ENTERPRISE INVESTMENT

How VeriProof Helps Protect Your AI Investment Through the Trough of Disillusionment

When the hype fades, the agents that matter are the ones that keep doing useful work under controls you can verify.

Document ID VP-INV-2026-01 | November 2026

For business owners, finance leaders, and AI program sponsors

Contents

01 Keeping useful work when expectations fall

02 Checking policy before action

03 Choosing tasks with clear limits

04 Staffing exception review

05 Evidence for compliance and verification

06 Preserving integration work

07 Measuring business value

08 Deciding which workflows to keep

AI investment is easy to defend when every demonstration looks extraordinary. It becomes harder when an agent encounters incomplete information, an unfamiliar exception, or a business action it should not take. Suddenly, the conversation shifts from what the technology might achieve to whether anyone can trust it with real work. Promising work stalls when leaders cannot tell which tasks they can safely delegate.

Gartner's 2026 Hype Cycle for Agentic AI and its implications for buyers

Topic in Gartner's research	Examples	What buyers can evaluate
Agent governance	Policies, oversight, and accountability	Who can authorize an exception before the agent acts?
Agentic AI	Agents that plan and carry out tasks	Which tasks finish correctly, and at what cost?

Gartner uses the trough of disillusionment to describe the drop in expectations after early experiments disappoint. For a funding decision, the practical question is which workflows still deliver enough value to keep running.

For an investment review, separate routine work from exceptions. A refund agent may save money on straightforward requests even when preparing unusual cases creates more work for a manager.

The Real Risk Is Losing Useful Work Along With the Hype—and Hitting a Fleet Bottleneck

A pilot can demonstrate a valuable capability without demonstrating a reliable business process. An agent may summarize accurately most of the time yet mishandle an exceptional request. It may prepare a sensible update but lack authority to commit it. If these distinctions remain invisible, leaders face an uncomfortable choice: accept poorly understood exposure or shut down automation that was genuinely helping.

A two-person governance team can follow one pilot through its logs and approval messages. As more agents come online, each separate queue adds work. Shared rules and review queues let the team spend more time on the exceptions themselves. The best guide to how many agents it can support is the number of exceptions and the time people need to resolve them.

Price each part of the workflow separately. If an agent can collect the evidence for a decision but lacks authority to make it, the value of that preparation depends on how much reviewer time it saves.

Put Controls Where Actions Become Consequences (Enforce)

For a refund, the consequential step is the call that sends money. Checking the amount before that call lets the sponsor choose a spending limit and see which requests would need a manager.

In Shadow mode, teams see which requests a proposed rule would allow, change, or send for review. That gives the business owner a way to tune the rule before enforcement.

Keep Agents Working on Bounded, Valuable Tasks

A customer service agent can check eligibility, assemble the order history, and prepare a refund. A retailer can fund those tasks together or approve them separately as it learns what each costs.

Other bounded tasks might include assembling internal evidence packs, identifying incomplete submissions, or drafting updates for authorized staff. Each needs its own quality tests and permissions. The investment thesis is that a portfolio of modest, repeatable tasks can be more resilient than a single ambitious promise of end-to-end autonomy. When enthusiasm cools, work that demonstrably saves effort has a stronger reason to remain funded.

Make Human Oversight an Operating Process Inside Existing Tools (Coordinate)

“Human in the loop” is not enough if nobody knows which human, what authority they hold, or what happens while a decision waits. An effective exception process needs a clear owner, sufficient context, and an explicit resolution. Otherwise, oversight becomes either a rubber stamp or a queue that consumes the productivity automation was meant to create.

Authorized reviewers can work in the VeriProof inbox or through connected Slack and Microsoft Teams workflows. Each decision stays linked to its request.

Count both the time a reviewer spends on a case and the time the customer spends waiting. A queue can look inexpensive in staff hours and still delay the refunds customers care about.

Evidence for compliance reviews and independent verification

During the trough, champions cannot rely on impressive demos or assurances that the model is improving. They need to answer concrete questions: What did the agent attempt? Which rule applied? Who reviewed the exception? What happened afterward? Without that evidence, a small incident can trigger broad suspicion because nobody can establish its scope or explain whether the control process worked.

Two kinds of evidence help teams answer those questions.

- Framework mappings connect decision evidence to the controls a compliance team reviews. In Vanta, governance tasks sync hourly. Teams can send selected evidence reports as drafts and schedule future deliveries after the first approved send.
- A signed evidence pack brings the relevant rule, review history, and reported outcome together. An auditor can check the files against the signed inventory and confirm the ledger anchor.

The signature protects the record's integrity. The business owner checks the quality of the underlying decision against the source system.

Protect the Integration Investment, Not Just the Model

An organization's AI investment includes more than model subscriptions. It includes application connections, redesigned workflows, test cases, staff knowledge, and the effort required to make governance acceptable. Disappointment with one model should not automatically invalidate all of that work. Where controls and evidence are consistently implemented, teams have a better basis for evaluating replacements without treating every change as an entirely new experiment.

VeriProof publishes integration paths for SDKs, agent frameworks, Copilot Studio, MCP, and browser workflows. When changing models or tools, retest the points where the application checks policy and reports the result.

Measure the Value That Survives Scrutiny

Use Shadow mode to see how a proposed rule would handle current traffic. Measure completion time, review effort, errors, and cost against the existing process once the workflow runs. A useful comparison counts correctly completed tasks.

Start with the shared VeriProof calculator's assumptions: 250 reviews a month, 18 minutes per case today, and six minutes with a prepared case and a named reviewer. The difference is 50 staff hours a month. At the model's loaded reviewer rate of $58.82 an hour, that time is worth $2,941 if the team puts all of it to use.

The Way Through Is Productive Control

Consider a refund pilot with those 250 monthly reviews. In this illustration, 200 standard cases take six minutes each with the agent's help. The other 50 need specialist judgment and still take 18 minutes. Only the standard cases release time: 200 × 12 minutes gives 40 hours a month.

The sponsor funds assistance for the 200 standard cases and retains the existing specialist review. At the calculator's starting assumption of 50% productive use, 20 of the 40 hours count toward the business case, worth about $1,176 a month before costs. The sponsor compares that value with the cost of running the agent and checks that refunds still settle correctly.

Try one workflow with your team

Choose a workflow with a known manual cost. During the trial, track correctly completed tasks, reviewer time, and operating cost. Use the comparison to decide which cases to automate.

Start your free 60-day trial

veriproof.app/start