watched by someone who is not paying attention to this category full time.

If this happens	It means	Where to watch
MAS issues the AI Risk Management Guidelines as final	The proposed 12-month transition clock starts, putting the first examination cycle in late 2027 for a guideline issued in 2026	mas.gov.sg consultations and guidelines pages
Any SAFR element is reissued inside a Notice rather than a Guideline	The per-action record moves from best practice to legally binding requirement for a class of institutions, with penalties attached	MAS Notices to banks and merchant banks
The NCSC replaces the 20 August post with formal guidance	UK public sector procurement and data protection arguments acquire a citable standard, which is how NCSC guidance becomes de facto mandatory	ncsc.gov.uk guidance and blog
The immutable-log language survives into that formal guidance	Log integrity becomes an assurance question in UK buying rather than an engineering preference	ncsc.gov.uk, and supplier assurance questionnaires
AWS raises the 24-hour trajectory window or ships trajectory export	The retention objection weakens materially and temporal policy becomes usable as evidence rather than only as enforcement	AgentCore release notes and what's-new feed
AWS opens Dogwood contributions or ships a production interpreter	The language is stabilising and third parties can build enforcement on it, which is when it stops being an AWS feature and starts being a standard	github.com/dogwood-policy/dogwood
Cedar's automated reasoning is extended to temporal conditions	Formal analysis returns and policy sets over sequences become provable rather than reviewable by hand	Cedar project releases, CNCF sandbox updates
The UAE publishes a task classification framework	The first public automate-or-not taxonomy exists, and given the absence of alternatives it will be copied well beyond the Gulf	UAE Cabinet and National Committee announcements
The USMC registry records tool grants and owners, not just agent names	Registries move from inventory to authority record, which is the step that makes them evidence	Marine Corps releases and defence trade press

If this happens	It means	Where to watch
A supervisor anywhere rejects an agent's own trace as insufficient evidence	The integrity problem acquires a price, and authenticated tracing moves from a design note to a purchase requirement	Enforcement notices in Singapore, the UK and the US

ASSESSMENT The second row is the one to watch hardest. Singapore has a clean mechanism for converting best practice into binding requirement, it has used it before in third-party risk management, and it now has an industry-developed specification sitting ready with named participants who helped write it. That is the shortest path from a voluntary framework to an examinable requirement anywhere in the set. [21] [2]

Scenarios to 2028

Point forecasts are worthless in a category where the two most important documents in this report were published fifteen days apart and one of them is explicitly interim. What follows is four structured futures with subjective probabilities. The probabilities are the least reliable content in this document. The earliest visible signs are the most useful, because they can be checked without me.

SCENARIO A Guidelines harden and tooling follows 40 percent

MAS finalises, the NCSC formalises, and platform vendors close the retention and trace gaps because regulated buyers make it a condition of purchase.

Consequence. Agent governance becomes a normal control domain with an audit programme, and the current crop of point products gets absorbed into cloud platforms.

Earliest visible sign. A major control plane publishes a retention schedule and a signed trace format in its release notes.

SCENARIO B The record stays voluntary 30 percent

MAS keeps SAFR as an industry reference, the NCSC's formal guidance stays advisory, and per-action evidence remains something responsible institutions do rather than something anyone checks.

Consequence. Two-tier market: banks and defence build the evidentiary layer themselves, everyone else runs agents on access logs and hopes. Vendor differentiation stays on policy features.

Earliest visible sign. MAS finalises the Guidelines with no per-action record requirement and no signal of a follow-on Notice.

SCENARIO C An incident forces the pace 20 percent

A public agent failure with financial or safety consequence occurs where the operator cannot reconstruct what happened, and the inability to explain becomes the story rather than the failure itself.

Consequence. Emergency guidance arrives inside months, written quickly and therefore badly, and compliance costs land on everyone including the operators who had done the work.

Earliest visible sign. Any enforcement or incident report whose findings centre on missing or untrustworthy agent logs.

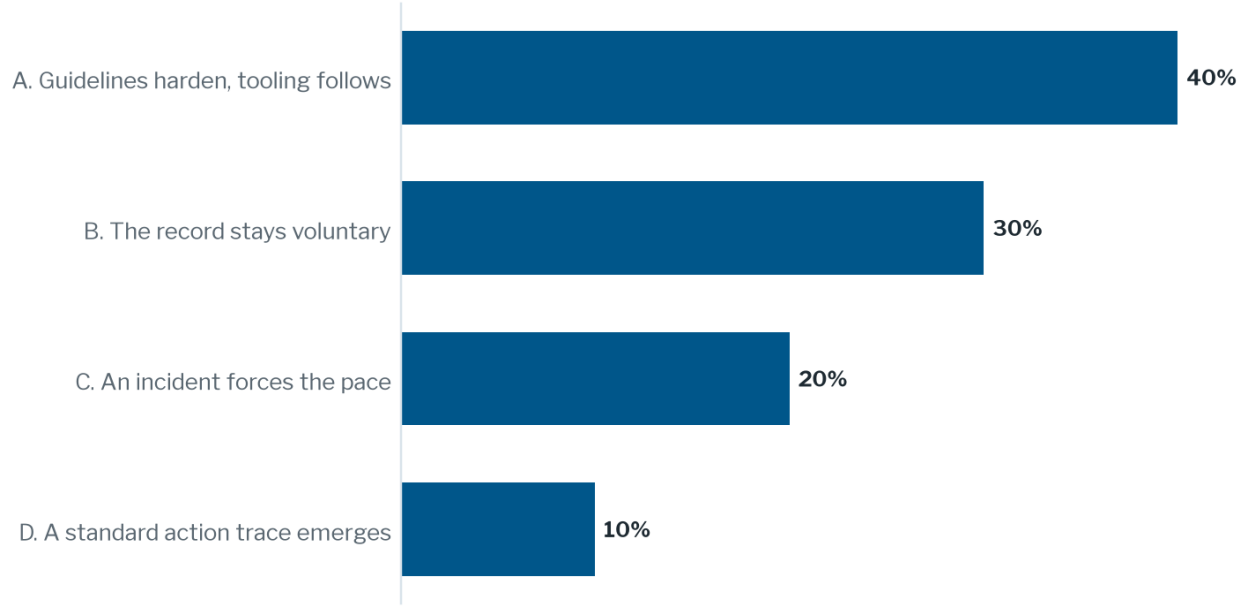
SCENARIO D A standard action trace emerges 10 percent

The trace format standardises through a protocol body rather than a regulator, most plausibly as an extension to MCP given its position at the tool-call boundary and its recent move toward header-level traffic visibility.

Consequence. The evidence problem becomes an interoperability win rather than a compliance cost, and portability between control planes becomes real.

Earliest visible sign. A trace, provenance or attestation extension appearing in an MCP specification draft.

FIGURE 6 Scenario probabilities to 2028



SCENARIO Subjective probabilities from the author, summing to 100 by construction. These are the least reliable content in the report and are included because refusing to quantify a judgment is its own kind of evasion. The tripwires above are the part worth acting on. [1] [4] [20]

Dated predictions

ASSESSMENT MAS issues the final Guidelines on AI Risk Management before 31 March 2027. Confidence: high. Resolution date 31 March 2027. [3]

ASSESSMENT The NCSC publishes formal agentic AI guidance superseding the 20 August 2026 post before 31 December 2027, and it retains explicit logging and immutability language. Confidence: moderate on the date, higher on the content. Resolution date 31 December 2027. [4]

ASSESSMENT No major agent control plane ships an authenticated, externally anchored action trace with a published retention schedule before 31 December 2027. Confidence: moderate. This is the prediction I would most like to be wrong about, and the one whose resolution is easiest to check. [7] [8]

ASSESSMENT The UAE does not publish a task classification taxonomy with a risk-based top-level split before 31 December 2027. Confidence: moderate to low, given how little visibility exists into the programme's internal documents. Resolution date 31 December 2027. [13] [14]

Implications

Four audiences with genuinely different problems, treated separately because blending them would produce advice that fits none of them.

For bank risk officers facing AI guidelines

ASSESSMENT Your instrument is a Guideline, which means no penalty for contravention and a direct effect on your supervisory risk assessment. Plan for a 12-month transition from issuance and work backwards. The two paragraphs to build against are 3.5, which lists the inventory attributes, and 4.18, which requires pre-deployment review by parties not involved in development. Both are achievable with existing model risk management machinery, extended rather than rebuilt. [1] [21]

ASSESSMENT The thing your current framework does not cover is the runtime record, because model risk management was built to validate a model before it goes live rather than to evidence decisions it makes in production. SAFR names that gap directly and your examiner will eventually ask about it. Read SAFR as the design specification your supervisor is already familiar with, and if your institution was among its named contributors, expect to be held to it sooner than the market is. [2] [3]

ASSESSMENT One practical instruction. Before the Guidelines are final, run the inventory attribute list from paragraph 3.5 against whatever you currently record for agents, and count the missing fields. That gap, measured in fields rather than in adjectives, is the most useful number you can put in front of a board this quarter. [1]

For public sector AI governance leads

ASSESSMENT The UAE and Marine Corps efforts bracket the realistic options. If you have the mandate to run a top-down classification programme, the UAE criteria are a reasonable starting screen and need one addition: a reversibility and contestability test that sits above the readiness criteria and can veto a task that scores well on all of them. Without it you will automate the well-documented decisions that citizens are most likely to appeal. [13]

ASSESSMENT If you do not have that mandate, and most do not, do the Marine Corps thing first. A registry that names owners and reduces the population is worth more than a classification scheme applied to agents you cannot enumerate. Maj. Clark's framing about Marines building tools that scrape their own email is the actual starting condition in most large organisations, whatever the policy says. [15]

ASSESSMENT Both routes need the same thing the vendors do not yet supply: a record that survives the agent and the session. Public sector records schedules are typically measured in years, and a 24-hour trajectory window is not a rounding error against that, it is a category difference. Put retention in the requirement document, not the evaluation criteria, so a non-compliant bid is non-responsive rather than merely lower scoring. [7]

For security architects buying agent control planes

ASSESSMENT The enforcement layer is in better shape than the evidence layer, and your procurement is probably weighted the wrong way as a result. Default deny at a gateway with Cedar policies is a solved problem you can buy today; sequence constraints over an agent session are newly solved and worth having. Neither answers the question you will be asked after an incident, which is what the agent did and on what basis. [9] [7]

ASSESSMENT Three architectural decisions follow. Route every tool call through the enforcement point, because anything the agent invokes directly from its own code is invisible to policy. Write aggregate limits over request events rather than response events, because the response-event version is defeated by concurrency and the two policies differ by one word. And treat the agent's self-reported trace as untrusted input until something outside the agent signs it. [11] [6] [2]

ASSESSMENT On zero standing privilege: adopt it, and do not let it be sold to you as an answer to the audit question. Ephemeral scoped credentials shrink the blast radius and give you an identity to attribute actions to. They do not produce the action trace, and a vendor conflating the two is telling you something about their roadmap. [19]

For internal auditors evidencing agent decisions

ASSESSMENT Your position is the strongest and least comfortable in this report. The strongest, because the standard audit question is exactly the one the tooling cannot answer, and you will be the first to find that out empirically. The least comfortable, because you will find it out during a walkthrough rather than in a planning meeting. [7] [8]

ASSESSMENT Design the test before the control exists. Pick one completed multi-step agent workflow, from more than a week ago, and attempt to reconstruct it end to end from retained records without asking the engineering team. What you can and cannot rebuild is your finding, and it is more persuasive than any control design assessment because it is a demonstration rather than an opinion. [7]

ASSESSMENT Two specific things to test that most audit programmes will miss. Whether the authorization control was actually evaluating, given that a permission gap can make log-only mode fail silently. And whether an approval at one workflow step is being treated as authority for subsequent steps, which SAFR explicitly prohibits and which is how most agent frameworks are built by default. [11] [2]

What this document cannot answer

Five questions that need primary research rather than more reading, listed because the gaps are load-bearing and hiding them would undermine the rest.

What supervisors are actually asking for in examinations. Published guidance and examination practice are different documents, and the second one is not written down. Answering this needs conversations with people who have sat through an AI-focused examination in Singapore or the UK in the last six months.

Whether the UAE task classification framework exists in a form I could not reach. My search was in English and covered public sources. A framework circulating as an internal Cabinet document, or published in Arabic on a portal I did not find, would overturn the central claim of the third section.

What the other control planes retain. This analysis is AWS-heavy because AWS documented its limits publicly and in detail, which is a reason to credit AWS rather than a reason to single it out. Microsoft, Google and the independent agent security vendors may retain more, less, or the same, and I could not establish it from published material at the same level of specificity.

Whether anyone has actually reconstructed a multi-day agent workflow from retained records. Every conclusion about the practical effect of the 24-hour window rests on the documented limit rather than on an attempt. A handful of real reconstruction attempts, successful or failed, would be worth more than the whole of this section of the report.

What escalation volumes look like in production. SAFR's warning about escalation exceeding reviewer capacity is the most operationally consequential sentence in the framework, and there is no public data on the ratio at any institution running agents at scale.

Half-life of this assessment

Six months. The AWS platform limits, the availability status of temporal policies, the specific quota numbers, and the status of the NCSC's interim advice. Any of these can change in a release note or a blog post, and the concurrency and timestamp warnings in the Dogwood repository are the kind of thing a maturing project fixes. Check the release notes before quoting any number from the second section.

Twelve months. The tabulation of what each regime demands, which will need redoing once MAS finalises and the NCSC formalises. The scenario probabilities, which should be re-derived rather than adjusted. The claim that no published automate-or-not taxonomy exists, which is the single claim in this report most likely to be overtaken by a publication.

Twenty-four months. The assessment that the market prices policy expressiveness over evidentiary quality. This is a procurement culture observation and those move slowly, but a single enforcement action centred on missing agent records would move it fast.

Architectural, and unlikely to decay. That an agent's own account of its reasoning is untrustworthy as evidence unless something outside the agent attests to it. That authority granted for one step does not extend to the next in a workflow that adapts. That readiness criteria and risk criteria are different axes and a screen built on one will mis-sort on the other. These follow from the structure of the problem rather than from any current product or document, and I expect them to hold after everything else here is stale.

Limitations

I have no commercial relationship with AWS, Britive, or any vendor named here, no position in any listed company mentioned, and this report was not commissioned, reviewed, or seen by any of them before publication.

The evidence base is asymmetric by jurisdiction. Singapore is covered from two full primary documents totalling 55 pages. The UK is covered from two blog posts. The UAE is covered from workshop readouts and ministerial statements in English trade and national press, with no primary programme document available to me. The Marine Corps is covered from a single conference report of one official's remarks. Conclusions about Singapore are substantially better supported than conclusions about the other three, and the side-by-side table should be read with that imbalance in mind rather than as four columns of equal weight.

The vendor analysis is AWS-heavy for the reason given above, and a reader should not infer that AWS is behind its competitors on any of this. The opposite inference is at least as defensible: AWS is the vendor whose limits can be analysed because AWS published them.

Every count in the charts derived from the regime tabulation is my coding of prose written by other people, and coding prose involves judgment. I have tried to count hedged language as

present rather than absent, which biases the counts upward and makes the gap I am describing look smaller than a stricter reading would.

The red-teaming coverage figures come from one paper by one group, published this month, with unsafe behaviour scored by model judges rather than human review. It is the best available quantification of benchmark incompleteness and it is a single unreplicated result. I have used it to support a directional claim and would not defend the specific percentages.

Agent population forecasts circulating in trade coverage, attributed to Gartner and quoted widely, are excluded from the analysis. The reported figures reach me through secondary coverage rather than a readable methodology, and no conclusion here needs them. A category can be important without a number attached, and borrowing an unverifiable one would have weakened rather than strengthened the argument.

Sources

[1] Monetary Authority of Singapore. "Consultation Paper on Proposed Guidelines on Artificial Intelligence Risk Management for Financial Institutions." 13 November 2025. Consultation closed 31 January 2026. Paragraph references in this report are to the proposed Guidelines annexed to that paper. *Regulator primary* mas.gov.sg

[2] Monetary Authority of Singapore and industry contributors. "Safeguards for Agentic Finance at Runtime (SAFR)." White paper version 1.0, July 2026. Published under the BuildFin.ai initiative. *Regulator primary* mas.gov.sg

[3] Gan Kim Yong, Deputy Prime Minister and Chairman of MAS. "Written reply to Parliamentary Question on agentic AI in financial services." 5 August 2026. *Regulator primary* mas.gov.sg

[4] National Cyber Security Centre. Toby W, Principal Security Architect. "Managing the cyber risk of agentic AI." 20 August 2026. Described by the NCSC as interim advice pending formal guidance. *Government primary* ncsc.gov.uk

[5] National Cyber Security Centre. Martin R and Dr Kate S. "Thinking carefully before adopting agentic AI." 15 May 2026. *Government primary* ncsc.gov.uk

[6] Brooker, Marc; Tassarotti, Joseph; Tristan, Jean-Baptiste. "Introducing Dogwood: runtime verification for AI agents." AWS Open Source Blog, 6 August 2026. *Vendor primary* aws.amazon.com

[7] Amazon Web Services. "Securing AI agents with temporal policies in Amazon Bedrock AgentCore." AWS Machine Learning Blog, 6 August 2026. Source of the 24-hour trajectory window, the single concurrent authorization request per session, and the session invalidation behaviour. *Vendor primary* aws.amazon.com

[8] dogwood-policy/dogwood. Reference parser and interpreter for the Dogwood policy language. Apache-2.0. Source of the production-use warning, the unvalidated timestamps, and the in-memory engine limits. *Vendor primary* github.com

[9] Amazon Web Services. "Policy in Amazon Bedrock AgentCore is now generally available." 3 March 2026. *Vendor primary* aws.amazon.com

[10] Amazon Web Services. "Control agent behaviors and cost beyond a single action: new capabilities in Amazon Bedrock AgentCore." 6 August 2026. *Vendor primary* aws.amazon.com

[11] Konishi, Hidekazu. "Amazon Bedrock AgentCore Policy Implementation Guide: Cedar-Based Agent Authorization and Default-Deny Design." 2026. Independent practitioner documentation of evaluation order, quotas, logging behaviour and the LOG_ONLY failure mode. *Practitioner primary* hidekazu-konishi.com

[12] Chen, Xuan; Yan, Lu; Zhang, Ruqi; Zhang, Xiangyu (Purdue University). "Who Tests the Testers? Systematic Enumeration and Coverage Audit of LLM Agent Tool Call Safety." arXiv:2603.18245v2, 18 August 2026. Preprint, not peer reviewed at time of writing. *Academic preprint* arxiv.org

[13] Aletihad News Center. "UAE Government holds agentic AI workshop with participation of 50 federal entities." 10 June 2026. Source of the assessment criteria and the ten named areas. *National press* aletihad.ae

[14] AI News. "Agentic AI in government just hit the hard part: deciding what a machine may decide." 20 August 2026. *Independent press* artificialintelligence-news.com

[15] GovCIO Media and Research. "Marine Corps Eyes AI Agent Registry to Rein in Shadow AI." 14 August 2026. Reporting remarks by Maj. Christopher Clark at the Federal AI Forum. *Independent press* govciomedia.com

[16] Mascellino, Alessandro. "NCSC Urges Stronger Controls for Agentic AI Systems." Infosecurity Magazine, 20 August 2026. *Independent press* infosecurity-magazine.com

[17] QA Financial. "MAS puts agentic AI testing under scrutiny." 18 August 2026. *Trade press* qa-financial.com

[18] InfoQ. "AWS Open-Sources Dogwood, Extending Cedar to Govern Sequences of Agent Tool Calls." 16 August 2026. Source of the concurrent-transfer example defeating a $5,000 cap. *Trade press* infoq.com

[19] Poghosyan, Art (CEO and co-founder, Britive). "Agentic AI Is Outrunning the Access Model It Inherited." Cybersecurity Insiders, 20 August 2026. Vendor-authored opinion. Contains no data and cites no external source; used here only as a statement of the zero standing privilege position. *Vendor-authored opinion* cybersecurity-insiders.com

[20] Model Context Protocol. "The 2026-07-28 Specification." Introduces required Mcp-Method and Mcp-Name headers, giving gateways and proxies visibility of agent tool traffic without inspecting the body. *Standards body primary* modelcontextprotocol.io

[21] Monetary Authority of Singapore. "Supervisory Approach and Regulatory Instruments." Source of the distinction between Notices, which impose legally binding requirements, and Guidelines, which do not but affect MAS' overall risk assessment of an institution. *Regulator primary* mas.gov.sg

[22] Baker McKenzie Connect On Tech. "Singapore: MAS publishes consultation paper on proposed guidelines on AI risk management for financial institutions." 2025. Law firm summary, used only to cross-check the structure of the proposed Guidelines against the primary text. *Law firm commentary* bakermckenzie.com

Rights and permitted use