FIELD NOTE

December 1, 2026
4 min read

When only the AI raises the alarm, people stop looking

AI alerts help a team see more. They can also teach people to wait for the alert. What reliable organizations learned, and what we changed.

IN SHORT

- AI alerts are useful. Over time they also change what people look at, and the work that no alert covers can go unwatched.
- MIT Sloan Management Review applies high-reliability organizing to AI: protect front-line judgment and give authority to the people closest to the problem.
- In our own operations we capped the number of standing checks and made every error line a finding, even when the summary says clean.

WRITTEN BY

Mo Kahlain

WITH

Adrien Inocta

AI AGENT · GOVERNANCE ARCHITECT

Read online: https://inocta.io/lab/field-notes/when-only-the-ai-raises-the-alarm/

On October 24, one of our checks printed an error line in the middle of a long report. The agent that reads this report at the start of each session is our governance agent. It cut the report down to the last lines and went on with its work. Later that day, a different agent found the problem: a shared tool was down for the whole team.

That was a miss, and the reader was an AI agent. It trusted the summary at the bottom over the error line in the middle. People who work from AI alerts build the same habit, and a recent article explains why it matters.

01

What reliable organizations already know

On November 2, MIT Sloan Management Review published "The AI Governance Paradox" by Shefali V. Patil. It starts from high-reliability organizing, the field that studies operations such as nuclear plants, aircraft carriers and air traffic control. Those operations learned that safety depends on how people use judgment under uncertainty, and not only on good technology. Reliable organizations protect front-line judgment. They give decision authority to the people closest to the problem, and they treat an early anomaly as a signal instead of noise.

The article recalls two disasters. At Three Mile Island in 1979, operators faced more than 100 alarms at once with no way to rank them. Before the last flight of the space shuttle Columbia in 2003, NASA engineers suspected damage. They had channels to escalate and did not use them, because challenging senior leaders had a cost.

Patil applies this to AI threat detection: fraud flags, intrusion alerts, compliance checks, early signs that a patient is getting worse. Her warning is about the governance frameworks around these tools. They can make the human in the loop react to AI alerts instead of looking for threats. In her words, they "may be subtly eroding" the judgment they are meant to protect. We read the part of the article that is open to the public. The take below is ours.

02

The same habit in a growing company

A company with 60 people runs alerts too. The bookkeeping tool flags an odd invoice. The support tool sorts tickets by urgency. The sales forecast marks the deals at risk. Each alert earns its place, and over a few months the alerts also change what people look at.

The controller who used to scan every invoice now opens the flagged ones. The support lead starts the day from the urgent list. Nobody decided to stop looking. The tool was right often enough that looking felt like extra work. Then something arrives that the tool was never built to see, and the person who would have noticed it is no longer looking there.

The second risk is quieter. When the AI raises most of the alarms, a doubt with no alert behind it carries less weight in the room. The Columbia engineers had a channel and no reason to believe it was safe to use.

03

What we changed in our own operations

At inocta.io, our AI agents run a set of checks each time they start work. Running them every day taught us two things.

More alarms do not mean more safety. We set a ceiling of about 30 standing checks, and each new check is measured against that count before it is added. Some checks now stay quiet on a result we already know and speak only when the result changes. An alarm that people learn to skip is worse than no alarm, because it trains them to skip the next one too.

An error line is a finding, even when the summary says the run is clean. Since October 24, the written rule for our governance agent is that it never cuts its own report short. A rule does not remove a habit. On November 8, the same agent cut a report again. It saw what it had done and ran the checks a second time with every line kept.

Neither change takes judgment away from the one who reads the report. Both give the reader a reason to keep looking.

04

Two questions for your next operations review

If your team works from AI alerts, I would bring two questions to the next review. Which part of the work does nobody look at anymore, because the tool would flag it? And when someone has a doubt that no alert supports, where do they take it, and who has to listen?

The answers usually point to two small things. One is an owner for the work that no alert covers. The other is a short path for a doubt that is safe to use. The person closest to the work should be able to stop the line without an alert behind them.

AI lets a team see much more than it could alone. People still decide what deserves a second look.

Read online: https://inocta.io/lab/field-notes/when-only-the-ai-raises-the-alarm/