PULSEMETER

Techstrong Group
a Futurum company

AI in Observability: The Context Gap Limiting Trust and Action

by Guy Currier, Analyst, Futurum Group

SPONSORED BY

NeuBird AI

The promise of AI in IT operations has long centered on speed: faster detection, faster root cause analysis, faster resolution. But as AI capabilities have matured and enterprise observability environments have grown more complex, a significant gap between what AI is asked to do and what it knows about the operational environment has been revealed as a fundamental challenge.

Incident Response: The Context Assembly Problem

Incident response has always been a race between signal volume and human comprehension. In the past several years, the raw material of incident response—logs, metrics, traces, alerts—has scaled massively, to the point of obstruction. Engineers struggle to connect these signals into a coherent picture of what is broken, what caused it, and what the appropriate remediation path looks like.

The pain points that observability practitioners report form a coherent pattern. Alert noise and slow root cause analysis represent the input and output sides of the same underlying issue: too many signals entering the system without sufficient structure to make them actionable, and too much time required to translate those signals into an understanding of what happened. The consequences are significant, and include ineffective automation, inadequate cross-tool visibility, and exhausting manual correlation. But they are not independent problems so much as cascading effects of operating without a unified, contextually rich view of the environment.

Research Insights

Techstrong Research recently polled our community of IT and security professionals to understand their perspectives about the current state of AI adoption in observability and incident response. The results paint a picture of enterprises that have moved past AI experimentation but have not yet arrived at AI confidence.

Key Takeaways

- Incident response must address a context assembly problem: teams are not failing to collect signals; they are failing to connect signals quickly enough to act.
- The vast majority of organizations using or evaluating AI in observability report that it lacks the cross-system context required to move reliably from investigation support to trusted operational action.
- AI usage in observability drops sharply as organizations move from letting AI investigate to letting AI take action.
- Human oversight is the preferred model in AI-driven remediation.
- Production risk is distributed across context gaps—in configuration, infrastructure, and system visibility—and those gaps can keep teams locked in reactive operations.

Analyst View

The challenges practitioners report in incident response are not isolated frustrations; they're symptoms of an operational environment that generates more signal than it can contextualize. The top pain points cluster around alert noise, slow root cause analysis, and manual correlation, describing an environment that has scaled beyond humans' capacity to assemble context at the speed incidents demand. That is exactly the problem AI is supposed to solve. And the rest of the data explains precisely why it has not yet done so at scale.

A substantial share of respondents has introduced AI into their observability workflows, but the distribution of use cases exposes a critical constraint: the further along the remediation chain AI is asked to operate, the fewer organizations trust it to do so. Investigation assistance is the most common use case. Automated triage is far less common. Autonomous remediation remains rare. This pattern reflects a rational, calibrated response to a specific limitation. Teams allow AI to look because looking carries minimal risk. They withhold authorization to act because acting in a production environment carries consequences that demand a level of reliability AI has not yet demonstrated for most organizations.

Only 12% of respondents report that the AI tools they have used or evaluated carry rich, cross-system context. Context is critical in AI—an obvious enough statement; given that the majority report partial context, isolated tool context, or context so generic as to be operationally irrelevant, AI adoption limitation can be said to stem from limitations in context. This is an information architecture problem. AI operating on siloed, incomplete, or stale context cannot reliably distinguish between a deployment-induced regression and a capacity event, or between a cascading failure and a coincident alert storm. In environments where the cost of incorrect remediation can exceed the cost of the incident itself, that is disqualifying.

The dominant preference expressed by respondents to our survey, AI remediating with human approval, reflects a sophisticated understanding of where AI adds value today and what conditions would need to change for that boundary to expand. Respondents favor AI that either recommends actions or executes them with approval, reinforcing the preference for explainable, auditable, human-in-the-loop (HITL) automation. Organizations are saying: show me what you know, show me why you think what you think, and let me validate before you act.

Configuration drift, infrastructure issues, and lack of system visibility are the three leading risk categories respondents report, and they are all forms of context loss. They represent gaps in the operational picture: things that changed without adequate tracking, dependencies that behaved unexpectedly, system states that were unknown or unknowable from available tooling. The persistence of reactive operations as the dominant mode for a meaningful share of teams can be attributed, in no small part, to how much they are operating without sufficient context to prevent incidents. AI-assisted observability that does not address the underlying context deficit will improve response times, but not nearly as much as AI that genuinely assembles and reasons across the full operational environment has the potential to.

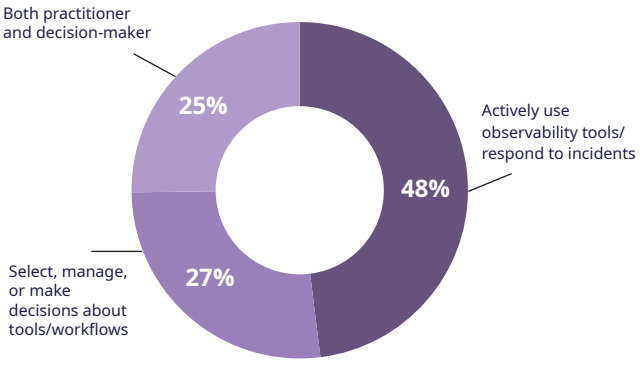
Context is the variable that each of our takeaways shares—its absence as a cause of incident-response failures, its inadequacy as the limiting factor in AI adoption, and its deficit as a structural condition that sustains reactive operations. The opportunity for AI in observability is not another layer of alerting or another investigation interface; it's AI that can assemble, reason across, and act on the full operational context of a production environment. Until that bar for context is consistently met, organizations will continue to apply rational constraint on AI use: AI will be allowed to look and recommend, but the decision to act will remain with the slower human, more likely to be stressed and overwhelmed, who alone carries the context AI doesn't yet reliably hold.

Recommendations

- **Prioritize adding context to AI for observability.** Context inadequacy is a critical and frequently unappreciated limiting factor. Before adding more AI capabilities, teams should assess whether their tools can connect alerts, services, infrastructure, recent changes, and team knowledge into a coherent incident picture.
- **Employ human-in-the-loop remediation.** The 78% in our poll who favor HITL models are not resisting AI—quite the contrary. They are ensuring trust through transparency. Initial AI automation should recommend, explain provenance, and execute only with explicit approval.
- **Start by replacing manual correlation.** Manual correlation is among the highest-burden activities in incident response and among the most accessible to automate. Teams should identify where engineers spend the most time on investigation and implement workflows to enrich context.
- **Treat observability maturity as an operational model, not a tooling checklist.** The risk landscape respondents describe—configuration drift, visibility gaps, skills concentration, staffing dependencies—reflects operational knowledge fragmented across people, tools, and systems, lacking a common context layer. Addressing it requires investment in process and coordination as much as tooling implementations.
- **Take AI as your chance to move further toward prevention.** Reactive operations are self-reinforcing: incident volume consumes the bandwidth that would otherwise go toward prevention. AI has a role at both ends—reducing mean time to resolution now, and surfacing recurring failure patterns over time.

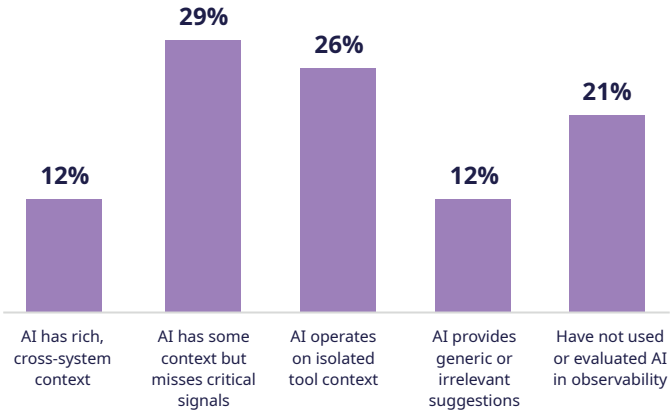
Research Results

Which of the following best describes your involvement with observability or incident response in your organization?



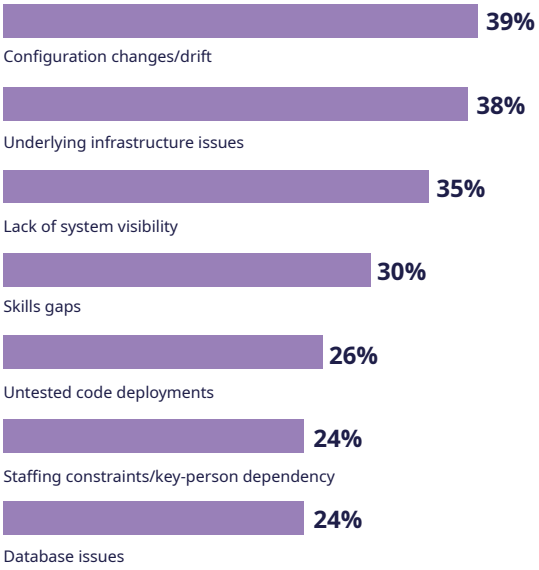
The survey drew from a practitioner-weighted respondent base, with 48% actively using observability tools and responding to incidents as part of their daily work. Another 25% occupy both practitioner and decision-maker roles simultaneously, and 27% focus primarily on tool selection and workflow governance. This distribution matters for interpreting what follows: three-quarters of respondents have direct operational exposure to the pain points the survey explores, which grounds the findings in day-to-day production reality rather than theoretical preference.

When your team uses or has evaluated AI-assisted observability, how well does the AI include the full context of an incident—including relationships across services, infrastructure, recent changes, and team knowledge?



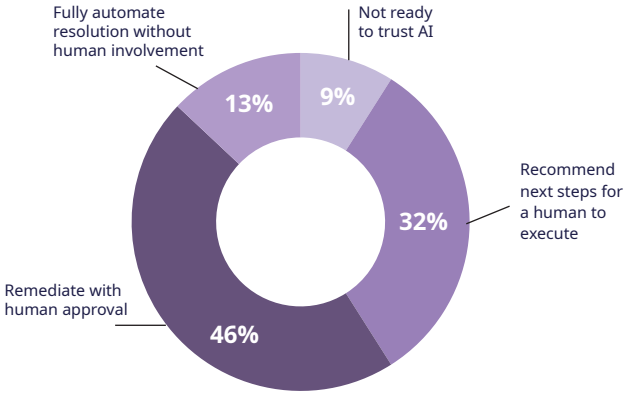
Only 12% of respondents who have used or evaluated AI in observability report that it delivers rich, cross-system context. The remaining 88% distribute across four less encouraging positions: 29% say AI has some context but misses critical signals, 26% say AI operates only on isolated tool context, 12% say AI produces generic or irrelevant suggestions, and 21% have not used or evaluated AI in observability at all. Combined with the data on the use of AI in operations, these numbers explain the maturity cliff directly: organizations are withholding autonomous action because fewer than one in eight have encountered AI that knows enough about the full operational environment to justify trusting it.

What are the two biggest risks to your production environment today?



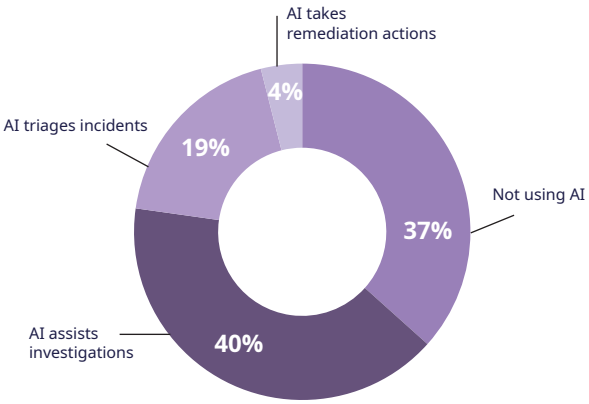
The three leading production risks—configuration drift (39%), underlying infrastructure issues (38%), and lack of system visibility (35%)—share a defining characteristic: each represents a gap in operational context rather than an isolated technical failure. Skills gaps (30%) and staffing/key-person dependencies (24%) reinforce that critical operational knowledge frequently resides in individuals rather than in systems. Collectively, this risk landscape describes environments in which the context required to understand, predict, and prevent incidents is fragmented—and in which AI that cannot assemble that context will add another data source without solving the underlying problem.

If an AI agent could investigate incidents and identify root cause in minutes, what would you most want it to do next?



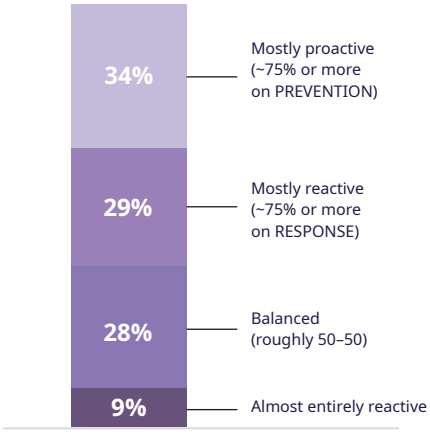
Given a hypothetical AI agent capable of identifying root cause in minutes, respondents concentrated their preferences around human-in-the-loop models. Remediation with human approval leads at 46%, followed by AI recommendation with human execution at 32%, a combined 78% expressing a preference for AI action paired with human oversight. Fully automated resolution without human involvement attracted 13%, while 9% are not ready to extend meaningful trust to AI for these decisions at all.

Which of the following best describes your team's current use of AI in observability and operations?



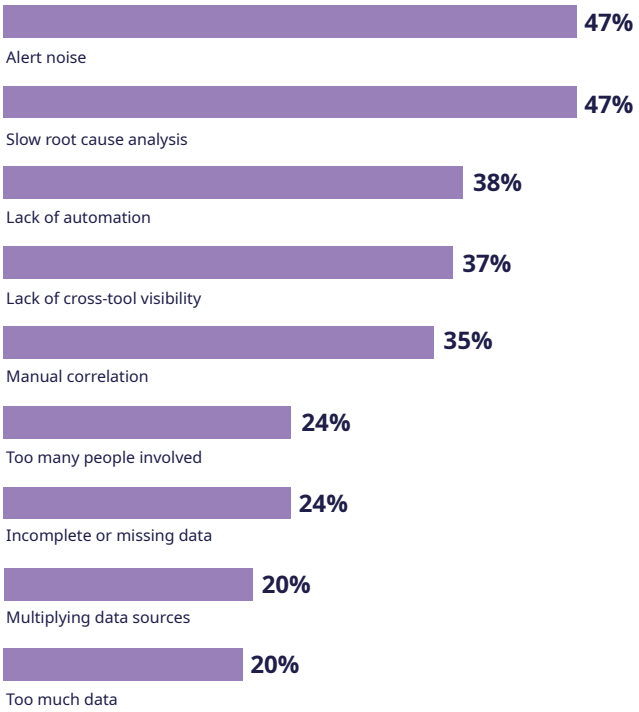
AI use in observability shows a clear progression, with adoption concentrated at the investigation stage and dropping sharply as AI involvement expands into remediation. Investigation assistance is the most common current use case at 40%, reflecting a low-stakes, generally tolerable level of AI involvement. Automated triage drops the usage rate to 19%, and autonomous remediation reaches only 4%, a tenfold decline from AI-assisted investigation to independent AI action. 37% of respondents are not using AI in observability at all. The combination of a large non-adopter cohort with a very small autonomous-action cohort indicates that the market is moving but has not yet found the conditions under which it trusts AI to operate without frequent human checkpoints.

What percentage of your team's operational time is spent on reactive incident response versus proactive prevention and improvement?



Respondents are distributed across the reactive-to-proactive spectrum with no single operating model dominating, but the balance tilts toward reactive. A combined 38% operate in mostly or almost entirely reactive modes, compared to 34% who are mostly proactive and 28% who describe themselves as balanced. The gap is modest in absolute terms but meaningful: a larger share of enterprise observability teams remains primarily consumed by incident response. This matters for AI adoption because reactive operations deplete the organizational bandwidth required to evaluate, validate, and mature AI tooling. Teams spending 75% or more of their time fighting fires have limited capacity to build trust in AI systems that would eventually reduce those fires.

What are your biggest pain points in incident response today?



Alert noise and slow root cause analysis each drew 47% of respondents; lack of automation (38%), lack of cross-tool visibility (37%), and manual correlation (35%) complete a cluster of pain points that collectively describe a single underlying failure: environments that generate more signal than teams can contextually process fast enough to act on with confidence. What is striking about this distribution is its coherence—these are five angles on the same problem. The lower-ranked responses, including too many people involved and incomplete data (each at 24%), reinforce that this is as much a coordination and knowledge problem as it is a tooling one.

This Techstrong PulseMeter is sponsored by NeuBird AI.
To learn more about NeuBird AI, visit neubird.ai.