It Ran Smoothly

Why AI decisions land in the wrong box — and why only the outlier finds them

An ERP upgrade went in, untested. One stock item's parameters were pointed at the wrong stock master file for the balance sheet update, and everything ran smoothly for a couple of months. Nobody picked it up — because there was nothing to pick up. The system produced its balances, month after month, and the balances were wrong.

Stock on hand in the balance sheet was understated by $62,500. Year-end did not catch it either. The financial accounts were drawn up, income tax was calculated on the incorrect profit, and penalties were paid. And the worst part: this was a subsidiary of a company traded on the stock exchange, and the error had to be fudged in the following reporting period.

That sentence — "it ran smoothly" — is the whole problem. A system that is wrong in a consistent way throws no error. It does not stumble, it does not warn, it does not ask for help. It runs smoothly while it is wrong, and every check inside it confirms that everything is fine, because every check is testing the system against itself.

The wrong box

MIT's Center for Information Systems Research published a framework in June 2026 that every executive handing decisions to AI should read first. Sebastian, Weill, Haskamp and vom Brocke sort decisions by two dimensions: ambiguity — how clearly the data determines the answer — and risk — what it costs to be wrong. Their AI Decision Matrix gives four boxes, and four different instructions.

Decision	Ambiguity and risk	What MIT says to do
Routine	Low ambiguity, low risk	Automate it
Consequential	Low ambiguity, high risk	Manage it, with humans monitoring and a named risk owner
Exploratory	High ambiguity, low risk	Enable it, with humans framing and learning
Strategic	High ambiguity, high risk	Lead it, with humans owning the framing and the learning

The automated stock posting in my opening story was put in the first box. It belonged in the second. The posting logic was clear — low ambiguity — so it looked routine. But the consequences of being wrong were listed-company accounts, tax on the wrong profit, and penalties. That is consequential, and consequential decisions are not handed over and left. They are managed, monitored, and owned by a named human who can change them.

MIT splits every decision into three parts: frame it, act on it, learn from it. Frame means defining the problem, the assumptions, the constraints and what success looks like. Act means weighing options and executing. Learn means monitoring outcomes, being accountable for them, and updating the system. In the upgrade story, learning failed twice. Year-end — the checkpoint that exists precisely to catch this — missed it. And when it surfaced, it was fudged into the next period instead of owned. That is the opposite of learning.

Gartner's numbers, published in October 2026, say this is not one company's bad luck. Only 13% of CIOs in their survey of 2,002 technology leaders say their enterprise has realised significant or transformational value from AI so far, and the likelihood of any single AI use case succeeding is one in five. A separate Gartner survey found only 27% of executives have a comprehensive AI strategy. Gartner's own line for the gap is blunt: "Volume is not value." I agree with the number. I disagree with where most people look for the cause. They look at the model. They should look at the box the decision was put in — and at the data that put it there.

When the data deletes the choice

Dirty data hides ambiguity by deleting it. I was running a statistical analysis in a hardware business when I found that one stock code covered screwdrivers from different manufacturers — same selling price, different gross profits — and that stock item codes had been reused for disparate hardware items entirely. One code, one supposed answer, many different realities underneath it. The differences had been erased at classification, before anyone made a decision at all. By the time a manager — or an AI agent — looked at the item, there was no visible choice left to make. The ambiguity had not been resolved. It had been deleted upstream, by the data itself.

That is a different industry from the ERP story, and that is the point. The same failure pattern, in two unrelated industries: the data presents one clear answer, so the decision is classified as routine, so it is handed over — and the hand-over is where the error gets its speed.

Dirty data hides risk the same way, from the other side. In the ERP story the wrong posting looked safe because it ran smoothly. A wrong-but-consistent posting throws no exception, so nothing in the system signals danger. The risk stayed invisible until it landed outside the system, in drawn accounts, in tax penalties, and in the standing fear of the stock exchange becoming aware of it. Dirty data does not announce risk. It defers it, to the place where it is most expensive.

Governance is not sexy

Most people think data governance is a panacea for correct, accurate reporting. I think that faith is flawed, because governance, like cleaning data, is not sexy.

That is not a joke about glamour. It is a mechanism, and it is the one MIT's framework does not name. MIT designs decision rights on day one: set the rights, name the owner, build the foundations. But governance lives in time, and time is unkind to it. The excitement of doing the exercise loses its shine. The checking becomes boring. Repetitive work is suited for AI, not humans — and humans are bad at sustained repetitive vigilance, so the checking gets handed over as processing, to the hands with the least context, the least authority and the least mandate to question it.

A check handed over that way is not a check. It is a routine. And a routine inside a closed transaction loop cannot find the loop's own problems, because it is testing the transactions against the same data and the same assumptions that produced them. The loop confirms itself. That is exactly what happened in the ERP story: every internal check passed, because every internal check was part of the system that was wrong.

What finds the problem has to sit outside the loop. That is the subject of Chapter 11 of *Dirty Data* — "Outliers: The Data That Refuses to Fit." The outlier is not an error to be smoothed away. It is the data point that refuses to fit the story the loop is telling, and it is the only signal the loop cannot manufacture for itself. I found the screwdrivers that way: not by auditing the transactions, but by looking at the distribution from outside them, and letting what did not fit speak.

Rank is not the villain here. The person who would have caught either of these errors is not defined by seniority. They are defined by a stance: they question the data before trusting it, they trace consequences, and they do not bend to the senior narrative. Hand that person a boring check and they will still question it. Hand a check to someone whose job is defined as processing, with no mandate to question, and the check will pass whether it is true or not. Governance fails when judgement is delegated away — and judgement delegated away is judgement lost.

What the leader does

A decision handed to AI on data nobody has interrogated can look routine while it is actually consequential, because the dirty data has hidden both its ambiguity and its risk — and the system will run smoothly while it is wrong. So before a leader hands AI a decision, he takes ownership of the data behind it and the outcome it produces, and he puts a check outside the loop to prove both.

Ownership means three concrete things. First, classify the decision honestly: which box is it really in, and what would have to be true about the data for that box to be right? Second, name the human who owns the outcome — not a role, not a committee, a person who can be asked, and who can change the system. Third, build the outside check before the decision goes live: a statistical view, a distribution, an outlier report that does not depend on the loop's own rules, and a person whose job is to read it.

Gartner's winners, in that same October survey, are more likely to set stopping rules — guidelines for when a pilot should not move to production, and for abandoning low-value initiatives already in production. That is Gartner's finding, and it is consistent with everything above: the leader who owns a decision also owns the decision to stop it. I have not lived that particular discipline myself, and I will not pretend otherwise. But I have lived its absence, in a balance sheet that ran smoothly while it was wrong.

MIT asks where AI can be used. The better question is the one this brief has been circling: before AI is given a decision, who owns the data it will be decided on, who owns the outcome, and what check outside the loop will catch the data point that refuses to fit?

Structure before intelligence — and ownership before automation.

Sources: Gartner, "Gartner Survey Finds Only 13% of CIOs Say Their Enterprise Has Realized Significant or Transformational Value from AI", 19 October 2026 (2027 CIO and Technology Executive Agenda survey, n=2,002, May–July 2026); Gartner, "Gartner Predicts by 2027, 50% of Enterprises Without a People-Centric AI Strategy Will Lose Their Top AI Talent", 13 May 2026 (Global Labor Market Survey 1Q26, n=12,004; CxO survey December 2025, n=197); MIT Center for Information Systems Research, Sebastian, Weill, Haskamp and vom Brocke, "Designing Decision Rights for AI", Research Briefing, June 2026 (30 interviews, 27 executives, nine companies, illustrated by One New Zealand). Gartner stopping-rule figures are Gartner's findings, cited as such.

Gary Segal is the author of Dirty Data and Switching ERP for Embedded AI. He writes the weekly Structure Before Intelligence™ executive brief at dirtydataexpert.com.