TECH mahindra

WHITEPAPER

Agentic Data Sentry: The Case for Autonomous Data Quality and Self-Healing Pipelines

How enterprises move from reactive data monitoring to governed, self-correcting data operations

Executive Summary

Enterprise data quality has become the primary blocker for AI adoption. Poor data quality costs organizations an average of $XX million a year, and that figure grows as more AI systems are built on unreliable inputs.1 This makes it one of the leading barriers to AI adoption cited by data and analytics leaders, ahead of talent and budget constraints.2 Rule-based monitoring and manual oversight were built for a slower data environment and cannot keep pace with pipelines spanning cloud, on-premises, and edge systems. The result is a widening gap between how fast enterprises want to deploy AI and how much they can trust the data behind it.

What separates enterprises that scale AI from those in pilot mode is a shift from reactive monitoring to autonomous data quality, governed by clear boundaries. Monitoring has stalled what autonomous data quality, self-healing pipelines, and agentic data operations look like in practice, and how enterprises can govern that autonomy without slowing it down. Data leaders who build that governance will spend the coming years scaling AI successfully.

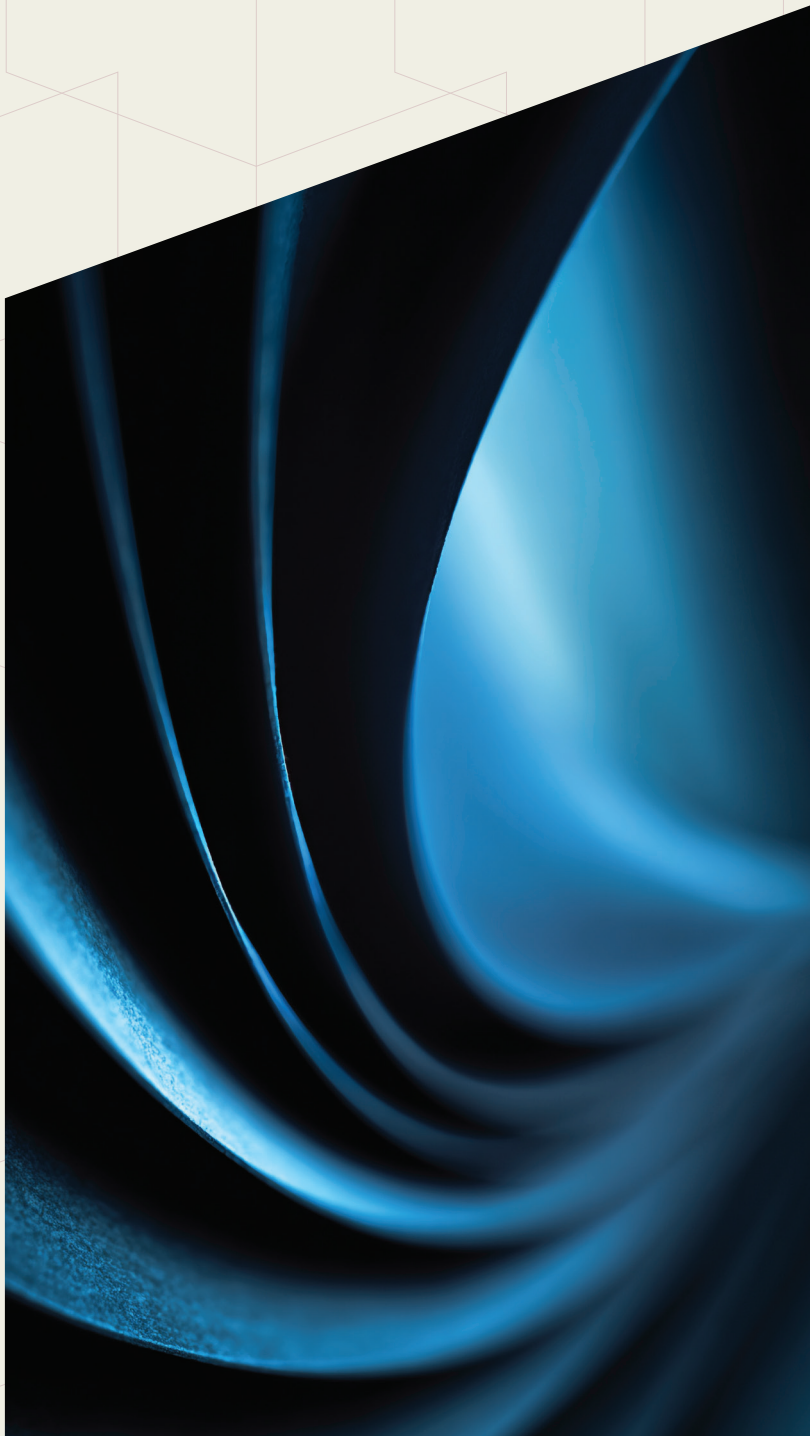


Table of Contents

Your Data Is Lying to You	4
Why Monitoring Alone Is Not Enough	5
From Guardian to Agent: The Next Layer of Data Intelligence	6
Effective Governance: Turning Autonomy Into a Safe Operating Model	7
Governed Autonomy Keeps Ethics and Risk Non-Negotiable	8
The Enterprise That Governs Its Data Governs Its Future	8
About the Author	9
References	9

Your Data Is Lying to You

Every enterprise AI initiative rests on one assumption: the data feeding it is accurate, complete, and current. Gartner estimates poor data quality costs companies an average of $12.9 million a year, resulting in lost productivity, flawed decisions, and remediation once errors surface.[1] Any model or workflow trained on that data repeats and multiplies the error.

In a reporting environment, a flawed record produces a wrong number on a dashboard. In an AI environment, it trains a model, shapes a recommendation, or triggers an automated action at scale, often with no human oversight to catch it beforehand.

Enterprise data volumes are growing faster than most organizations' capacity to govern them, and most organizations only learn their data is unreliable once something breaks downstream. This is a governance failure with a measurable balance-sheet impact.

The pressure increases when enterprises move past pilots. Research from McKinsey on the state of AI found that a growing share of organizations are shifting from experimenting with generative AI to deploying agentic systems that take action with less human review at every step.3 Every one of these agents inherits the data quality of the systems it draws from. An organization that has not solved its data quality problem before deploying agents is not removing risk from the process, it is scaling that risk. Because an agent trusts bad data as much as good data.



Why Monitoring Alone Is Not Enough

Most enterprises already run some form of data quality monitoring, which makes it worth asking why the problem persists regardless. The answer lies in what these systems are built to do. Rule-based monitoring checks data against predefined conditions: a field should not be null, a date should fall within a valid range, and a total should reconcile against a source system. It does not catch problems that fall outside existing rules, and in a fast-changing data environment, these cause the vmost damage.

Modern data pipelines rarely live in one place; they span cloud, on-premises, and edge systems at once. These pipelines are often maintained by teams that joined the organization long after the initial architecture was designed. When a schema changes, an API updates, or a source system migrates, downstream pipelines can fail silently while continuing to pass data through.

By the time a monitoring alert fires or a business user spots a number that looks off, the damage has usually already spread. That's the limitation of reactive monitoring: it's damage control in the rearview mirror. It flags the error after the fact, but it doesn't prevent it or stop it from spreading.

This isn't a dashboard problem; it is an architectural one. Closing this gap requires systems that act on what they detect in real time, rather than waiting on manual interpretation.



From Guardian to Agent: The Next Layer of Data Intelligence

The underlying shift is simple: governance is moving from oversight to action. Traditional data observability tools give teams visibility, dashboards, lineage maps, and alerts that flag when something is wrong. But visibility alone assumes a person will read the alert, diagnose the cause, and fix the pipeline manually. Autonomous data quality closes this loop. It pairs the same visibility with agents that detect an exception, determine the likely cause, and correct it, all within guardrails a data team defines and reviews in advance.

This approach to autonomous data governance rests on three pillars.

- **Autonomous data quality:** Agents continuously monitor data, detect anomalies as they emerge, and correct issues before flawed records reach downstream systems.
- **Self-healing pipelines:** Pipelines detect their own failures, reroute data around a broken step, and restore normal operation without waiting for a person to intervene.
- **Agentic data operations:** These manage the pipeline end to end, from ingestion to delivery, inside guardrails a data team sets and continues to refine.

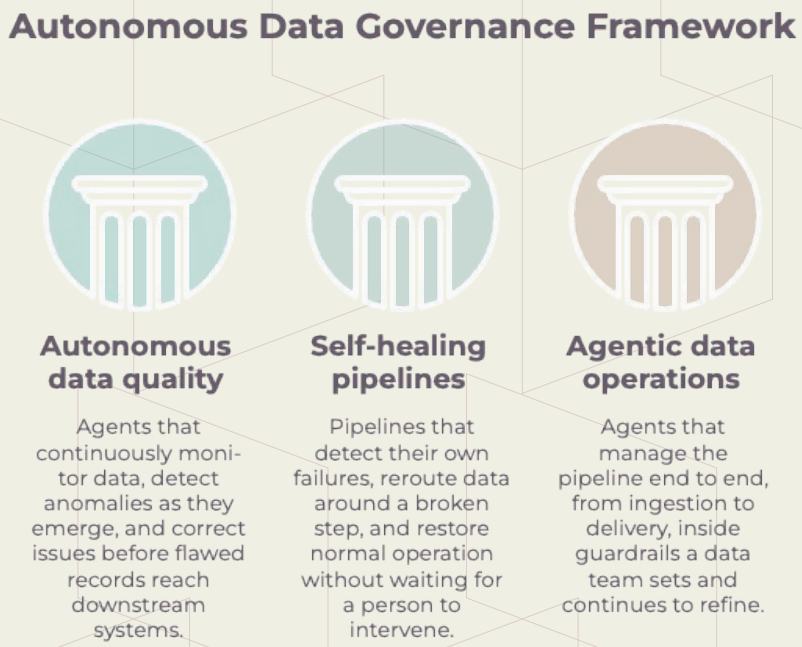


Fig. 1: The Three Pillars of Autonomous Data Governance

Analysts widely expect a significant share of data management tasks to be automated using AI within the next few years. A signal that this shift is a near-term expectation for enterprise data, not a niche experiment. This isn't about having AI; most observability tools already use it to catch unusual patterns. It's about what happens next.

Autonomous data governance goes a step further: it takes a defined, auditable action and escalates to a person only when a situation falls outside its guardrails. Getting the escalation threshold right depends on how much autonomy a pipeline has earned.

Effective Governance: Turning Autonomy Into a Safe Operating Model

Autonomy without structure is just a different kind of risk. None of this works without guardrails that hold up under audit. Effective governance for agentic AI in data operations rests on a few concrete mechanisms:

- Decision boundaries define what an agent can fix on its own and what requires a data owner's approval, based on how significant the fix is.
- Audit trails record what an agent did, what data it used, and why, so the full sequence of events can be reviewed later.
- Rollback mechanisms undo a flawed correction in the same time it took to make.
- Access controls limit which systems and data an agent can reach, using the same access principles applied to employees.

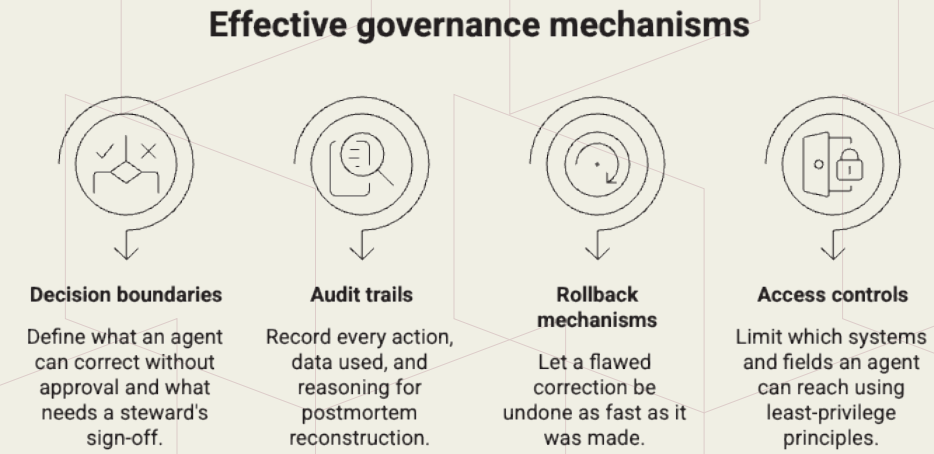


Fig. 2: Governance Mechanisms for Agentic AI in Data Operations

Deloitte's 2026 State of AI in the Enterprise survey found that only 21% of organizations have a mature governance model for agentic AI in place, even as agentic usage is expected to scale quickly over the next two years.[4] Governance lags behind adoption, and closing that gap is what turns autonomy from a risk into an asset.



Governed Autonomy Keeps Ethics and Risk Non-Negotiable

Data governance is no longer an IT concern managed quietly in the background. It is a boardroom question, because every future AI initiative, every model, every agent, every automated decision, will run on the governance foundation an enterprise builds today.

Governance carries an assumption: more oversight means more friction, more approvals, more reviews, and more delay. In practice, the opposite holds true.

Acting without oversight raises real, high-stakes questions. When an agent corrects a record, that correction can shift a customer's credit decision, often without a human reviewing the change before it takes effect. Ethical risk and operational risk are the same risk; without guardrails, a technical fix becomes a liability.

Enterprises building autonomous data governance need to test not just whether an agent's corrections are accurate, but whether they are fair across the populations and use cases the underlying data represents, and revisit that testing as the data and the business context change.

Good governance doesn't slow things down. It's what lets teams move fast without breaking things. The companies asking hard questions about data governance now are the ones that won't have to pause their AI plans later.

The Enterprise That Governs Its Data Governs Its Future

Data governance is no longer purely an IT concern. As AI systems take on more decisions with less human review, the reliability of the data behind them becomes an important question, with direct exposure to revenue, compliance, and customer trust. A model is only as dependable as the pipeline feeding it, and the pipeline is only as dependable as the governance built around it.

Governance has a reputation for slowing teams down. In practice, once it shifts from a manual checkpoint to an automated, self-correcting capability, the opposite holds: teams stop losing time to firefighting flawed data, and that time goes into building the next initiative instead of repairing the last one.

Autonomy is no longer a question of if; it is already running, in every pipeline that corrects itself before anyone notices the error. What remains open is whether that autonomy is governed well enough to earn trust or moving fast enough to lose it. The enterprises that get this right will not be the most automated; they will be continuing to build on their AI investment successfully.

Author



Ravi Pandikunta,

Global D&A Delivery IBG Head- DIG & Manufacturing

With over 25 years of experience in the IT industry, Ravi is a seasoned expert in AI, generative AI, data analytics, automation, and digital transformation. As Vice President and Global Delivery Head at Tech Mahindra, he leads global delivery portfolio, driving AI-powered solutions, advanced analytics, cloud transformation, and intelligent automation to accelerate efficiency, scalability, and business impact for enterprises worldwide. Known for delivering growth and building high-performance team, Ravi has played a pivotal role in shaping digital strategy, expanding global capabilities, and advising Fortune-class organizations across BFSI, manufacturing, retail, CPG, travel, logistics, oil & gas, utilities, and public services. A respected global thought leader with former experience from HCL Tech, HPE, S&P Capital IQ along with 350+ industry talks, Ravi remains committed to enabling future-ready enterprises and mentoring the next generation of AI leaders.

References

1. **Dataversity. (2024, January 19). Understanding the impact of bad data.**
   https://www.dataversity.net/articles/putting-a-number-on-bad-data/
2. **AI & Data Analytics Network. (2025, February 10). Data quality & availability top list of AI adoption barriers.**
   https://www.aidataanalytics.network/data-science-ai/news-trends/data-quality-availability-top-list-of-ai-adoption-barriers
3. **McKinsey & Company. (2025, September 26). The agentic organization: Contours of the next paradigm for the AI era.**
   https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/the-agentic-organization-contours-of-the-next-paradigm-for-the-ai-era
4. **Deloitte. (2026, April 24). Business and IT leaders report AI agents are scaling faster than their guardrails.**
   https://www.deloitte.com/us/en/insights/topics/emerging-technologies/ai-agents-scaling-faster.html

About **Tech Mahindra**

Tech Mahindra (NSE: TECHM, BSE: 532755) offers technology consulting and digital solutions to global enterprises across industries, enabling transformative scale at unparalleled speed. With 150,000+ professionals across 90 countries, Tech Mahindra provides a full spectrum of services including consulting, information technology, enterprise applications, business process services, engineering services, network services, customer experience & design, AI & analytics, and cloud & infrastructure services. It is the first Indian company in the world to have been awarded the Sustainable Markets Initiative's Terra Carta Seal, which recognizes global companies that are actively leading the charge to create a climate and nature-positive future. Tech Mahindra is part of the Mahindra Group, founded in 1945, one of the largest and most admired multinational federation of companies. For more information on how TechM can partner with you to meet your Scale at SpeedTM imperatives, please visit https://www.techmahindra.com/.

www.techmahindra.com
www.linkedin.com/company/tech-mahindra
www.x.com/tech_mahindra

Copyright © Tech Mahindra Ltd 2026. All Rights Reserved.

Disclaimer: Brand names, logos, taglines, service marks, tradenames and trademarks used herein remain the property of their respective owners. Any unauthorized use or distribution of this content is strictly prohibited. The information in this document is provided on "as is" basis and Tech Mahindra Ltd. makes no representations or warranties, express or implied, as to the accuracy, completeness or reliability of the information provided in this document. This document is for general informational purposes only and is not intended to be a substitute tor detailed research or professional advice and does not constitute an offer solicitation, or recommendation to buy or sell any product, service or solution. Tech Mahindra Ltd. shall not be responsible for any loss whatsoever sustained by any person or entity by reason of access to, use of or reliance on, this material. Information in this document is subject to change without notice.