AI GOVERNANCE GAME · PRODUCTION SIGN-OFF ON AN AGENTIC DEPLOYMENT

# Logs are not a leash.

An AP agent has matched invoices and released payments under a threshold for four months, unattended. Error rates fell. Finance wants the threshold lifted. They bring their evidence: every release logged, a signed rollout, a prompt-review process, a dashboard green across the board. Then a payment clears to a supplier deactivated three weeks ago — a cached vendor record, acted on. No human saw it. **The dashboard is still green.**

OPTION A · LOWEST

## Trust the controls

Approvals, logs and the dashboard are working as designed. Approve with a monthly review.

**High risk. Auditable, not constrainable. The next stale record clears the same way.**

OPTION B · LOW

## Add more monitoring

A second dashboard, daily log review, an alert on any large payment.

**Elevated risk. You will know sooner. You still will not have stopped it.**

OPTION C · HIGHEST

## Ask one question first

If the agent cannot confirm the vendor's authority is current and scoped to this payment, does it stop on its own, or proceed and log it? Require a fail-closed boundary before scaling.

**Governable. Not a safety guarantee — the precondition for one.**

OPTION D · MIDDLE

## Pull the agent

Too risky for production. Go back to manual release.

**Risk avoided, value forfeited. Defensible if you cannot yet answer C.**

RUN THE FIVE-QUESTION TEST AGAINST THE ACTION, NOT THE ROLLOUT

**Who authorised this?** A named, current authority for *this payment* — not the pilot sign-off four months ago. **Was the authority current?** The cached record is the whole failure. **Was it scoped to this action?** Authority to release payments is not authority to pay a deactivated supplier. **Was the evidence sufficient?** It acted on a cache, not verified state. **Would it fail closed?** It failed open.

THE DISTINCTION THE WHOLE SCENARIO TURNS ON

01

**Logs are detective.** They tell you what happened, after it happened.

02

**An execution boundary is preventive.** It decides whether the action is allowed at all.

03

Finance brought five forms of observability and **zero enforcement**.

> A SIEM full of alerts is not a firewall.
> A monitored system is not a controlled one.

**Source:** AI Governance Game scenario "Logs Are Not a Leash", filed 21 June 2026 — agentic accountability, observability versus enforcement. Scenario is illustrative and composite, not a specific incident. Options are scored on governance quality rather than on caution, which is why pulling the agent scores in the middle and not at the top. Play the full branching version at fredriklindstrom.info/governance-game/. Status labels and column assignments above are this sheet's reading of the cited source, not the source's own framing.


FOLLOW FREDRIK LINDSTROM FOR MORE

## Fredrik Lindstrom

AI and cybersecurity governance for boards, CISOs and senior leaders. 25+ years in enterprise security.

*A practitioner to guide you.*

ARTICLES & BOARD ADVISORY
**fredriklindstrom.info**
Long-form governance analysis and toolkits

LINKEDIN
**in/fredriklindstrom**
Daily commentary and article launches

YOUTUBE
**@Promise-and-Risk-AI**
AI education without the hype or fear

COMPANION SITE
**promiseandrisk.ai**
Episode notes, sources and explainers

NEWSLETTER
**The Governance Memo**
Monthly, board-facing. Subscribe at fredriklindstrom.info