PITHY NOTES PUBLICATIONS

PithySignal

AGENT GOVERNANCE SIGNAL

SIGNAL BRIEF | ISSUE 01

Agent Runtime Approvals Need a Governance Pattern, Not a Patch

A signal brief on why production agents need durable approval, escalation, and intervention patterns before enterprises scale high-impact workflows.

Created by Pithy Notes Publications | Published under Pithy Signal

KEY SECTIONS

- Executive Summary
- 1. Runtime Approval Is Becoming a Governance Control
- 2. The Broken Assumption: Human-in-the-Loop Is Not a Complete Control

Category: Agent Governance Signal
Publication Date: 2026-05-18
Issue: 01
Tags: Agent Governance | Runtime Approvals | Human Oversight | Workflow Control

This publication was created by Pithy Notes Publications and published as part of the Pithy Signal governance intelligence series.

Executive Summary

As AI agents move from experimental assistants into operational workflows, enterprises are beginning to face a practical governance problem: when should an agent be allowed to act on its own, and when should a human approve the action before execution?

Many organizations are approaching this as a workflow issue.

That is too narrow.

Runtime approvals are not merely interface prompts, confirmation buttons, or exception queues. They are a governance pattern for managing delegated authority in AI-enabled systems.

Agentic systems may summarize information, trigger workflows, call tools, access data, interact with APIs, create records, modify systems, or initiate downstream actions. Each of these activities may carry different levels of operational, legal, security, privacy, financial, or reputational risk.

The governance challenge is not whether humans should always be "in the loop."

The challenge is determining where human approval is required, what evidence must be captured, what authority the agent has been delegated, and how approval decisions are monitored over time.

Hyper TPRM programs will increasingly need to assess whether vendors can support runtime approval governance for AI-enabled and agentic functionality.

The question is shifting from:

"Does the vendor use AI?"

to:

"How does the vendor govern AI actions at runtime?"

1. Runtime Approval Is Becoming a Governance Control

Traditional governance models often rely on policies, contracts, access controls, and periodic evidence reviews.

Agentic AI introduces a new requirement: action-level oversight.

When agents can execute tasks, interact with systems, or initiate workflow steps, approval governance must move closer to runtime operations.

A runtime approval is not simply a user confirmation.

It is a control point that determines whether an AI system may proceed with a specific action under specific conditions.

Effective runtime approval governance should define:

- which actions require approval
- who may approve them
- what evidence must be presented
- what decision criteria apply
- how approvals are logged
- when escalation is required
- how exceptions are reviewed

Without this structure, organizations risk creating fragmented approval prompts that look like governance but do not operate as governance.

2. The Broken Assumption: Human-in-the-Loop Is Not a Complete Control

Many AI governance discussions rely on the phrase "human-in-the-loop."

The phrase is useful, but insufficient.

A human review step does not automatically create effective governance.

Organizations must ask:

- Is the human qualified to approve the action?
- Does the human understand the risk?
- Is the approval decision logged?
- Is the agent's recommendation explainable?
- Are approval thresholds defined?
- Are repeat approvals monitored for drift or rubber-stamping?
- Can approvals be audited later?

Without defined approval patterns, human oversight can become ceremonial.

Hyper TPRM requires more than asking whether a vendor has human review.

It requires understanding how human approval is operationalized.

3. Approval Thresholds Should Be Risk-Based

Not every agent action requires the same level of oversight.

A low-risk summarization task may not need human approval.

A high-risk action involving customer data, financial movement, contractual commitments, production changes, regulatory reporting, or external communications may require explicit approval before execution.

Runtime approval models should be based on:

- data sensitivity
- action reversibility
- financial impact
- legal or regulatory exposure
- customer impact
- security risk
- operational criticality
- degree of agent autonomy
- external system interaction

The governance pattern should distinguish between:

- allow automatically
- notify after execution
- require approval before execution
- escalate to specialist review
- block entirely

This moves approval governance from ad hoc intervention to structured delegated authority management.

4. Vendors Must Be Assessed for Runtime Governance Capability

As vendors embed AI agents into platforms, TPRM teams will need to evaluate more than model use or AI disclosures.

They will need to assess whether vendor platforms support runtime governance controls.

Key vendor questions include:

- Can agent actions be classified by risk?
- Can approval thresholds be configured?
- Can human approval be required before high-risk actions?
- Can approvals be logged and exported?
- Can approval decisions be tied to user identity?
- Can the system preserve evidence of agent reasoning or recommendation context?
- Can customers restrict agent permissions?
- Can agent actions be monitored over time?
- Can approval patterns be tested before deployment?
- Can exceptions be escalated and reviewed?

These questions move AI vendor oversight from policy attestation toward operational assurance.

5. Evidence Generation Is Central

Runtime approvals only become meaningful governance controls when they generate evidence.

An approval system should capture:

- the agent action requested
- the triggering condition
- the data or systems involved
- the recommendation or rationale presented
- the approving individual
- approval timestamp

- decision outcome
- escalation path
- resulting action
- exception handling

This evidence supports:

- audit readiness
- incident investigation
- regulatory inquiry response
- operational monitoring
- vendor assurance
- control testing

Without evidence, runtime approvals may reduce immediate risk but fail to support long-term governance accountability.

6. Hyper TPRM Implications

Runtime approval governance is becoming a core Hyper TPRM concern because agentic vendor systems may operate across changing workflows, identities, APIs, and business processes.

Traditional TPRM asks:

- Is the vendor secure?
- Does the vendor have controls?
- Has the vendor completed an assessment?

Hyper TPRM increasingly asks:

- What actions can the vendor's AI perform?
- What authority has the AI been delegated?
- Which actions require approval?
- How are approval decisions evidenced?
- Can the customer configure approval thresholds?

- Can runtime behavior be monitored continuously?

This is a structural shift in vendor oversight.

The focus moves from static control existence to runtime control operation.

7. Strategic Implications

Organizations should expect runtime approval governance to become a key buying criterion for enterprise AI platforms.

Platforms that support configurable approval thresholds, action-level logging, identity-aware governance, and evidence export will be better positioned for regulated and risk-sensitive environments.

Governance teams should prepare for increased convergence across:

- AI governance
- third-party risk management
- identity and access management
- security operations
- compliance evidence management
- procurement evaluation
- operational resilience

The organizations that define approval patterns early will be better prepared to scale agentic AI without relying on improvised controls.

8. What Organizations Should Do Next

Organizations should begin defining agent runtime approval patterns.

Recommended actions:

- classify agent actions by risk level
- define approval thresholds
- assign approval authority by action type
- require evidence capture for high-risk actions
- test approval workflows before production use

- monitor approval patterns over time
- review exceptions and overrides
- include runtime approval questions in AI vendor assessments
- require vendors to demonstrate approval logs and control evidence

The goal is not to slow every AI workflow.

The goal is to govern delegated authority with precision.

Closing Signal

Agent runtime approvals should not be treated as temporary workflow patches.

They are emerging governance patterns for managing autonomous action, delegated authority, and operational accountability.

As AI agents gain the ability to act across enterprise and vendor ecosystems, approval governance becomes a foundation for trust.

Signal Sources

- NIST AI Risk Management Framework
- ISO/IEC 42001 Artificial Intelligence Management Systems
- Cloud Security Alliance - Agentic AI Governance and AI Risk Management Guidance
- Microsoft - Agent Governance and Runtime Oversight Guidance
- Google Cloud Office of the CISO - Shadow Agents and AI Governance
- Palo Alto Networks - Agentic AI Governance
- OECD AI Principles
- EU AI Act Governance and Human Oversight Requirements

Created by Pithy Notes Publications Published under Pithy Signal

Governance intelligence publications focused on AI governance, risk management, compliance, and Hyper TPRM.