

Nobody signed for the outcome.

A mid-size insurer puts an agentic system on the simpler tier of quote requests — read it, price it, issue it, no human in the loop. It clears about 70% of volume and the economics look excellent for two quarters. Then an upstream feed changes format and the agent issues a run of underpriced quotes. **They are binding.** Nobody saw them. The loss lands in the next operating review, and the board asks who owns it. The vendor says the model performed to spec on the inputs it received. IT says uptime and error rates were green. The business owner says they signed off on the use case, not the data plumbing. **Every answer is true, and none of them is an owner.**

OPTION A · PARTIAL

Tighten the model

Retrain on the new format, add validation on the input feed, redeploy. The technical answer, and a necessary one.

Necessary, insufficient — and the trap, because it looks finished. The agent got more accurate; nobody got more accountable. The next drift arrives to the same silence.

OPTION B · HIGHEST

Assign the outcome first

Before any retraining: name one accountable owner for the agent’s decisions, define the escalation path for input drift, and set the human gate on binding actions. Then fix the feed.

Governable. The feed will break again in a way nobody predicted — this is the only option that decides in advance who answers when it does.

OPTION C · LOW

Pull the agent

Revert the whole tier to human processing until confidence returns.

The over-correction. It trades away 70% of the volume to avoid the accountability question rather than answer it — and the question comes back with the next deployment.

OPTION D · LOWEST

Send it to the vendor

The model mispriced it. Pursue the contract, the indemnity and the service credits.

The quotes were issued in your name and they bind you, not the vendor. You cannot contract out an outcome you already shipped. Worth doing in parallel; worthless as the answer.

THE FORK WAS NEVER WHETHER TO RETRAIN

It was whether **accountability comes before accuracy**. Every option here except one treats the incident as a defect to be repaired — in the model, in the deployment, or in the contract. The defect was real and none of those repairs touch what actually failed: a system was given the authority to bind the company, and the authority to answer for it was never assigned to anyone. **The more autonomous the system, the larger the human design around it has to be.** Decision rights and escalation paths are not the paperwork around the deployment. On the day the feed changed, they were the missing control.

THREE THINGS THE UNOWNED OUTCOME DECIDES

01

Monitoring was green and the loss still happened.

Uptime and error rates measure whether the system ran, not whether it was right. Neither would have caught a valid-looking quote at the wrong price.

02

Binding actions are a different class from

recommendations. The gap between advising a price and issuing one is the gap between an error and a liability — and it is the line a human gate belongs on.

03

Ownership is assigned before the incident or

discovered during it. Four parties answered the board honestly and none of them was the owner. That was decided at design time, by omission.

The agent performed to spec.
The spec never said who answers when the inputs move.

Source: AI Governance Game scenario “The Outcome Nobody Owned”, the agent-execution-boundary family. The underlying dilemma — when agents execute work, who owns the outcome — is item 6 of BCG’s *10-20-70* “70% Playbook”, verified 29 June 2026. The scenario is **illustrative and composite**, not a specific incident. **Option D is this sheet’s addition** and does not appear in the filed scenario; it is included because pursuing the vendor is the most common real response and it fails in a way worth showing. Options are scored on governance quality rather than on caution, which is why pulling the agent scores low rather than safe. **Maps to:** NIST AI RMF GOVERN (accountability structures, roles and responsibilities) and MANAGE (escalation and response); ISO/IEC 42001 roles, responsibilities and authorities, and operational controls over autonomous decisions. **EU AI Act:** the Article 4 AI-literacy duty has been in force since 2 February 2025, and any use that proves to be Annex III high-risk carries stand-alone obligations from 2 December 2027 — context for the scenario, not a deadline driving it. Play the full branching version at fredriklindstrom.info/governance-game/. Status labels and option scoring 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