

Testing Emergency Suspension Across Buyer and Provider Agents

A Practical Rehearsal for Service Managers

Phil Hatch | Akholi

October 23, 2026

Service managers should test emergency suspension as a sequence of observable events. A contractual stop right is usable only when the command reaches every relevant action path and the firms can establish what has stopped.

Rehearse the Stop Before It Is Needed

The service-governance manager should schedule a joint suspension rehearsal with the provider for each material agentic workflow. The objective is to establish whether authorized staff can contain harmful activity and account for unfinished work. A successful button click or acknowledged ticket is only one event in that sequence.

The study distinguishes immediate stopping capability from later remediation and routine approval. Respondents described uncertainty over who could stop an agent and gaps in out-of-hours authority. These accounts justify testing the actual rota and escalation path, rather than assuming the provider's incident policy covers the buyer's need.¹

Agree the rehearsal's scope, safe environment, operational safeguards, and evidence before starting. Where production testing could affect customers or essential services, use a representative environment and identify what remains unproven. Managers should not claim production containment from a tabletop discussion alone.

¹ Phil Hatch, *Agentic AI Outsourcing at Tier 1 Financial Institutions: Full Study Report*, version 1.8.2 (Akholi, forthcoming manuscript dated November 1, 2026), Part 3, Co-Management, sec. 2.4, "Speed of Stopping a Misbehaving Agent," indicator C143, <https://doi.org/10.2139/ssrn.7164559>.

Define What Suspension Must Accomplish

A stop can mean preventing new tasks, withdrawing credentials, isolating a tool, cancelling queued work, or pausing an entire service. The joint plan should distinguish these effects. The appropriate action depends on the suspected harm and the consequences of interrupting work already in progress.

For a hypothetical outsourced account-update service, the immediate objective might be to prevent further writes while preserving read access for investigation. That choice differs from disconnecting the complete platform. Managers should define the intended safe state with system owners and record which actions cannot be interrupted once committed.

Measure time from recognition to authorized decision, command issuance, enforcement at the relevant systems, and confirmed cessation of prohibited activity. These intervals reveal different problems. A rapid command with slow enforcement requires a different correction from an unreachable decision-maker. Agree thresholds from exposure and service needs rather than borrow an arbitrary industry target.

Test the Paths Beyond the Main Agent

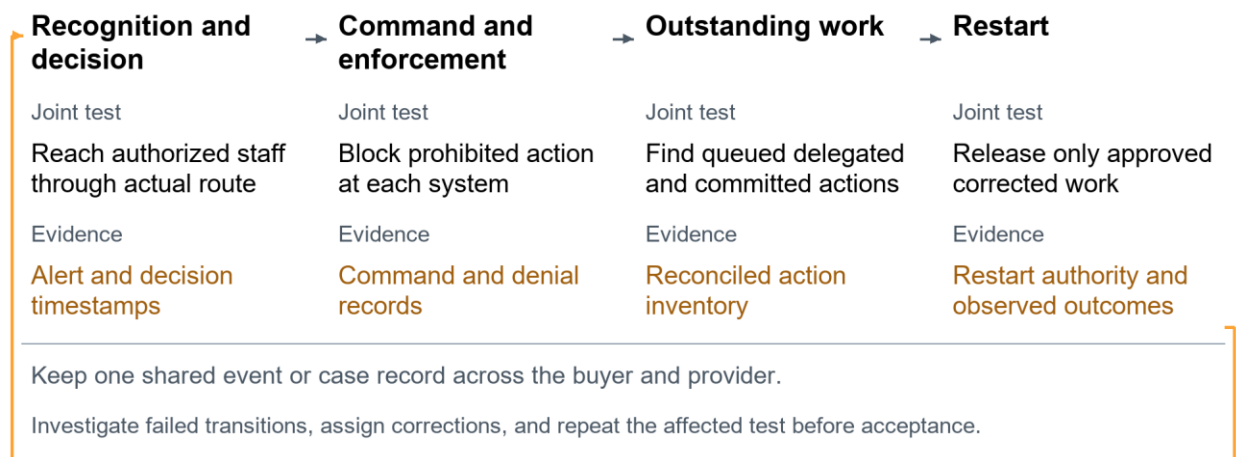
Figure 1 provides a rehearsal sequence. Include delegated agents, cached credentials, scheduled jobs, retries, and downstream systems. Suspending the provider's orchestrator may leave work executing elsewhere. The firms should reconcile outstanding action identifiers and verify the resulting state at the system where harm could occur.

The allied security guidance recommends containment, fail-safe defaults, and incident-response testing. It supports a rehearsal that examines the complete system rather than the model alone. It does not guarantee that a particular stop mechanism is safe for every workflow or suitable for physical processes.²

Use a clearly identified test event to distinguish pre-stop work from post-stop activity. Preserve the original timeline and state changes. If clocks differ, record and reconcile the difference rather than present a misleadingly precise elapsed time. The evidence should let both firms determine whether an action was authorized, queued, or newly initiated.

² Australian Signals Directorate's Australian Cyber Security Centre et al., *Careful Adoption of Agentic AI Services* (Australian Signals Directorate, May 1, 2026), "Manage third-party components," "Identity management," "Resilience," and "Human in the loop," <https://www.cyber.gov.au/publication/careful-adoption-of-agentic-ai-services>.

Figure 1. Measure the full suspension sequence



Author's proposed method; adapt the sequence and decision rights to the service.

Source: Phil Hatch, proposed framework developed for this paper.

Expose People and Communication Failures

Run at least one scenario through the actual out-of-hours contact route, with prior authorization for the exercise. The purpose is to test coverage, not surprise staff into unsafe action. Confirm that the reachable person has the authority, access, and information to make the required decision.

Test a provider contact who cannot be reached and a buyer contact who disputes the incident's severity. The escalation design should identify who can impose a temporary restriction while uncertainty is resolved. Managers should record whether a commercial approval or ticketing dependency delays containment of a defined threat.

The study's visibility concerns also affect the exercise. If each firm sees only its own tool status, neither can confirm complete suspension. The rehearsal should produce a joint event record linking the provider command to buyer-side enforcement and the final state of outstanding work.³

Keep Suspension and Restart as Separate Acceptance Tests

Stopping harmful action does not establish readiness to resume. The manager should preserve the incident evidence, identify affected cases, and agree how uncertain outcomes will be reconciled. Restarting an unresolved queue can repeat the original error or duplicate actions whose status was not known at suspension.

³ Hatch, *Agentic AI Outsourcing*, Part 2, sec. 5.1, "Agent Actions Visible Only Inside Each Firm's Own Tools."

If the exercise finds that the main agent stops but a delegated task continues, record the stop as incomplete for that scenario. Do not average the successful and failed paths into a reassuring overall score. The owner should identify the remaining authority, remove or constrain it safely, and rerun the scenario. Management can then accept a narrower demonstrated capability while the broader design is repaired, with the unproven path kept outside production authority.

The restart test should require an approved configuration, resolved exposure, verified permissions, and a controlled release of work. Use a limited initial batch and independently check results before normal operation. Define who can approve restart and who can veto it when their domain remains unsafe.

A cross-border service needs local escalation and notification analysis. Swiss, UK, EU, and US entities may have different incident obligations, even when the same operational suspension protects them. The joint incident record should feed those processes without assuming that a single provider message meets every legal duty.

Close Findings with a Repeated Demonstration

The rehearsal report should identify the tested configuration, observed intervals, residual actions, evidence gaps, and named corrective owners. A control owner should rerun the failed portion after remediation. Changing a procedure or updating a contact list is not enough to close a finding about whether suspension works.

The manager should accept the stop capability only within its demonstrated scope. Schedule a new rehearsal after changes to authority, tools, delegation, or continuity design. A tested suspension sequence gives the firms a common basis for containment while preserving a clear distinction between what was observed and what remains assumed.

Evidence and Source Note

The study is a nonrepresentative cross-section of 850 vendor delivery-team respondents on 150 engagements serving Tier 1 financial institutions, collected during January–June 2026. Neither banks nor vendors corroborated the accounts. It establishes reported experience, not causation, legal breach, or market prevalence. Its application here is a mechanism to investigate, not an estimate of conditions in other industries. The author-supplied version 1.8.2 is dated November 1, 2026 and is cited as a forthcoming manuscript consulted October 23, 2026.⁴

This is an analytical decision paper based on the supplied study and selected primary public sources checked through October 23, 2026. The original contribution is a joint suspension rehearsal with separate decision, enforcement, reconciliation, and restart tests. The proposed method has not been independently validated. Legal requirements depend on the entity, activity, jurisdiction, and effective provisions; supervisory observations, research, and professional commentary have distinct authority.

Bibliography

Australian Signals Directorate's Australian Cyber Security Centre et al.. *Careful Adoption of Agentic AI Services*. Australian Signals Directorate, May 1, 2026. <https://www.cyber.gov.au/publication/careful-adoption-of-agentic-ai-services>.

Hatch, Phil. *Agentic AI Outsourcing at Tier 1 Financial Institutions: Full Study Report*. Version 1.8.2. Akholi. Forthcoming manuscript dated November 1, 2026. Author-supplied manuscript consulted October 23, 2026. <https://doi.org/10.2139/ssrn.7164559>.

⁴ Hatch, *Agentic AI Outsourcing*, Introduction, “Executive Summary” and “Evidentiary Basis and Its Limits.”