

The Hidden Cost of Unoperated AI

AI agents drift. Knowledge gaps accumulate. Routing breaks on untested edge cases. Unoperated AI is slow, invisible decay nobody measures.



L5 Team

September 19, 2025

Nobody writes a post-mortem for a slowly declining deflection rate. There is no incident when an AI agent's routing accuracy drops from 94% to 81% over four months. No alert fires when the knowledge base article that used to resolve password reset tickets becomes outdated after an identity provider migration. The degradation is silent. The cost accumulates invisibly.

This is the hidden cost of unoperated AI: not a sudden failure, but a quiet drift that compounds until someone notices the ticket queue is larger than it was last year and cannot explain why.

How agents drift

AI agents in an ITSM environment operate on three inputs: routing logic, knowledge, and escalation thresholds. All three exist in an environment that changes constantly.

Teams restructure. Assignment groups change. New products are deployed. Old integrations are retired. Processes that were documented at go-live evolve in the hands of the people running them. New hire classes bring different types of tickets that were not part of the original configuration.

The agent does not know any of this happened. It continues routing based on the logic that was correct six months ago. It continues surfacing knowledge articles that were accurate before the migration. It continues escalating based on thresholds that made sense for a team that has since been reorganized.

"You built an agent and walked away. What do you think is happening to the prompts that the agent is doing? Who is watching it? Somebody needs to monitor it, somebody needs to govern it."

Pulkit Keshar, COO, L5

- **6 mo** Typical time before AI agent performance visibly degrades without an active operating cadence.
- **30%** Knowledge base articles that become materially outdated within 12 months without active maintenance.^[1]
- **Zero** Organizations that measure AI agent drift weekly without a dedicated operations function.^[3]

What drift looks like in practice

The signs of agent drift are not dramatic. They are the kind of thing that gets attributed to other causes: seasonal volume, a product launch, team turnover, before the pattern becomes undeniable.

- Ticket deflection rates that were 65% at go-live are now 44%, three quarters later
- Agents who trusted the AI routing six months ago are now manually re-routing a significant fraction of what the agent sends them
- Employees who used the self-service portal at launch have started emailing IT directly because the bot "never helps"
- First-contact resolution rates have declined but nobody has correlated this to a specific agent failure
- A knowledge article that generates high deflection traffic is 14 months out of date and the process it describes no longer exists^[1]

Each of these is individually explainable. Together, they are the signature of an unoperated AI environment: one where the agents were deployed at go-live and have been running on their original configuration ever since, accumulating drift in every direction as the world around them changed.

The governance gap

Most organizations that deploy AI in their ITSM platform do not have a function responsible for AI governance. The platform team manages the platform. The IT operations team manages incidents. The knowledge management team, if it exists, manages articles. Nobody owns the performance of the AI layer as a whole.

"It is not just AI by itself. It is a tag team. There is a caring and feeding to this AI world that people do not understand fully. You have to have a governance model and you have to have checks and balances, audits and controls."

Ron Mechling, CRO, L5

The Governance Gap

The governance gap is the space between deploying AI features and operating AI outcomes. Most ITSM deployments fall into this gap. They have agents that are technically active but are not monitored, tuned, or held accountable for what they produce. Closing the governance gap requires a weekly operating cadence with defined metrics and a team responsible for acting on what those metrics show.

The Drive as the antidote

L5 closes the governance gap through the Drive. Every week is a structured work cycle with a defined output: an agent tuned, a knowledge gap closed, a routing rule updated, a new edge case handled. The Drive does not wait for problems to surface in ticket volume data. It reviews agent performance proactively and takes action before the drift becomes visible to end users.

A typical Operate Drive on a Zendesk environment includes:

- **Deflection rate review:** which categories are declining and why
- **Routing accuracy audit:** what the agent sent versus where it should have gone
- **Knowledge base scan:** articles with high traffic that have not been reviewed in 60+ days
- **Escalation pattern analysis:** are humans overriding the agent in consistent ways that indicate a logic gap
- **New issue type detection:** ticket categories emerging in the last week that the agent has not seen before

Each of these produces an action. The action is the **DRIVE output**: a concrete change to the agent configuration that is live in production by the end of the week. The following week's Drive measures whether it worked.

An MSP waits for your ticket. L5 already ran the Drive. The difference between reactive support and active operation is the cadence. Weekly Drives mean drift is caught in days, not discovered when a leader asks why ticket volume is up 30% year over year.

What unoperated AI actually costs

The cost of unoperated AI is not usually calculated because nobody is tracking the right metrics. But the components are real.

Every deflection point lost to agent drift is a ticket that a human agent handles instead. At a fully-loaded cost of \$15 to \$25 per ticket,^[2] a deflection rate decline from 65% to 44% on 10,000 monthly tickets is \$31,500 per month in unnecessary labor cost. Misrouted tickets waste agent time and introduce resolution delay. Employees who stop trusting the self-service portal generate direct contact volume that bypasses deflection entirely. Compliance gaps that persist because the monitoring agent is surfacing false positives nobody acts on represent audit and security risk.

None of these costs appear on a line item that says "unoperated AI." They show up as labor, escalations, SLA misses, and employee satisfaction scores. Which is exactly why they go unaddressed: because the cause is invisible to everyone not watching the right metrics on a weekly cadence.^[4]

The hidden cost of unoperated AI is not hidden because it does not exist. It is hidden because nobody is looking for it. **Zendesk's Resolution Learning Loop** is where that looking happens on a Zendesk platform, provided someone runs the cycle.

Sources

1. Gartner. "IT Knowledge Management and Self-Service Optimization." Content decay research shows 25-40% of knowledge articles become materially outdated within 12 months without active maintenance.
2. HDI (Help Desk Institute). "The Cost of a Service Desk Ticket." Fully-loaded per-ticket cost benchmarks for human-handled IT service desk contacts, 2024.
3. L5. *Modern ITSM for Mid-Market Organizations* (White Paper). 2026.
4. L5 internal delivery data. Deflection rate measurement across Operate subscription engagements, 2025.