Silent Performance Failure™

The Invisible Threat Inside Your Clinical and Operational AI

An Executive Intelligence Report by SmartSigma AI

Richard G. Greenhill, DHA, FACHE
Principal and Founder, SmartSigma AI

Release Date: March 10, 2026

Executive Summary

9%	Only 9% of FDA-approved AI/ML-enabled medical devices included a prospective study for post-market surveillance.[1]

Read together, these figures describe an industry scaling risk faster than it is scaling accountability.

1,250+	As of July 2025, the FDA's public database listed more than 1,250 AI-enabled medical devices authorized for marketing in the United States — a number that continues to grow rapidly.[2]

Our healthcare industry is deploying artificial intelligence (AI) at scale. Across the spectrum of clinical care and operations (clinical decision support, revenue cycle, operational efficiency, and workforce management), AI tools are going live in health systems every day. What is happening at a slower pace — or in most cases not at all — is the governance infrastructure required to detect when those systems quietly begin to fail. The gaps in governance are not the product of negligence. They are predictable results of an ecosystem focused on deployment rather than stewardship.

This Executive Intelligence Report (EIR) produced by SmartSigma AI experts and practitioners introduces the Silent Performance Failure™ Classification Framework - a repeatable board-level taxonomy, causal chain, and accountability model that translates technical degradation into governance action few have operationalized at the board level.

Developed and trademarked by SmartSigma AI, the framework addresses the gap that every other AI governance framework has acknowledged but none has filled.

The verdict is straightforward. Every high-risk industry that has deployed complex systems at scale has learned the same lessons: critical process failures rarely announce themselves. Aviation, nuclear power, and financial services each responded by building governance infrastructure specifically designed to detect, name, and classify failures before they compound. The discipline of naming failures – giving them taxonomy, causal structures, and accountability – is what separates industries that govern risk from those that merely react to it. The healthcare industry deploying AI in clinical operational systems is now at that inflection point. Silent Performance Failure™ is not a future risk. It is a *present condition* in health systems that have deployed AI without a comprehensive post-deployment monitoring infrastructure.

This paper documents the evidence, names the phenomenon, introduces the classification framework, and describes what responsible post-deployment governance requires. It is written for health system C-Suite Leaders, Clinical leaders, and Board Members who are responsible for the safe and effective use of AI in their organizations — those who will be accountable when the evidence of silent failure eventually surfaces.

The Problem: AI Deployment Without a Safety Net

The governance gap is not the product of negligence – it is a predictable result of an environment focused on deployment rather than stewardship. This happens because vendor contracts dilute monitoring accountability, regulatory frameworks have not yet evolved to require post-deployment surveillance, and health systems focus is on the pace of competitive market demand rather than a pace that sound governance would require.

The core problem is not that AI systems fail. It is that they fail silently. Unlike a system crash or an error message, performance degradation in deployed AI produces no alarm. The system continues operating. Outputs continue flowing into clinical and operational workflows. Decisions continue being made based on those outputs. All the while, the model's performance may be quietly diverging from the parameters under which it was recently validated — eroding its predictive value, amplifying bias, or generating recommendations that no longer reflect the provider's wishes, operational mandate, or patient population it is meant to serve.

> ***Existing monitoring approaches in healthcare AI are often manual, sporadic, and reactive — making them ill-suited for the dynamic environments in which clinical models operate.[3]***

The broader data science community has recognized the gaps. Researchers, regulators, and quality leaders have all identified post-deployment monitoring as a critical unmet need. What has not existed — *until now* — is a specific framework that classifies what is actually happening inside those gaps: a taxonomy of how silent failure occurs, why it accumulates undetected, and what the consequence chain looks like from infrastructure absence to patient harm - to institutional crisis.

A Note on What "Monitoring" Actually Means in Practice

Recent federal data appears, at first glance, to complicate the urgency of this framework. The 2024 ASTP Data Brief reported that among hospitals using predictive AI, 82% evaluated models for accuracy, 74% for bias, and 79% conducted post-implementation monitoring.[4] These numbers deserve closer examination — because what they describe is not what most healthcare leaders assume.

What the survey did not ask – and what these percentages cannot tell us – is how many models in each hospital were never evaluated.[4] A health system evaluating one model out of fifteen could report "yes." Responses were typically provided by chief information officers, and likely not clinical or quality leadership. No standardized methodology, no shared definition of "monitoring," and no classification framework governed what counted as evaluation was mentioned. The American Heart Association, in a November 2025 advisory published in *Circulation*, confirmed that while governance frameworks exist, few health systems have comprehensively translated them into practice — and that actual evaluation reflects considerable variability in local validation, bias assessment, and post-deployment monitoring.[5]

Most critically: those hospitals reported *that* they were monitoring. None had a framework that told them *what they were looking for* — no taxonomy of failure types, no causal chain from infrastructure gaps to patient harm, no shared vocabulary for what Silent Performance Failure™ looks like when it is occurring.

Measuring without a classification framework is the clinical equivalent of taking a patient's temperature without knowing what fever means.

It is worth noting that what constitutes 'adequate' post-deployment surveillance remains undefined in practice — there is no shared standard for what monitoring sufficiency looks like, the boundary between FDA-regulated medical devices and operational AI software is actively shifting, and many of the highest-stakes AI tools in health systems fall outside current device regulation entirely. These gaps do not diminish the urgency; they amplify it.

That is precisely the gap this framework addresses. The question is not whether hospitals are performing some form of evaluation. The question is whether that evaluation is designed to detect the specific failure mechanisms that accumulate silently — and whether it would recognize Silent Performance Failure™ if it were occurring right now in your deployed systems.

The Silent Performance Failure™ Classification Framework maps the complete causal architecture — from absent infrastructure to institutional consequence.

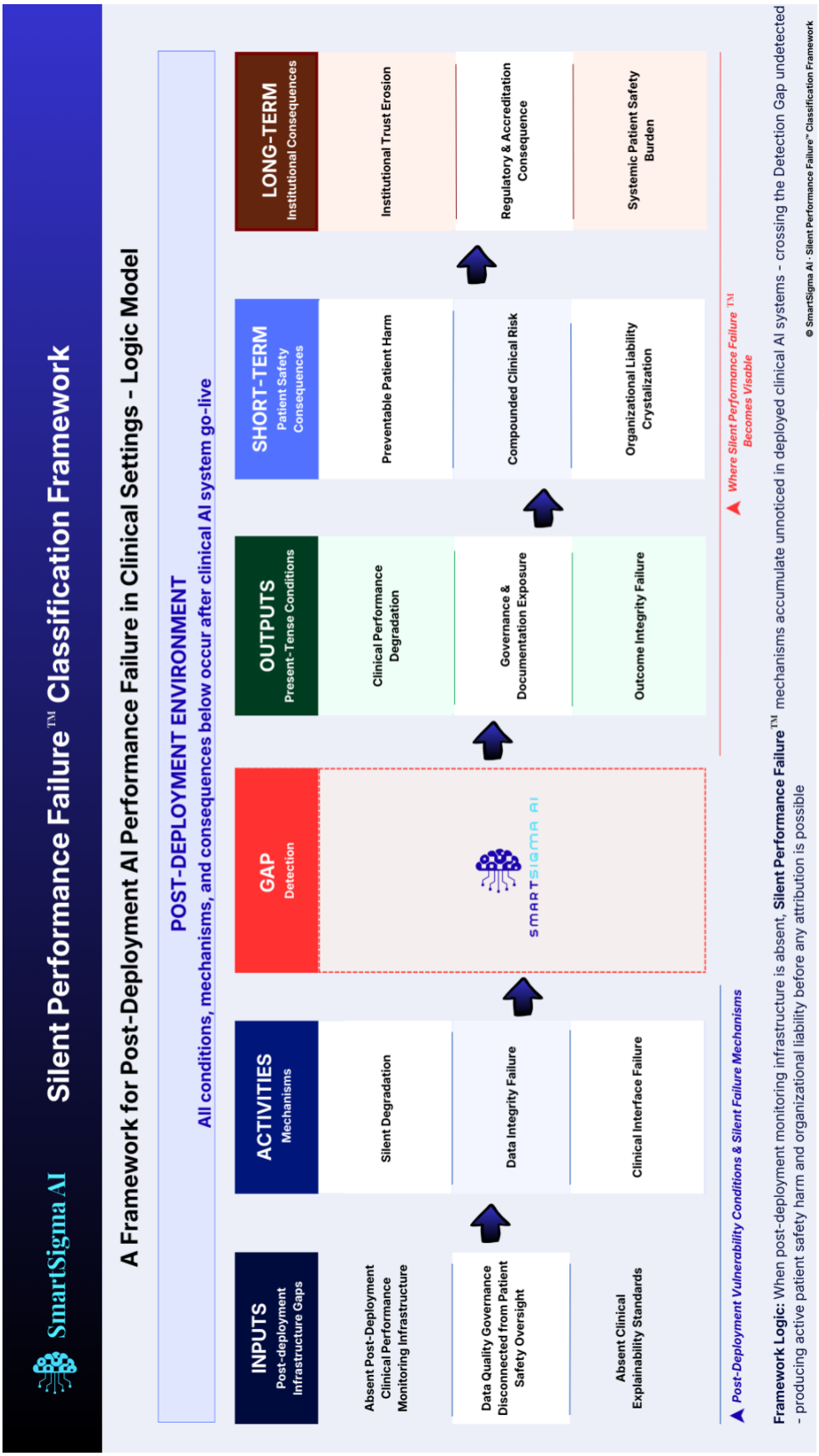

Figure 1. Silent Performance Failure™ Classification Framework Logic Model. © 2026 SmartSigma AI. All rights reserved

This diagram represents a single governing reality: when post-deployment monitoring infrastructure is absent, failure does not wait for permission to accumulate. The three mechanisms in the center — Silent Degradation, Data Integrity Failure, and Clinical Interface Failure — are not sequential. They operate simultaneously, invisibly, and without organizational awareness.

The detection gap at the center is not a technical problem. It is the point at which your governance either intercepts harm or allows it to compound into patient safety consequences and institutional liability. Every AI system your organization has deployed without post-deployment surveillance infrastructure is currently somewhere on this causal pathway. The question this framework answers is not whether that is true. It is where.

How Other Industries Solved This Problem

Before examining the emerging evidence of silent failure in healthcare AI, it is worth noting that this is not an unprecedented issue. Every industry that has deployed complex algorithmic and automated systems at scale has confronted the same fundamental problem: how do you recognize when a system that appears to be functioning has quietly begun to fail? The industries that have answered this question well share a common feature. They built governance and detection infrastructure before catastrophic harm made it unavoidable.

Aviation

Commercial aviation is the most frequently cited model for healthcare patient safety — and for good reasons. Commercial aviation's safety record was not achieved by building better aircraft alone. It was achieved by focusing the systems to continuously monitor aircraft performance after deployment, detect anomalies before they became failures, and create mandatory reporting and review processes that turned near-misses into learning rather than liability. Healthcare continues to strive towards these tenets through focusing on high reliability.

Flight data recorders, real-time telemetry monitoring, and post-flight analysis protocols are in place because the industry understood that complex systems operating in dynamic environments will degrade in ways that ground-based validation cannot fully predict. The insight that transformed aviation safety was not technological. It was governance-oriented and based on the adage: you cannot manage what you do not continuously measure.

Our healthcare systems that leverage AI in managing clinical decisions carry consequences comparable to aviation. The post-deployment monitoring infrastructure does not.

Nuclear Power

The Nuclear Regulatory Commission does not ask facility operators whether they would like to monitor post-deployment system performance. They mandate it. Every operational reactor in the United States operates under continuous performance monitoring requirements, mandatory anomaly reporting protocols, and structured periodic review cycles.

The nuclear industry's foundational insight is directly applicable to AI deployed in healthcare systems related to patient safety and quality: silent degradation in a high-consequence system is not a maintenance issue. It is a precursor event. The question is not whether degradation will occur in complex systems operating over time. The question is whether your governance infrastructure will detect it before it becomes a patient safety incident.

Financial Services

Algorithmic systems used in trading are monitored in real time for model drift and performance degradation. When a trading algorithm begins operating outside its validated and prescribed parameters, automated circuit breakers and risk management systems respond. The 2010 Flash Crash — in which the Dow Jones Industrial Average dropped nearly 1,000 points in minutes before partially recovering — illustrated what happens when algorithmic systems operating at scale interact with fragile market conditions faster than any governance structure can respond. The SEC/CFTC joint review documented the event's complexity and pointed to circuit breakers and real-time monitoring enhancements as part of the remediation response — not as failures that caused the crash, but as governance infrastructure whose absence made recovery slower and attribution harder.[6]

The parallel for systems with embedded AI in healthcare delivery is stark. Clinical decision support systems, revenue cycle algorithms, and staffing optimization tools are the health system's equivalent of trading algorithms. They operate at scale, at speed, and with real-world consequences. Like trading algorithms, their performance can drift in ways that produce patient harm and institutional liability before anyone with accountability has noticed.

The question a Chief Medical Officer (CMO), Medical Director, or Chief Quality Officer (CQO) must ask is not whether the AI model was validated before deployment – it is whether a surveillance infrastructure exists to know if it is still performing as intended today.

Key questions for healthcare leaders to consider are: *do our current governance processes detect silent AI algorithmic degradation in our systems before it reaches a patient? Or are we waiting for a sentinel event to find out the system was not working as intended?*

Manufacturing and Industrial Operations

Predictive maintenance exists because the manufacturing industry focuses on silent failure before it becomes catastrophic failure. Industrial operations do not wait for equipment to break to discover it is degrading. Sensors monitor performance continuously. Anomaly detection algorithms identify drift from baseline. Maintenance is triggered by performance signals, not by failure events.

In healthcare we have borrowed this logic for clinical equipment — biomedical devices are maintained on scheduled intervals rather than waiting for failure. However scheduled maintenance is still predominantly preventative and not predictive.

> **Every industry that has deployed complex algorithmic systems at scale has confronted what this framework defines as *Silent Performance Failure™* - the silent degradation of high-consequence systems before detection. Every industry that has done so successfully built detection infrastructure before catastrophic harm occurred.**

Healthcare AI: The Inflection Point

Healthcare has had the benefit of watching other industries navigate this challenge. The evidence from aviation, nuclear operations, financial services, and manufacturing is not historical curiosity. It is a roadmap — and a warning. The roadmap shows that post-deployment monitoring infrastructure built proactively prevents catastrophic harm. The warning is what happens when oversight infrastructure had been designed for a different era of AI – fewer tools, lower clinical stakes, and consequences that were operational rather than patient-facing.

Healthcare AI is at that inflection point now. The deployment curve has steepened dramatically. The governance curve has not followed. Researchers studying the deployment pattern of clinical AI have described it as a "linear model" - models are developed, validated, deployed, and then effectively frozen, with parameters locked and little evidence of systematic post-deployment evaluation occurring in practice.[7]

> **An AI Accelerator of the IHI Leadership Alliance—part of our foremost quality organization — framed post-deployment AI monitoring as an open question in October 2025, asking: "How do we monitor AI's performance post-deployment and signal when it begins to underperform?"[8] That question had not yet been directly answered with specifics of a governance framework. Until now.**

Our industry knows the gaps exists. What it has not had is a framework that classifies what is happening inside that gap — a taxonomy of failure types, a causal chain from infrastructure absence to patient harm, and a vocabulary that health system leaders rather than data scientists can act on.

The following section shares evidence that Silent Performance Failure™ is not a theoretical future risk. It is the present condition in health systems that have deployed AI without the infrastructure to detect it.

The Evidence: Silent Performance Failure™ Is Already Occurring

The following evidence is drawn from peer-reviewed literature and documented research. In each case the pattern is the same: an AI system was deployed, it operated within existing workflows, and its performance degraded in ways that were not detected by the governance infrastructure in place. The harm accumulated. Attribution came late — if at all.

Silent Degradation in Clinical Settings

Researchers studied AI performance in medical image classification and documented distribution shifts between training data and real-world deployment data. These shifts produced silent failures that none of the existing confidence-based detection methods reliably prevented.[9] The systems kept running. They kept generating output. Those outputs would continue influencing clinical decisions. The failure was silent because nothing in the operational environment signaled that something had gone wrong.

A study from 2025 of AI system degradation in healthcare found that calibration drift — a gradual shift in the relationship between model confidence and actual accuracy — can silently compromise clinical decision-making over time. The review concluded that ongoing monitoring and timely recalibration are essential for safe AI deployment, and that without them, deployed systems cannot be assumed to be performing within their validated parameters.[10]

The clinical and operational implications are direct: a sepsis prediction model, a readmission risk tool, or a deterioration alert system may be generating outputs that clinicians are acting on while the model's actual predictive accuracy has quietly declined from its validated baseline. No alert fires. No dashboard turns red. The system is operating. It is no longer working as intended.

Data Integrity Failure: The Bias Case

In 2019, researchers published what became one of the most cited studies in healthcare AI governance: a documented analysis of a commercial algorithm used across hundreds of health systems to manage the care of populations. The study found that the algorithm systematically underestimated the health needs of Black patients — not because of an error in deployment, but because the data feeding the model had diverged from the population it was meant to represent.[11]

The algorithm had been operating across healthcare systems for years before the bias was identified — through external academic research, not internal monitoring. This is textbook Data Integrity Failure under the Silent Performance Failure™ taxonomy: the data governance system was disconnected from patient safety oversight, the failure accumulated silently, and the detection came from outside the governance structure rather than from within it.

This is not a solitary case. A review of cardiovascular AI systems found that only approximately 17% had been tested in randomized clinical trials, and just one-third shared code or data for external scrutiny.[12] This challenge is compounded for AI tools operating within proprietary systems such as electronic health records. Health systems frequently assume that vendor-embedded AI carries vendor accountability for performance monitoring. In practice, there is no standardized expectation for what vendor contracts should guarantee regarding AI performance surveillance. The health system owns the patient's outcome but whether visibility into ongoing model performance is contractually assured, informally assumed, or entirely absent varies by vendor, contract, and health system sophistication.

Interface Failure: The Accountability Gap

Far beyond model drift and data integrity, a third failure mechanism operates at the point where an AI output meets human decision-making. When clinicians cannot explain or interrogate the basis of an AI recommendation, the system's outputs become a governance liability regardless of the underlying model's technical performance.

Revenue cycle AI systems – such as those managing coding accuracy and authorization prediction – are exposed to continuous changes in payer rules, coverage policies, and billing requirements. Benchmarking documented denial rate increases from 10.15% in 2020 to nearly 12% in late 2023.[13] In this environment, provider-side AI model degradation becomes difficult to detect – rising denial rates are often attributed to payer behavior, which obscures whether the organization's own AI tools are performing as intended. The financial erosion is real.

A recent review of enterprise AI governance in healthcare identified that health systems deploying AI without cross-functional governance committees — including data scientists, clinicians, compliance officers, and ethics experts reviewing performance throughout the AI lifecycle — are operating with a structural accountability gap that no amount of pre-deployment validation can compensate for.[14]

> **In each of these documented cases, the AI system continued operating. The unintended outcome accumulated. Attribution came late — if at all. This is Silent Performance Failure™.**

Naming What Governance Has Been Missing

This framework builds on the foundational literature. Machine learning researchers — including Gama and colleagues on concept drift — have documented post-deployment degradation as a technical phenomenon. The FDA has issued post-market surveillance guidance for AI/ML-based Software as a Medical Device. Patient safety researchers have studied diagnostic error and automation complacency. This work is not a gap. It is the starting point.

The NIST AI Risk Management Framework, the FAIR-AI framework, and the drift detection literature have collectively established that post-deployment monitoring is necessary, what governance structures should support it, and what statistical signals data teams should track. This framework addresses what that body of work leaves unanswered: when silent failure is already occurring in a deployed system, what exactly is the organization looking at — and who is accountable for it? That is not a data science question. It belongs to the C-suite and the board, and no existing framework has answered it at that level — with a named taxonomy of failure mechanisms and a causal chain that health system leaders, not data teams, can act on.

> **Drift detection is a data science problem. Silent Performance Failure™ is a governance problem.** *They are related. They are not the same. Only one has a framework built for health system leaders – not data scientists.*

Regulatory frameworks describe the surveillance infrastructure that should exist. Silent Performance Failure™ names what happens to patients and institutions when it does not. One is prescriptive. The other is descriptive of present reality in most deployed clinical AI environments. The logic model format used to communicate this framework is intentionally borrowed from established healthcare quality practice — the theoretical contribution is not the structure; it is the taxonomy and the causal chain it encodes.

The patient safety literature has studied how clinicians over-trust automated systems and stop independently verifying their outputs. That is a human behavior problem. Silent Performance Failure™ is an organizational infrastructure problem — it addresses whether the institution itself ever knows the automated system has begun to fail, regardless of how any individual clinician responds to it. A clinician can be appropriately skeptical of an AI output and still have no mechanism to surface that concern to the people accountable for governing the system. That gap — between individual clinical judgment and organizational accountability — is precisely what this framework is designed to close.

The Three Mechanisms of Silent Performance Failure™

The Silent Performance Failure™ Classification Framework identifies three distinct mechanisms through which deployed AI systems fail silently in healthcare settings. The diagnostic indicators, subcategory taxonomy, detection protocols, and governance response criteria for each mechanism are proprietary and constitute the operational foundation of SmartSigma AI's post-deployment assessment and advisory practice.

Mechanism 1: Silent Degradation

Silent Degradation occurs when an AI system's predictive performance declines gradually after deployment — without any signal to the organization that decline is occurring. The model continues generating outputs. Clinicians and operators continue acting on them. No feedback loop exists to detect that what the model was validated to do and what it is currently doing has quietly diverged.

What this looks like in practice:

A sepsis prediction model deployed 18 months ago begins generating alerts at a rate meaningfully different from its validated baseline — either flagging fewer high-risk patients or generating more alerts that do not result in clinical intervention. No error message fires. The model is technically operational.

What is at stake: Clinical decisions are being made on outputs that may no longer reflect the model's validated predictive value — and the organization has no mechanism to know when that divergence began.

Mechanism 2: Data Integrity Failure

Data Integrity Failure occurs when the data feeding a deployed AI system drifts from the characteristics of the data on which the model was trained and validated — while the system continues operating as if the data remains representative.

What makes this mechanism particularly consequential is that the model itself may appear technically stable. The failure is not in the model. It is in the relationship between the model and its inputs — a relationship that most health system governance structures are not designed to monitor. When that relationship breaks down, model errors do not occur randomly. They occur in patterns — and patterns in AI error become patterns in clinical and operational decision-making at scale.

What this looks like in practice:

A readmission risk model was trained on patient population data from 2019–2021. The health system has since acquired two community hospitals serving demographically distinct populations. The model continues generating risk scores across all patients — including those from the newly acquired facilities — without any adjustment for the changed input population.

What is at stake: The model may be generating systematically unreliable outputs for a defined patient subgroup — producing patterns of error that are not random but directional, and that accumulate undetected across the care continuum.

Mechanism 3: Clinical Interface Failure

Clinical Interface Failure occurs when the connection between AI system outputs and human decision-making breaks down in ways that compromise the safe and accountable use of those outputs — regardless of whether the underlying model is technically performing within validated parameters.

This mechanism is distinct from the first two because the failure is neither in the model nor in its data. It is in the organizational conditions governing how AI outputs reach human decision-making — and what happens at that intersection. A system can be technically performing within validated parameters while the organization has lost meaningful human accountability for the decisions it is influencing.

For non-clinical AI — revenue cycle, operations, workforce — the equivalent failure is the absence of documented human accountability for AI-driven decisions. That absence becomes a liability exposure independent of model performance.

What this looks like in practice:

A clinical deterioration alert system is generating recommendations that clinicians have quietly learned to override — not because the underlying model is wrong, but because the alert fires too frequently and the rationale is not visible. Alert fatigue sets in. The overrides go untracked.

What is at stake: The organization has lost meaningful human accountability for the decisions the AI is influencing — and has no documented evidence of oversight, which becomes a liability exposure independent of whether the underlying model is performing correctly.

The Consequence Chain: From Silent Failure to Institutional Crisis

The Silent Performance Failure™ Classification Framework is not only a taxonomy of how AI systems fail. It is a causal architecture that connects infrastructure absence to institutional consequence. Understanding the consequence chain is essential for communicating this governance challenge at the board level — where the conversation must be about risk, liability, and organizational resilience, not about model performance metrics.

Present-Tense Conditions: What Is Happening Now

When the three failure mechanisms operate without detection infrastructure, three present-tense conditions accumulate simultaneously in the organization. Clinical performance is degrading — AI systems are influencing care decisions in ways that no longer reflect validated performance. Governance and documentation exposure is growing — the organization is operating AI systems without the audit trail that demonstrates responsible oversight. Outcome integrity is failing — the connection between AI-assisted decisions and actual patient outcomes is neither monitored nor documented.

These are not future risks. They are current conditions in any health system that has deployed AI without post-deployment monitoring infrastructure. The question is not whether they are occurring. The question is whether the organization has the infrastructure to see them.

Patient Safety Consequences

> **Preventable patient harm does not announce itself. It accumulates in the gap between what an AI system was validated to do and what it is actually doing — undetected, in clinical workflows that have come to depend on outputs that are no longer trustworthy.**

The patient safety consequences of undetected Silent Performance Failure™ are not theoretical. They are the logical and documented outcome of deploying clinical decision support systems without monitoring post-deployment performance. Preventable patient harm occurs when clinical decisions influenced by degraded AI outputs produce outcomes that would not have occurred had the AI been performing within its validated parameters — or had the clinician known it was not.

Compounded clinical risk occurs when multiple AI systems in the same patient pathway are simultaneously experiencing silent failure without detection. In a health system where sepsis prediction, deterioration monitoring, and medication management AI all operate without coordinated performance oversight, failures can compound across the care continuum before any single point of failure is identified.

Institutional Consequences

> **Organizational liability crystallizes not at the moment of harm but at the moment attribution becomes possible. The question regulators, payers, and plaintiffs will ask is not whether you deployed AI — it is whether you governed it.**

Regulatory and accreditation consequence is no longer a distant possibility. CMS, The Joint Commission, and state health departments have begun signaling movement toward AI oversight expectations — through advisory statements, public comment processes, and evolving accreditation conversations — even as formal requirements remain nascent.

> **Institutional trust erosion is the long-term consequence that no balance sheet can easily capture but every board must understand. Health systems that cannot demonstrate responsible AI governance will face increasing scrutiny from patients, payers, accreditors, and communities. Trust, once compromised by documented AI governance failure, is extraordinarily difficult to rebuild.**

What Responsible Post-Deployment Governance Requires

Responsible post-deployment AI governance requires a structured response to each mechanism of Silent Performance Failure™ — across monitoring infrastructure, data governance, explainability standards, and enterprise-wide scope. The requirements are not aspirational. They are baseline conditions for operating AI systems that influence decisions affecting patient safety.

Your board will ask three questions: First, is our deployed AI still performing as validated? Do we know when it stopped? Is there documentation that demonstrates our governance posture? Responsible post-deployment governance allows a confident answer of "YES" to all three – before a sentinel event, a regulatory inquiry, or a plaintiff's attorney makes those questions unavoidable.

The health systems that govern AI responsibly — before regulatory mandates formalize expectations, before patient harm events force attribution, before board-level scrutiny arrives — will be the ones that demonstrate what trustworthy AI deployment actually looks like.

SmartSigma AI's Silent Performance Failure™ governance assessment gives health system leaders a structured, evidence-based answer to the question your board will eventually ask.

*Begin your **complimentary** Silent Performance Failure™ governance assessment by contacting us at* admin@smartsigmaai.com

SmartSigma AI offers a structured executive governance briefing — designed for health system leadership teams and board quality committees — that applies the Silent Performance Failure™ framework directly to your deployed AI inventory.

This is not a vendor presentation.

It is a governance conversation, conducted under confidentiality, that gives your leadership team a defensible answer to the question your board will eventually ask: *were we governing our AI before something went wrong?*

Request a private governance conversation at admin@smartsigmaai.com

About SmartSigma AI

SmartSigma AI was founded on a straightforward observation: healthcare organizations are deploying artificial intelligence at a pace that their governance infrastructure cannot yet support. The consequences of that gap are not theoretical. They are accumulating now, in clinical and operational AI systems operating without the post-deployment monitoring, data governance, and accountability structures that responsible deployment requires.

About the Author

Richard G. Greenhill, DHA, FACHE

Richard G. Greenhill, DHA, CPHQ, HACP, FACHE is an internationally recognized healthcare quality and patient safety expert and the architect of the Silent Performance Failure™ Classification Framework. He serves as Editor-in-Chief of the International Journal for Quality in Health Care, Communications, published by Oxford University Press in partnership with the International Society for Quality in Healthcare.

He was elected by his international peers to the prestigious International Academy for Quality and Safety — an honor previously conferred on global leaders including Sir Liam Donaldson, Donald Berwick, and Jeffrey Braithwaite. He serves as a subject matter expert for hospital resilience with the World Health Organization Eastern Mediterranean Regional Office and is a sought-after national and international speaker on AI use in healthcare delivery, quality, patient safety, and hospital resilience.

Dr. Greenhill's perspective on AI governance is grounded in more than thirty years of clinical and operational leadership. His clinical background spans nursing and laboratory medicine. He has served as a C-suite executive across health systems, accreditation organizations, and government healthcare initiatives — with direct accountability for the quality, operational, and strategic functions that AI governance now demands.

Dr. Greenhill is a published researcher and author whose scholarly work includes 9 textbooks — among them *Introduction to Quality Management*, now in its fifth edition, published by Health Administration Press — along with 36 peer-reviewed publications and 183 scholarly citations. He serves as national faculty for the American College of Healthcare Executives, a board member and expert for the International Society for Quality in Healthcare (ISQua) and is the Principal and Founder of SmartSigma AI.

Contributing Sources

1. Muralidharan V, et al. A scoping review of reporting gaps in FDA-approved AI medical devices. npj Digital Medicine. 2024 Oct 3;7:273. DOI: https://doi.org/10.1038/s41746-024-01270-x Full text: https://www.nature.com/articles/s41746-024-01270-x

2. FDA AI-Enabled Medical Device List (updated periodically). https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices

3. Statistically Valid Post-Deployment Monitoring Should Be Standard for AI-Based Digital Health. June 2025. https://arxiv.org/abs/2506.05701

4. Chang W, Owusu-Mensah P, Everson J, Richwine C. Hospital Trends in the Use, Evaluation, and Governance of Predictive AI, 2023–2024. Office of the Assistant Secretary for Technology Policy. Data Brief: 80. September 2025 https://www.ncbi.nlm.nih.gov/books/NBK618497/

5. Schwamm LH, Jain SS, et al. Pragmatic Approaches to the Evaluation and Monitoring of Artificial Intelligence in Health Care: A Science Advisory From the American Heart Association. Circulation. November 2025. https://www.ahajournals.org/doi/10.1161/CIR.0000000000001400

6. 2010 Flash Crash — SEC/CFTC Joint Report. https://www.sec.gov/news/studies/2010/marketevents-report.pdf

7. Rethinking clinical trials for medical AI with dynamic deployments of adaptive systems. npj Digital Medicine. May 2025. https://www.nature.com/articles/s41746-025-01674-3

8. IHI Leadership Alliance. AI Governance: Maximizing Benefit and Minimizing Harm for Patients, Providers, and Health Systems. October 6, 2025. https://www.ihi.org/library/blog/ai-governance-maximizing-benefit-and-minimizing-harm-patients-providers-and-health

9. Understanding Silent Failures in Medical Image Classification. arXiv 2307.14729. August 2023. https://arxiv.org/abs/2307.14729

10. Keeping Medical AI Healthy: A Review of Detection and Correction Methods for System Degradation. arXiv 2506.17442. June 2025. https://arxiv.org/html/2506.17442v1

11. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366:447-453. https://www.science.org/doi/10.1126/science.aax2342

12. Design framework for trustworthy AI in healthcare — cardiovascular AI evidence. Science Direct. October 2025. https://www.sciencedirect.com/science/article/pii/S1566253525008747

13. Kodiak Solutions, "The Healthcare Waiting Game," RCA Benchmarking Analysis, December 2023

14. Scaling enterprise AI in healthcare: governance risk mitigation. PMC. 2025. https://pmc.ncbi.nlm.nih.gov/articles/PMC12075486/

