AI GOVERNANCE MEASUREMENT FRAMEWORK

Version 1.0.1
Public Reference Standard

DOI
10.5281/zenodo.17552324

Author

Russell Parrott

Release Date
7 November 2025

Canonical Repository
https://github.com/russell-parrott/Standardized-Definition-of-AI-Governance

CREATIVE COMMONS ATTRIBUTION–NONCOMMERCIAL–NODERIVATIVES 4.0 INTERNATIONAL (CC BY-NC-ND 4.0) LICENCE.

This license lets people share this work but not change it or make money from it. Every permission and restriction flows from those three core terms:

Attribution (BY)

Anyone who shares this work must give the author proper credit. That means naming the author exactly as specified, linking to the original source if available and stating that the work is under the CC BY-NC-ND 4.0 license. The credit must be visible wherever the work appears. They cannot claim authorship, hide the author's name or present it as their own material.

NonCommercial (NC)

The work cannot be used for commercial advantage or monetary compensation. That includes:

- Selling it directly or bundling it inside something sold.
- Using it in a paid course, subscription or membership product.
- Republishing it on a platform or in marketing that generates profit, sales leads or brand value.

Even indirect benefit, such as using this work to promote a service, attract customers or enhance a commercial brand, counts as commercial use.

Educational, research or personal sharing that brings no financial or promotional gain is allowed.

NoDerivatives (ND)

The work must be shared exactly as it is, with no alterations.

No one may remix, translate, excerpt, adapt or build upon it. That includes editing text, combining it with other works or producing "summaries" that reuse your material. They can distribute it only in its complete, unmodified form, with the author's and license intact.

WHAT YOU CAN AND CANNOT DO

What people can do

- Download, copy and redistribute this work as long as they credit the author.
- Share it privately or publicly, online or offline.
- Use it in teaching, study, or non-profit communication, provided it's unchanged and unmonetised.

What people cannot do

- Sell, license or otherwise profit from it.
- Include it in a commercial product, campaign or event.
- Edit, adapt, translate or partially quote it as if it were their own.
- Repost it without attribution or without the license notice.

EXECUTIVE SUMMARY

AI Governance is the system of principles, policies, processes and accountability mechanisms that direct and control the entire lifecycle of artificial-intelligence systems, from design through decommissioning, to ensure that they operate lawfully, safely and in alignment with human values, stakeholder rights and societal objectives.

Purpose

AI Governance exists to:

- Align AI outcomes with declared legal, organizational and societal objectives
- Preserve meaningful human agency and authority
- Prevent, detect and remedy harm across technical, social and systemic dimensions

Core Conditions

Effective AI Governance demonstrates:

- **Accountability** – decisions and outcomes traceable to named human or institutional responsibility
- **Transparency** – processes, data sources and decision pathways explainable and auditable
- **Human Oversight** – authorized personnel able to intervene, override or halt AI actions
- **Risk Management** – continuous identification and mitigation of technical, social and systemic risks
- **Ethical Foundations** – fairness, privacy, non-discrimination and respect for human rights
- **Legal Compliance** – operation consistent with applicable law and regulatory obligations
- **Lifecycle Continuity** – governance applied consistently across design, development, deployment, operation, modification and decommissioning

THE PROBLEM

Unverifiable AI governance costs economies **~0.1% of GDP annually** through:

- Job displacement without transition (NTAD)
- Public cleanup costs (PRE)
- Delayed adoption due to lack of trust (VAG)

The Framework

Seven measurable dimensions of structural trust (CER, PSI, PER, LCR, TSI, CRS, MGI) that predict where economic leakage occurs.

The Evidence

- **Singapore financial sector** (US$76.6B output): Estimated losses of **US$76.6M annually**
- **Germany** (€4.66T GDP): Estimated losses of **€4.6B annually**
- **UK** (£2.9T GDP): Estimated losses of **£2.9B annually**

The Solution

Make these seven dimensions mandatory compliance metrics. Institutions that improve scores from 0.7 to 0.95 can recover millions annually.

The ROI

Example institution recovered **$7.5M/year** with ~5-month payback on remediation investment.

THE SYSTEM IN SIMPLE TERMS

What It Is

"The System" is a way to measure how much control, trust and accountability exist inside AI-driven societies or industries, and how much money is lost when those controls fail.

It's built from two parts:

1. **Seven Dimensions of Structural Trust** – seven tests that show whether governance is working
2. **Three Loss Channels** – three ways those failures turn into real economic damage

Together they turn abstract ethics and "AI principles" into something measurable, like checking a country's blood pressure or a company's carbon footprint.

The Seven Dimensions (The "Health Checks")

Each dimension looks at one part of how an AI system interacts with people, law, and evidence:

If these seven dimensions score high, AI governance is solid. If they score low, trust collapses and the country starts losing value.

The Three Loss Channels (Where Money Actually Disappears)

When the seven dimensions fail, the damage shows up in three simple, visible places:

1. **Job loss (NTAD)** – when automation happens with no retraining or transition plan
    - → People lose income that could have stayed in the economy
2. **Public clean-up cost (PRE)** – when governments must spend money fixing AI messes
    - → Taxpayers pay for legal cases, investigations, and damage control
3. **Trust slowdown (VAG)** – when organisations delay using AI because they don't trust it
    - → Growth and innovation slow down until safety and traceability improve

Add them up and you can estimate the total economic loss from weak governance.

How It Works in Practice

1. **Measure the seven dimensions** – e.g., how traceable decisions are, how stable data is, how often responsibility breaks at borders
2. **Identify where the scores fall below safe thresholds** – those weak spots create the losses
3. **Quantify the losses through the three channels**: lost wages, public clean-up spending, and missed productivity
4. **Sum them up** and express the result as a share of GDP or sector output
    - → That number is the "Governance Loss Index", the real cost of untraceable AI

What It Shows

Even small failures in governance can cost a country around **0.1% of GDP every year**; millions or billions of dollars that simply vanish through inefficiency, mistrust, and unaccountable automation.

Why It Matters

This system changes the conversation from moral debate to measurable economics:

- It treats governance as infrastructure, not red tape
- It gives regulators and policymakers a way to prove when AI control is saving or wasting national value
- It lets citizens and investors see whether AI systems are genuinely trustworthy, or just pretending to be

Summary

The system measures whether AI is operating under real, traceable control and if not, it calculates how much that failure costs in jobs, money and lost confidence. It turns "ethics" into economics and "trust" into something you can test.

SEVEN DIMENSIONS OF STRUCTURAL TRUST

Who Measures Them

Think of three distinct but connected layers of measurement. Each layer checks the one above it, that's how integrity is kept traceable.

No single actor owns the whole picture. Each layer keeps the other traceably honest.

How the Measurements Are Taken

Each dimension has a specific, testable signal. They don't rely on self-reporting; they use evidence that can be cross-checked.

Each check can be recorded, timestamped, and made part of a traceable compliance chain, so there's no "we'll check later."

Where It Happens

- **Inside organisations**: through automated governance dashboards tied to their AI systems (many can log these indicators in real time)
- **Across sectors**: through regulatory sandboxes or registries that collect comparable data — like MAS's Veritas or the EU's AI Act conformity assessments
- **At national and international levels**: through standards bodies or AI observatories (OECD, ISO, IEEE, or new "Structural Trust Councils") that aggregate and publish results

This layered approach creates traceable accountability:

- Companies measure themselves daily
- Regulators cross-check periodically
- Oversight bodies ensure the whole framework isn't being gamed

Why This Structure Is Defensible

- It mirrors financial supervision: firms self-report under Basel or Solvency rules, regulators audit, and independent boards verify the system
- It fits existing law: most jurisdictions already demand explainability, traceability, and auditability — this simply standardises how those are measured and proven
- It ensures no single party controls the narrative — the audit trail itself is the governance

Summary

The "health checks" are measured by everyone in the chain; operators measure themselves, regulators verify and a higher oversight layer ensures the measuring process itself can't be corrupted.

That's how trust becomes traceable, testable, and enforceable instead of just promised.

THE THREE LOSS CHANNELS

NTAD – No-Transition Avoidable Displacement

What It Is

The economic loss that occurs when automation replaces people without a lawful, traceable transition plan — no retraining, redeployment, or duty of care.

Why It Matters

When governance is weak, firms externalise the cost of automation. Displaced workers lose income; the state absorbs the fallout through welfare and lost tax revenue. That is a governance failure, not a market one.

How It's Calculated

NTAD = A × [Number of displaced workers × Average annual wage × (1 − Re-employment rate)] × (1 + κ_1 × (1 − N))

Where:

- A = AI adoption factor — share of the workforce or sector exposed to AI-driven automation
- N = Native custody factor — proportion of AI systems under domestic legal and technical control
- κ_1 = small amplification constant (typically 0.05–0.20) reflecting how weak local custody increases transition delays or enforcement gaps

Example:

NTAD = 0.4 × [6,000 × €20,000 × (1 − 0.5)] × (1 + 0.1 × (1 − 0.6)) = €26,400,000

It measures avoidable wage loss, income that would remain circulating if transition duties were enforced and systems were locally accountable.

PRE – Public Rectification Expenditure

What It Is

The cost to the state of cleaning up AI failures that lacked a clear authority chain or audit trail — investigations, regulatory actions, court time, cybersecurity responses, compensation funds.

Why It Matters

Every euro spent fixing preventable AI incidents is a sovereign subsidy of private recklessness. Proper traceability and liability continuity (PSI + LCR) would have stopped those costs before they reached taxpayers.

How It's Calculated

PRE = A × [Regulatory budgets + Legal administration costs + Compensation payouts + Cybersecurity and investigation spending] × $(1 + \kappa_2 \times (1 - N))$

Where:

- A = AI adoption factor — proportion of national or sectoral activity reliant on AI
- N = Native custody factor — proportion of AI systems under domestic jurisdiction
- κ_2 = amplification constant (typically 0.20–0.50) reflecting how foreign custody multiplies public clean-up cost when PSI and LCR break across borders

Example:

PRE = 0.5 × [€8,000,000 + €6,000,000 + €5,000,000 + €4,000,000] × $(1 + 0.3 \times (1 - 0.5))$ = €13,125,000

VAG – Verification and Adoption Gap

What It Is

The productivity and output lost because organisations, sectors or citizens delay or limit AI use due to lack of verifiable trust — they can't prove safety, reliability or accountability.

Why It Matters

This is the hidden drag on growth. When systems can't be audited (low PSI/LCR/TSI), decision-makers slow or block deployment. Governance clarity converts that hesitation into economic gain.

How It's Calculated

VAG = Σ ($GDP_s \times A_s \times$ Adoption $gap_s \times$ Share of gap caused by lack of verifiable $safety_s \times$ Potential $unlock_s$) × $(1 + \kappa_3 \times (1 - N_s))$

Where:

- A_s = AI adoption factor per sector
- N_s = Native custody factor per sector
- κ_3 = amplification constant (typically 0.10–0.30) reflecting how foreign custody lowers confidence and prolongs delay

Example:

VAG = (€28,000,000,000 × 0.4 × 0.12 × 0.55 × 0.10) × (1 + 0.2 × (1 − 0.6)) = €296,352,000

Summary of Loss Channels

- **NTAD** exposes how ungoverned automation leaks private income
- **PRE** captures how missing traceability leaks public money
- **VAG** measures how lack of trust leaks future growth

Together they quantify the full cost of untraceable AI governance, now scaled by AI adoption (A) and custody (N), showing not only how systems fail, but how far those failures reach across economies.

THE FORMULAS

Core Principle

Loss = f(Trust failure × Exposure)

Untraceable governance causes economic loss that can be decomposed and measured. Exposure depends on how widely AI is used and where that AI is controlled.

Base Formula

Annual Governance Loss (AGL) = NTAD + PRE + VAG

Where:

- NTAD = No-Transition Avoidable Displacement (lost wages)
- PRE = Public Rectification Expenditure (clean-up cost)
- VAG = Verification & Adoption Gap (lost productivity due to lack of trust)

All three depend on the Seven Dimensions of Structural Trust (SD_1…SD_7).

- When the system is healthy, SD scores approach 1
- When governance is weak, they drop, creating loss

Expanded Structural Form

$$AGL = A \times N \times [\alpha_1(1 - PSI) + \alpha_2(1 - LCR) + \alpha_3(1 - CER) + \alpha_4(1 - PER) + \alpha_5(1 - TSI) + \alpha_6(1 - CRS) + \alpha_7(1 - MGI)]$$

Where:

- **A** = AI adoption factor – share of national or sector GDP that actually depends on AI systems in a way that can cause loss (Example: if 40% of production uses AI, A = 0.40)
- **N** = Native custody factor – share of those AI systems under domestic legal and technical control (Example: if only 60% of those systems are hosted or governed locally, N = 0.60)
- Each **$(1 - SD_i)$** represents a trust deficit — the fraction of operations failing that dimension
- Each **α_i** is a weight translating that deficit into an economic effect (in money or GDP points)

Thus A × N scales the model by real exposure and control:

Countries with high adoption but low native custody lose more when governance fails.

Linking Back to GDP

Governance Loss Ratio (GLR) = AGL / GDP

For your base case:

GLR = 0.001 = 0.1%

Meaning: for every €1,000,000,000 of GDP, €1,000,000 leaks through untraceable AI governance.

When adoption and custody are known, refine it as:

$$GLR = (A \times N \times \Sigma\, \alpha_i(1 - SD_i)) / GDP$$

If those data are unavailable, retain the conservative floor of 0.1% GDP.

Recovery Equation (The "Good" Case)

Governance Recovery Value (GRV) = $AGL_{(baseline)} - AGL_{(remediated)}$

Or equivalently:

GRV = GDP × ($GLR_{(baseline)} - GLR_{(remediated)}$)

Example from Institution A:

- Baseline GLR = 0.0105% (≈ US$8M / US$76B)
- After remediation GLR = 0.00065% (≈ US$0.5M / US$76B)
- GRV = 76B × (0.000105 − 0.0000065) ≈ **$7.5M/yr**

That's the Return on Governance — the cash recovered by improving PSI and LCR.

Simple Operational Shortcut

For communication, you can use the practitioner's version:

AGL % of GDP = β × A × N × (1 − $S\bar{D}$)

Where:

- **$S\bar{D}$** = average of the seven trust scores
- **β** ≈ 0.7–1.0 (empirical scaling constant from your baseline model)
- **A** = AI adoption factor (portion of GDP driven by AI)
- **N** = native custody factor (portion of AI systems under local control)

Example:

If the average trust score is 0.85, adoption A = 0.4, and native custody N = 0.6,

AGL % of GDP = 0.8 × 0.4 × 0.6 × (1 − 0.85) = 0.0288% GDP ≈ the conservative lower bound once adjusted for real exposure.

Interpretation

The formula treats trust deficits as economic leak coefficients, scaled by how much of the economy actually runs on AI and how much of that AI is locally controllable.

It allows regulators, companies, or nations to calculate both:

- **Loss function**: what ungoverned AI costs now
- **Recovery function**: what improved governance returns later

Final Condensed Form

$AGL = A \times N \times \Sigma(i=1 \rightarrow 7)\ \alpha_i \times (1 - SD_i)$

And:

$GLR = AGL \div GDP$

Or verbally:

"Annual governance loss equals the weighted sum of trust deficits across the seven dimensions, scaled by AI adoption and domestic custody, expressed as a share of GDP."

That remains the formal backbone of the doctrine, governance treated as a measurable fiscal variable, not a moral claim.

KEY TAKEAWAYS

---

1. **The System quantifies AI governance failures as economic loss** – making abstract ethics measurable
2. **Seven dimensions measure structural trust** – from traceability to legal continuity
3. **Three loss channels capture the damage** – jobs, cleanup costs, and delayed adoption
4. **The baseline cost is ~0.1% of GDP annually** – scaled by adoption and custody factors
5. **Improving governance scores delivers measurable ROI** – institutions can recover millions per year
6. **The framework is auditable and defensible** – mirroring financial supervision models

This transforms AI governance from philosophical debate into financial infrastructure.