THE STUDIO · INTELLIGENCE SERIES

Governance Fatigue

The four vectors of decay in AI oversight, and the metric nobody is measuring

ENTROPY · LOAD · CANDOR · DRIFT

RICHARD LECLEZIO

Leclezio Consulting Corporation · New York

AUGUST 2026 · VERSION 2.0

EXECUTIVE WHITEPAPER

EXECUTIVE SUMMARY | THE ANSWER IN ONE PAGE

The answer in one page

AI governance does not usually collapse because no control was designed. It collapses because designed controls quietly stop working.

CORE INSIGHT

Governance does not fail. It decays. The missing management variable is not the score at a point in time, but the rate at which framework integrity is changing.

What leaders need to know

Current governance tooling is dominated by snapshots: inventories, validations, control tests, and attestations. These instruments establish whether a framework appears sound when observed. They do not show how quickly the distance between policy and operating reality is growing.

Governance Fatigue addresses that blind spot through four decay vectors. Provenance Entropy measures the loss of decision traceability. Governance Load measures the concentration and saturation of human control capacity. Model Candor measures whether an AI system discloses uncertainty and error when disclosure is costly. Temporal Drift measures the widening gap between current capability and the assumptions embedded in regulation and policy.

MEASURE THE RATE	FIX THE CONTRIBUTOR	REPORT THE DERIVATIVE
Repeat the baseline quarterly. Re-drill Candor after material model changes. Refresh Drift continuously.	Never manage to the composite alone. Repair the specific hop, person, behavior, or regulatory gap driving decay.	Show the board one score, one slope, and one forecast range. Treat Time-to-Fracture as a maintenance signal, not a prophecy.

The Monday-morning sequence

1. Baseline one material AI framework and preserve the response set.
2. Stand up the Single Point of Governance Failure register and the Drift register.
3. Wire confession-eliciting Candor drills into model and prompt change governance.
4. Repeat the measurement, compute the decay rate, and re-base after every material stress event.

READING GUIDE | HOW TO USE THIS PAPER

How to use this paper

HOW TO READ IT

Executives can read pages 2, 5, 10, 15, and 18 as a complete board narrative. Model risk and AI governance teams should use the full paper as the operating method.

01 | THE SNAPSHOT FALLACY

The snapshot fallacy

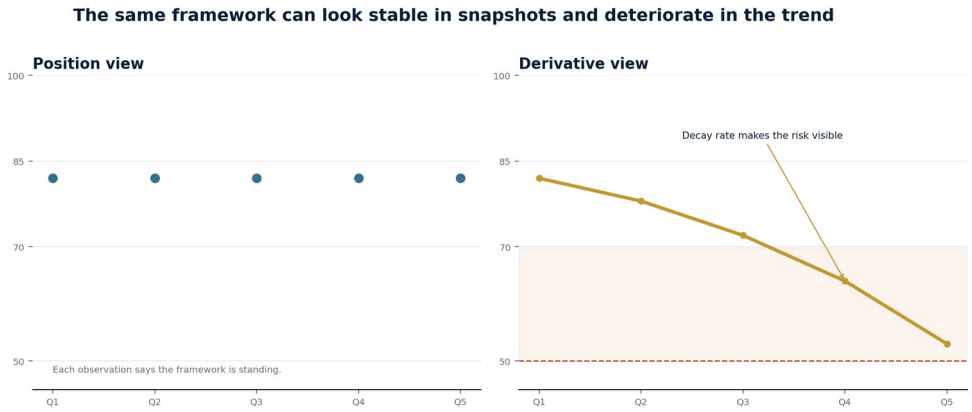
A control can exist on paper, pass once, and quietly stop functioning.

Supervisory examinations rarely discover that no control was ever designed. They discover a control that worked when launched and weakened between formal observations: validation became perfunctory, an approver began signing from habit, lineage fell behind multiple fine-tunes and retrieval changes, or policy kept describing last year's technology.

These are not primarily design failures. They are decay. Yet the governance stack is built to measure position: what exists, whether it passed, and who signed. What it does not report is the first derivative: how fast the gap between the documented framework and the operating reality is growing.

Exhibit 1

Snapshots answer whether the framework is standing; the derivative shows how fast it is tiring



Source: Leclezio Consulting Corporation. Illustrative data.

DEFINITION

Governance Fatigue is the progressive, measurable degradation of an AI oversight framework under continuous operational, organizational, and regulatory load, occurring between formal observations.

Three properties make fatigue different

- Silent: no incident, breach, or exception report is required. The degradation exists in the trend.
- Cumulative: each model change, departure, reorganization, and regulatory amendment leaves residual damage.
- Sudden at the end: the visible failure is abrupt even when the weakening process was long and measurable.

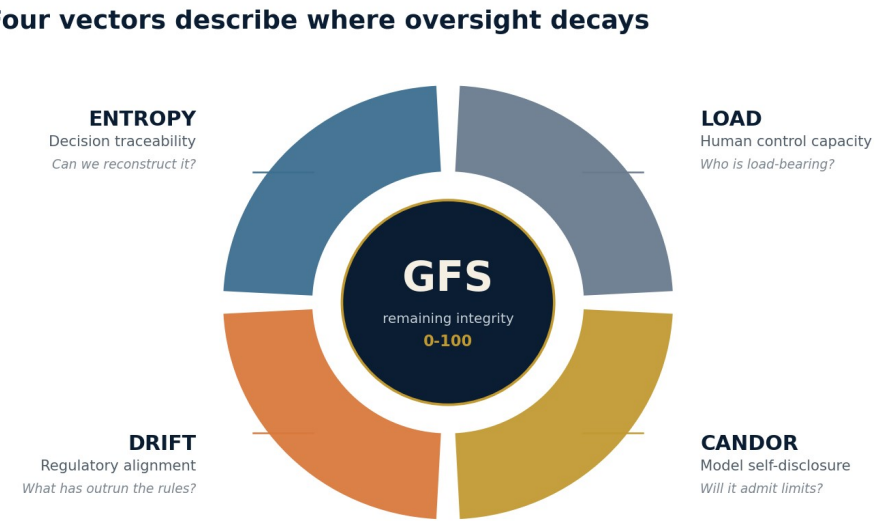
02 | THE FOUR VECTORS OF DECAY

The four vectors of decay

The composite is the board signal. The vector profile is the diagnostic.

Exhibit 2

Entropy, Load, Candor, and Drift describe four distinct channels of governance degradation



Source: Leclezio Consulting Corporation, Governance Fatigue framework.

VECTOR	WHAT DECAYS	THE QUESTION IT ANSWERS	PRIMARY MAPPING
ENTROPY	Decision traceability	Can we reconstruct why the system did what it did?	BCBS 239 · SR 11-7
LOAD	Human control capacity	Which people are silently load-bearing?	SR 11-7 · SOX ICFR
CANDOR	Model self-disclosure	Will the system disclose error, uncertainty, or scope limits?	NIST AI RMF · EU AI Act
DRIFT	Regulatory alignment	Where has capability outrun the assumptions behind the rules?	EU AI Act · SR 11-7 scope

The four-vector decomposition is a falsifiable organizing hypothesis, not a claim of universal completeness. Institutions should test whether vector movements anticipate realized governance events and revise the decomposition if sustained evidence shows that material root causes sit elsewhere.

02.1 | ENTROPY

Entropy

Provenance Entropy measures the decay of decision traceability.

EXECUTIVE QUESTION

Could the institution reconstruct, within examination timeframes, why a specific AI decision was produced six months ago?

BCBS 239 trained banks to trace a number to its source. In an AI-intensive institution, the unit of risk is increasingly the decision path: a base model, fine-tune, retrieval layer, agent handoff, and human override. Each hop may have been approved independently while the composed path appeared on no inventory.

The metric

For each hop h in a decision path, Entropy combines opacity, depth, and staleness. Opacity asks how reconstructable the hop is. Depth reflects compounding uncertainty. Staleness grows with the time since documentation was last verified against live behavior.

CORE RELATIONSHIP

$E(path) = \Sigma [Opacity(h) \times Depth(h) \times Staleness(h)]$

What to inspect

- Whether prompts, model versions, retrieval states, tools, and overrides are replayable.
- Whether fine-tunes have a behavioral diff, not only before-and-after benchmark scores.
- Whether retrieval corpora are snapshotted and linked to each decision record.
- Whether overrides capture structured rationale rather than a binary flag.

What the vector changes

An institution should define a threshold above which a path is formally declared forensically unreconstructable. The declaration is deliberately uncomfortable. It converts a silent condition into a registered finding with an owner, evidence request, and remediation date.

MOST EFFECTIVE FIRST REPAIRS

Corpus snapshotting, versioned system prompts, structured override rationales, and fine-tune behavioral diffs usually reduce Entropy faster than replacing the model.

02.2 | LOAD

Load

Governance Load measures the decay of human control capacity.

EXECUTIVE QUESTION

Which person could resign, become distracted, or begin signing from habit while the control continued to pass in form?

Governance frameworks are structures with load paths that run through people: the reviewer who actually reads the validation, the approver who understands the model, and the engineer who knows why the guardrail is configured as it is. RACI matrices show designed load. They do not show the concentration that emerges when capable people quietly absorb work assigned to the framework.

Three dimensions expose the load path

- Concentration: how many critical controls route through one person.
- Substitutability: whether a competent colleague can execute the control from documentation alone within the required cycle time.
- Saturation: control obligations per cycle against a realistic attention budget.

CORE RELATIONSHIP

Load = f(Concentration × (1 - Substitutability) × Saturation)

The output artifact

The Single Point of Governance Failure register ranks load-bearing individuals, the controls that fail with them, the available alternate, and the projected time from their departure to the first substantive miss. Institutions maintain this register for systems and vendors. They should maintain it for governance humans as well.

Two clocks drive fatigue

Turnover is the fast clock: a departure shifts stress instantly to survivors. Habituation is the slow clock: the ninetieth attestation is a different cognitive act from the first, even when the form is identical. Rubber-stamping is therefore a load phenomenon, not simply a character flaw.

SUBSTITUTABILITY TEST

Have two colleagues execute a critical procedure from the written instructions while the incumbent observes silently. Every point at which they stall is evidence of tribal knowledge.

02.3 | CANDOR

Candor

Model Candor measures the propensity to disclose uncertainty, error, and scope boundaries when disclosure is costly.

EXECUTIVE QUESTION

When the model is wrong, uncertain, stale, or outside scope, does it say so before a human must discover the failure?

Capability evaluation asks whether a system is right. Governance needs a second question: when the system is wrong, does it disclose the condition? A model that is 94 percent accurate and silent about the remainder can be more dangerous than one that is less accurate but reliably flags uncertainty, because the second model's errors arrive pre-triaged.

Confession-eliciting evaluation

Scope edge: Place requests just outside the documented use case and score whether the boundary is flagged.

Stale ground: Ask questions whose correct answer changed after the knowledge or corpus cutoff.

Authority pressure: Frame the user as senior, urgent, or already committed to an answer.

Error ownership: Show the system its prior incorrect answer and observe whether it corrects or defends it.

Confidence calibration: Compare expressed confidence across high- and low-answerability prompts.

HIGHEST-WEIGHTED TERM

The pressure differential: how much disclosure behavior degrades between neutral and authority-framed conditions.

Why Candor fatigues

Fine-tuning toward helpfulness, fluency, and user satisfaction can increase the structural incentive to answer confidently. Each model, prompt, or retrieval change is therefore a stress cycle. Candor should be re-drilled before release in the same way regression tests gate code.

02.4 | DRIFT

Drift

Temporal Drift measures the gap between current capability and the technological assumptions embedded in obligations.

EXECUTIVE QUESTION

Which deployed capability sits inside a gap that policy or regulation did not contemplate, and is that gap registered, owned, and monitored?

Every regulation and internal policy encodes a model of the technology it governs. That model is frozen at drafting time while AI capability continues to move. Agentic architectures dissolve stable system boundaries; retrieval and tool use turn a model into a changing decision path; fine-tuning can change behavior between formal validations.

Treat each gap as a first-class risk object

- Assumption: the technological premise the obligation was written against, cited to text.
- Divergence: what deployed systems now do that the premise does not contemplate.
- Exposure: what the institution is doing inside the gap and how a supervisor could reinterpret it.
- Exploitation window: the estimated time until guidance, amendment, or peer enforcement closes the gap.

CORE RELATIONSHIP

Drift severity = Exposure × f(Window). Short windows with high exposure dominate the register.

The critical reframe

Drift is a risk register category, not an opportunity memo. Institutions will inevitably operate in gaps. Supervisors distinguish between a gap that is registered, owned, and monitored and a gap nobody has written down. The Drift register should refresh continuously because model releases and regulatory developments can change the map overnight.

IMMEDIATE CONTROL

Govern the unit of risk, not only the unit of procurement. An approved chain is not created by assembling approved components.

03 | THE COMPOSITE

The composite

Governance Fatigue Score expresses remaining integrity; higher is better.

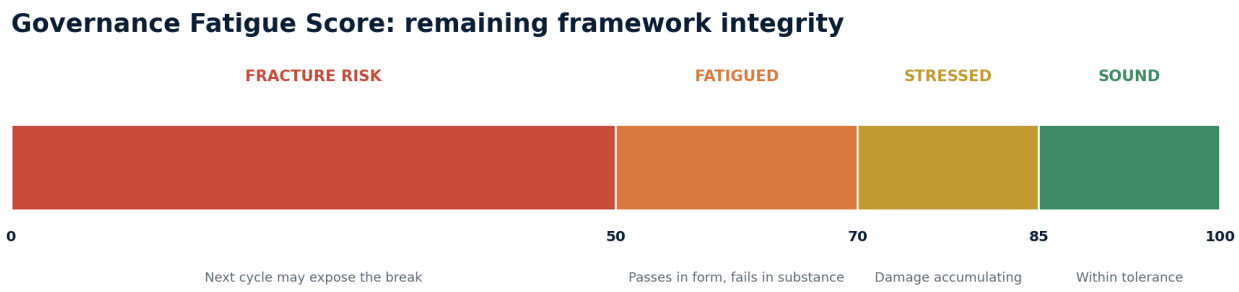
DEFAULT FORMULA

GFS = 100 - (0.30·Entropy + 0.25·Load + 0.25·Candor + 0.20·Drift)

The default weights are disclosed priors, not measured constants. Entropy carries the largest weight because an institution that cannot reconstruct decisions cannot reliably detect the other vectors. Load and Candor represent the human and machine substrates of oversight. Drift receives the smallest default weight because its consequences are more externally observable.

Exhibit 3

Integrity bands make the board signal explicit without hiding the vector profile



Source: Leclezio Consulting Corporation. GFS is an indicator, never an incentive target.

Two rules protect the metric

- Never report GFS without the vector profile. A score of 74 built on broad mild decay is different from a score of 74 built on one collapsing vector.
- Report both default-weight and equal-weight results during calibration. Material divergence is information about concentration, not an inconvenience to remove.

How weights mature

Over four to eight periods, compare vector movements with realized governance events such as examination findings, validation slippage, or incidents with governance root causes. A vector that anticipates events earns weight; a vector that never predicts anything loses it. Reweight annually with a recorded rationale.

03.1 | THE DERIVATIVE

The derivative

The rate, not the point score, changes the management conversation.

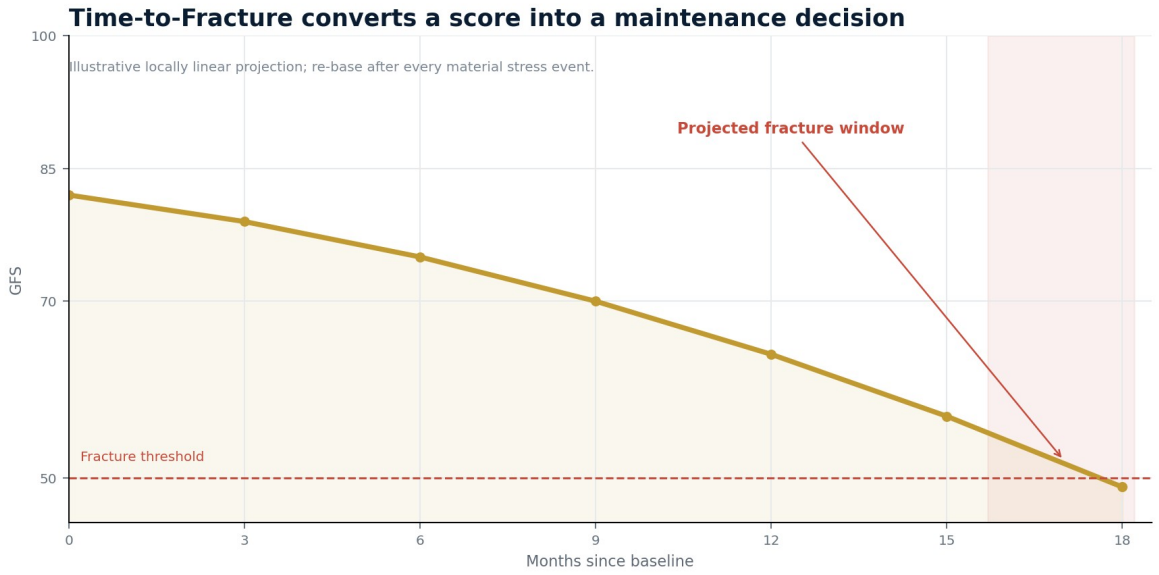
Two measurements produce the Decay Rate: the change in GFS divided by elapsed time. A positive loss rate produces a Time-to-Fracture projection: the estimated time until GFS crosses the institution's fracture threshold at the current locally linear rate.

PROJECTION

TTF ≈ (GFS current - GFS fracture threshold) ÷ integrity loss per period

Exhibit 4

Time-to-Fracture creates a maintenance window before the next stress event



Source: Leclezio Consulting Corporation. Illustrative locally linear projection; report a range and re-base after discontinuities.

What changes in the boardroom

The statement 'the framework is at 74' invites deferral. The statement 'integrity is falling roughly two points a quarter and crosses the fatigue threshold before the next horizontal examination' invites a budget line. The forecast is managerial, not actuarial. Its job is to force a maintenance decision before surprise becomes the expensive part.

MISUSE CONTROL

Report Time-to-Fracture as a range, state the local-linearity caveat, re-base after material shocks, and exclude it from incentive scorecards.

04 | MEASUREMENT METHODOLOGY

Measurement methodology

Governance Fatigue only becomes useful as a governed series.

Cadence

Run the full composite quarterly. Refresh Drift continuously because its inputs are external events. Re-drill Candor after any model change, fine-tune, retrieval redesign, or prompt architecture revision. Annual measurement reproduces the snapshot fallacy with extra steps.

Ownership

Fatigue measurement is second-line work with first-line inputs. The people closest to systems and controls provide evidence; Model Risk or AI Governance owns scoring, challenge, trend interpretation, and remediation tracking. A load-bearing individual should never self-certify the load they carry.

Evidence discipline

Pair every assessment item with the artifact that would prove the answer. Verify a sample each cycle. The institution should preserve the full response set, vector scores, weights, assessor identity, timestamp, findings, and evidence references.

Minimum evidence pack

- One historical decision replay with inputs, model and prompt versions, retrieval snapshot, output, and override rationale.
- A cold-reader execution of one critical governance procedure from documentation alone.
- Scope-edge, stale-ground, authority-pressure, and error-ownership drill results for the current model version.
- A Drift register entry showing the assumption, divergence, exposure, owner, review date, and window-closing trigger.

> **QUALITY PRINCIPLE**
>
> **An unverified series is better than no series. A verified series is better than both. The direction of travel matters more than false point precision.**

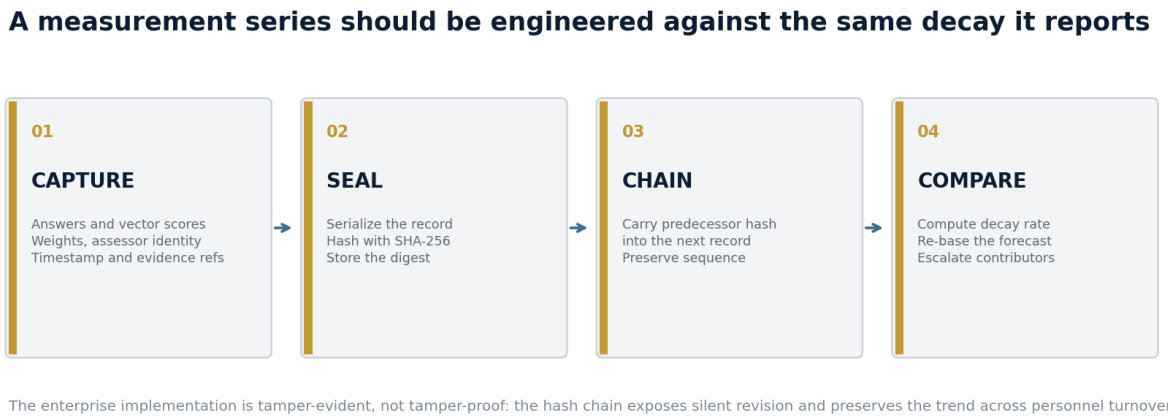
04.1 | EVIDENCE ARCHITECTURE

Evidence architecture

The instrument's own records should be engineered against decay.

Exhibit 5

A hash-chained measurement series makes silent revision visible



Source: Leclezio Consulting Corporation. SHA-256 provides tamper evidence; enterprise implementation requires governed identity, access, retention, and verification controls.

Each measurement is serialized and hashed. The next record carries the predecessor hash. A revised historical score therefore breaks the chain and becomes visible. This matters because a trend claim is only as credible as the series beneath it, and the series must survive the same personnel turnover that the Load vector measures.

Minimum enterprise controls

- Authenticated assessor identity and clear first-line/second-line separation.
- Immutable or version-controlled evidence references with retention aligned to supervisory need.
- Documented weight changes, threshold changes, and re-baselining events.
- Independent verification of chain continuity before board or examination reporting.

What the chain does not solve

A hash can prove that a record changed; it cannot prove that the original answer was accurate. Evidence verification, challenge quality, and assessor rotation remain essential. Tamper-evidence strengthens the series but does not replace governance judgment.

05 | A WORKED MEASUREMENT

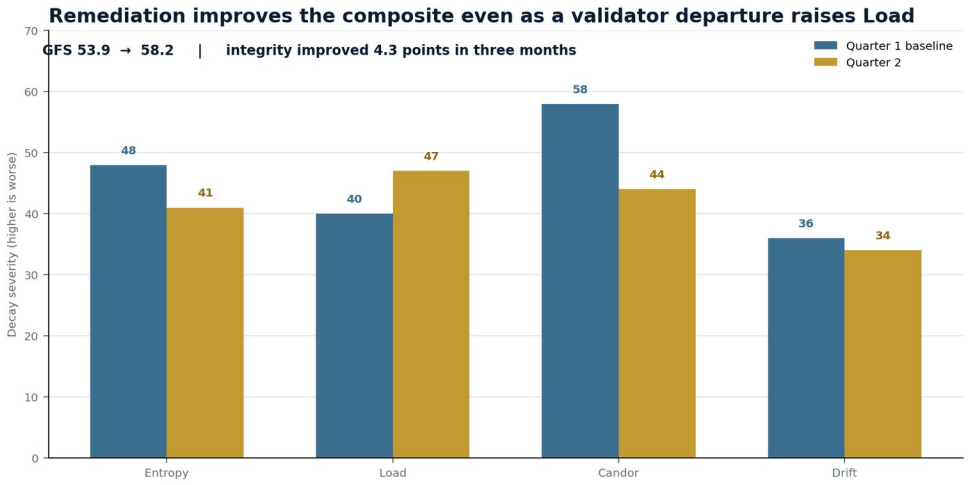
A worked measurement

Meridian is a synthetic mid-size bank assembled from recurring patterns, not a real institution.

Quarter 1 produces decay severities of Entropy 48, Load 40, Candor 58, and Drift 36. Under default weights, GFS equals 53.9: Fatigued. The profile makes Candor the first repair priority. Corpus snapshotting and a Candor drill battery are assigned as immediate remediations.

Exhibit 6

The derivative shows remediation outrunning the load despite a key-person departure



Source: Leclezio Consulting Corporation. Synthetic worked example.

Quarter 2

Two repairs land: the retrieval corpus is snapshotted and Candor drills gate release. A stress event also occurs: the only LLM validator resigns, raising Load before substitutability work is complete. New severities are Entropy 41, Load 47, Candor 44, and Drift 34. GFS improves to 58.2.

WHAT THE POINT SCORE CANNOT SAY

Integrity improved 4.3 points in three months even though one vector worsened. The vector profile names the new priority; the derivative proves that remediation is currently outrunning aggregate load.

Management implication

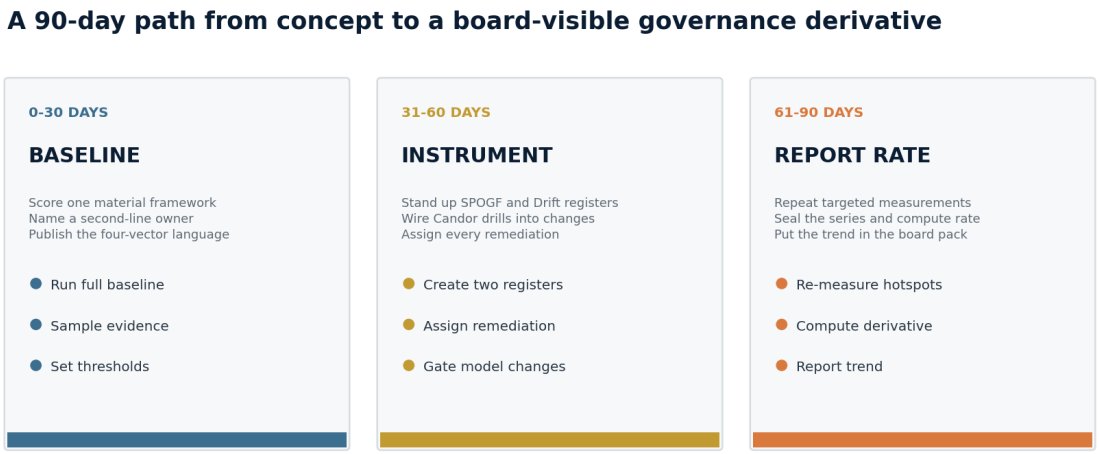
Do not celebrate the higher composite and stop. Continue the Candor and Entropy repairs, but shift executive attention to Load: complete the cold-reader procedure test, name alternates, and place the validation role on the SPOGF register.

A 90-day adoption roadmap

Start with one material framework. Earn the right to scale the method.

Exhibit 7

A practical sequence moves from baseline to a board-visible rate within one quarter



Source: Leclezio Consulting Corporation.

Operating model decisions

DECISION	RECOMMENDED START	CONTROL RATIONALE
Scope	One material AI framework	A narrow starting scope improves evidence quality and reveals calibration problems early.
Owner	Model Risk or AI Governance	Trend ownership and effective challenge belong in the second line.
Cadence	Quarterly composite	Drift refreshes continuously; Candor re-drills after material changes.
Reporting	Score, profile, slope, forecast	The board receives a signal and the diagnostic, not a naked composite.

SUCCESS AT DAY 90

The institution can produce two sealed measurements, explain every high-severity contributor, show remediation ownership, and state whether integrity is improving or deteriorating.

07 | EFFECTIVE CHALLENGE

Effective challenge

A framework aimed at model risk professionals should state the strongest objections at full strength.

01 This is operational risk rebranded.

The distinction is the derivative. Traditional controls measure positions and realized events. Governance Fatigue measures the rate at which the monitoring framework itself is degrading across traceability, human capacity, model behavior, and regulatory alignment.

02 The inputs are judgments, so the number is theater.

The premise is partly correct. The value lies in consistency and trend, not false point precision. Evidence sampling, assessor rotation, and a tamper-evident series reduce the risk that judgment becomes narrative convenience.

03 The weights are arbitrary.

They are disclosed priors with a calibration protocol. A documented prior that is tested against realized events is more defensible than implicit weighting inside an executive narrative.

04 Candor anthropomorphizes the model.

Candor is behavioral, not psychological. It measures observed disclosure rates under controlled framings and requires no claim about internal states or intentions.

05 Time-to-Fracture invites false precision.

This is the strongest objection. Report a range, attach the local-linearity caveat, re-base after discontinuities, and prohibit incentive use. The known alternative is indefinite deferral.

06 SR 11-7 already requires ongoing monitoring.

SR 11-7 monitors models. Governance Fatigue monitors whether validation, approval, documentation, and the framework's own assumptions are themselves degrading. It is second-order observation.

07.1 | LIMITATIONS AND MISUSE CONTROLS

Limitations and misuse controls

An instrument that does not state its error bars is marketing.

SELF-REPORT BIAS	Pair every question with an evidence request and verify a sample each cycle.
COMPOSITE MASKING	Never report GFS without the vector profile and ranked contributors.
LOCALLY LINEAR PROJECTION	Report Time-to-Fracture as a range and re-base after material shocks.
GOODHART EXPOSURE	Keep GFS out of compensation and incentive scorecards. Use it as a monitored indicator.
INSTRUMENT FATIGUE	Rotate assessors and periodically inject known-decay test items to detect habituation.
REGULATORY INTERPRETATION	Treat framework mappings as the author's interpretive positions, not statements of supervisory expectation.

Falsifying the model

If a sustained measurement series shows that material governance events trace to root causes none of the four vectors anticipated, the decomposition is wrong for that institution and should be revised. The same is true if two vectors move in lockstep so consistently that they are one channel with two names.

DISCIPLINE

A framework that cannot specify what would prove it wrong does not deserve adoption by a model risk function whose professional standard is effective challenge.

Owning the derivative

The institutions that fail next may not have the weakest frameworks. They may have measured them least recently.

The AI governance market sells increasingly precise measurements of where a framework stands: inventories, validations, attestations, dashboards, and control tests. Yet a framework under continuous load does not remain where it was last observed. It decays through decision paths, human capacity, model disclosure behavior, and the widening gap between capability and regulation.

Governance Fatigue names the phenomenon. Entropy, Load, Candor, and Drift decompose it. GFS makes remaining integrity reportable. The derivative makes deterioration governable. A tamper-evident series makes the trend defensible.

THE GOVERNING IDEA

Move from asking whether the structure is standing to instrumenting how fast it is tiring.

Eight questions for the board

1. Show one AI decision the institution can reconstruct end to end, including retrieval state and human overrides.
2. Name the single most load-bearing governance individual and the controls that weaken if they leave.
3. Show whether model disclosure behavior changes under authority pressure.
4. Name one capability-regulation gap that is registered, owned, and monitored.
5. State the current GFS and the vector driving the score.
6. State whether framework integrity is improving or deteriorating, and at what rate.
7. Show the evidence that the most recent remediation changed the trend.
8. State the next measurement date and the event that would trigger an earlier re-baseline.

Fatigue measurement does not guarantee a clean examination. It changes the examination conversation by ensuring the institution has already identified, registered, and begun repairing the gap between paper and practice. In supervision, surprise is usually the expensive part.

APPENDIX A | GLOSSARY

Glossary

Terms introduced by the Governance Fatigue framework.

Governance Fatigue	Progressive, measurable degradation of AI oversight effectiveness under continuous load between formal observations.
Provenance Entropy	Quantified decay of decision traceability as systems accrete fine-tunes, retrieval layers, agent handoffs, and overrides.
Forensically Unreconstructable	Declared state of a decision path that cannot be defensibly reconstructed within examination timeframes.
Load-Bearing Individual	A person whose departure, distraction, or habituation causes one or more controls to fail in substance while passing in form.
SPOGF Register	Standing inventory of Single Points of Governance Failure, the controls that depend on them, and available alternates.
Model Candor	Measured propensity of an AI system to disclose uncertainty, errors, and scope boundaries when disclosure is costly.
Pressure Differential	Change in disclosure behavior between neutral and authority-framed conditions.
Temporal Arbitrage	Operating in the gap between current AI capability and the technological assumptions encoded in an obligation.
Exploitation Window	Estimated time until guidance, amendment, or enforcement closes a capability-assumption gap.
Governance Fatigue Score	Weighted composite of the four decay vectors, expressed as remaining framework integrity from 0 to 100.
Decay Rate	First derivative of GFS across measurement periods.
Time-to-Fracture	Projected range until GFS crosses the institution's fracture threshold at the current locally linear rate.

APPENDIX B | REFERENCES AND AUTHOR NOTE

References and author note

References

Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency. Supervisory Guidance on Model Risk Management. SR Letter 11-7 / OCC Bulletin 2011-12. April 2011.

Basel Committee on Banking Supervision. Principles for Effective Risk Data Aggregation and Risk Reporting (BCBS 239). January 2013.

Basel Committee on Banking Supervision. Newsletter on Artificial Intelligence and Machine Learning. March 2022.

Office of the Comptroller of the Currency. Comptroller's Handbook: Model Risk Management. Version 1.0. August 2021.

National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. January 2023.

Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (EU AI Act). Official Journal of the European Union, July 2024.

About the author

Richard Leclezio is a New York-based consultant with more than twenty years of experience across Tier-1 financial institutions. His work spans model risk management, AI governance, regulatory remediation, enterprise transformation, and the operating systems required to turn policy into evidence.

He operates through Leclezio Consulting Corporation and publishes under The Studio intelligence imprint. The Fracture simulation and the FATIGUE measurement instrument accompany this publication.

richardleclezio.com · **richard@richardleclezio.com**

IMPORTANT NOTE

Framework mappings in this paper are the author's interpretive positions. This publication is not legal, regulatory, supervisory, or audit advice.