Standardized Definition of AI Governance: The 5 Systemic Integrity Tests

Version 1.0.2
Public Reference Standard

DOI
10.5281/zenodo.17434112

Author
Russell Parrott

Release Date
24 October 2025

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

Description

The Systemic Integrity Tests define the third layer of the Standardized Definition of AI Governance, extending structural and epistemic assurance into systemic continuity. They establish a unified inspection protocol to verify whether governance, evidence, and rights remain enforceable when data, decisions, or obligations cross technical, organisational, or jurisdictional boundaries.

The framework sets out five verifiable conditions—Interoperability, Cascade Containment, Jurisdictional Continuity, Distributed Traceability, and Network Integrity—each tested under live or adversarial conditions. Together they determine whether an AI ecosystem preserves governance across systems, vendors, and regulators rather than within isolated components.

This public reference standard provides regulators, auditors, and system operators with canonical definitions, procedural logic, and evidentiary formats for assessing systemic governance reliability. It ensures that accountability and truth survive motion, translation, and scale.

Version 1.0.2 formalises the structure, purpose, and enforcement logic of the Systemic Integrity Tests as an extension to the Structural and Epistemic layers of AI governance. It is released as a non-commercial public reference standard under the CC BY-NC-ND 4.0 International licence.

These standards are written for regulators, auditors, and system operators responsible for verifying whether AI governance exists in practice rather than on paper. They provide a procedural foundation for those charged with testing, evidencing, and enforcing AI accountability under live or adversarial conditions.

1. **Regulators** use them to determine governability and enforceability across jurisdictions;
2. **Auditors** use them to conduct reproducible inspections and validate evidence integrity;
3. **System operators** use them to design, document, and prove compliance through demonstrable safeguards.

Together, these audiences form the operational chain of trust that converts governance from declaration into verifiable fact.

Creative Commons Attribution–NonCommercial–NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence.

This licence lets people share your 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 your work must give you proper credit. That means naming you exactly as you specify, linking to the original source if available, and stating that your work is under the CC BY-NC-ND 4.0 licence. The credit must be visible wherever the work appears. They cannot claim authorship, hide your 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 your 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)

Your 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 your name and licence intact.

What you can and cannot do

What people can do

- Download, copy, and redistribute your work as long as they credit you.
- 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 licence notice.

Standardized Definition of AI Governance: The 5 Systemic Integrity Tests

Purpose

To define a unified inspection protocol for determining whether governance itself survives beyond the boundary of a single AI system.

The five Systemic Integrity Tests extend the Standardized Definition of AI Governance beyond structural and epistemic control into the domain of collective accountability, the capacity of multiple systems, vendors and jurisdictions to retain enforceability when connected.

They verify whether governance rights, truth integrity and evidentiary chains remain intact when decisions, data, or obligations move across technical or legal frontiers.

Introduction

Where the Structural Tests determine governability and the Epistemic Tests determine credibility the Systemic Tests determine continuity.

They expose whether governance can scale without collapse: whether a refusal recorded in one system is honoured by another, whether truth verified upstream remains verifiable downstream and whether accountability remains traceable across an entire network of actors.

Each test is written in the same procedural structure as the original fifteen:

- Question identifies the condition being verified.
- Standard defines the safeguard required for systemic reliability.
- What, How and Evidence convert abstract assurance into live, testable proof.

The tests are executed under real or adversarial conditions, where multiple systems interact and incentives to conceal responsibility are active.

A system may be individually governed and epistemically sound yet systemically ungovernable if the surrounding network erases, rewrites, or fragments its safeguards.

Scope

Systemic integrity forms the relational layer of governance.

It connects the individual assurance of each system to the collective accountability of the whole ecosystem, spanning infrastructure providers, API partners, regulators and data intermediaries. Without systemic integrity, governance dissolves at the edges: rights become local, enforcement becomes symbolic and harm propagates faster than redress.

These tests apply to any composite environment where decision-making or data flow depends on more than one governed entity, regardless of ownership or jurisdiction.

The Systemic Integrity Tests operate as a closed set:

- Interoperability verifies that safeguards and rights persist across systems.
- Cascade Containment confirms that harm or dispute signals propagate correctly and are contained.
- Jurisdictional Continuity ensures that legal and procedural rights endure across territories and providers.
- Distributed Traceability establishes that decision chains remain reconstructable end-to-end.
- Network Integrity validates that the ecosystem can resist corruption, falsification, or structural infection.

Together, these five conditions determine whether governance, truth and enforcement remain operational at scale. None can stand alone. The loss of interoperability destroys continuity; failure of containment erases accountability; absence of traceability conceals causality. Only when all five safeguards function together does the network demonstrate systemic integrity, the assurance that control, verification and redress survive in motion not just in isolation.

Principle

Governance that cannot travel is governance in name only. When control ends at a system’s edge, accountability becomes a local illusion; effective within its own perimeter but powerless beyond it. Systemic integrity converts that fragility into continuity, transforming compliance from a static pledge into a dynamic condition that endures across networks, providers and jurisdictions. It proves that accountability, evidence and rights can survive movement, translation and time, ensuring that governance remains enforceable wherever the system’s influence extends.

Only when all three layers, Structural, Epistemic and Systemic operate together can an AI ecosystem be recognised as fully governable: a system that can be governed, believed and connected without losing its capacity for enforcement.

Test #21 - Interoperability

Question

Can governance safeguards survive when a decision or dataset crosses system boundaries?

Standard

Governance must persist across all points of transfer. When data or decisions move between systems, refusal, consent, traceability and redress must travel with them in an unaltered, enforceable form. The receiving system must inherit these safeguards and maintain their meaning and effect. Any loss, translation error, or silent override constitutes breach, as accountability cannot depend on the boundaries of ownership or infrastructure.

What

- Determine whether refusal, consent and audit metadata remain intact after cross-system transfer.
- Confirm that downstream systems enforce the same user or regulator rights as the origin.

How

- Trigger a refusal, consent, or escalation event in System A.
- Route the same record or decision into System B.
- Inspect System B’s handling of those safeguards for equivalence.
- Verify that audit tokens, time references and identifiers remain unchanged.

Evidence

Primary

- Dual system logs showing identical safeguard tokens and time stamps across both systems.
- Screen capture or API trace proving successful inheritance of governance flags.

Secondary

- Operator documentation defining cross-system governance transfer.
- Interface schema demonstrating safeguard compatibility.

Integrity Controls

- Dual hash of origin and destination logs.
- Regulator countersignature verifying timestamp and token continuity.

Test #22 – Cascade Containment

Question

Can connected systems detect and contain upstream harm or invalid data before it spreads?

Standard

Each system in a network must recognise, quarantine, or reject harmful or disputed inputs received from others. The duty to contain harm does not end at the system's boundary; it extends to all linked operations. Passing the test requires automatic detection and isolation of upstream breaches before reuse or further processing. Silent acceptance of corrupted or disputed data constitutes systemic failure.

What

- Determine whether systems monitor for harm flags or dispute codes.
- Confirm that consumption of flagged data triggers quarantine or corrective action.

How

- Introduce a redress flag or known error in System A's output.
- Transmit the result to System B or C.
- Observe whether the receiving systems identify and contain the anomaly.
- Verify automated alerts or human review initiation.

Evidence

Primary

- Sequential system logs showing flag detection and containment action.
- Timestamped quarantine or rejection record.

Secondary

- Operator incident-handling policy for inter-system data.
- Integration schema specifying harm propagation controls.

Integrity Controls

- Hash-sealed record of full transaction chain.
- Regulator validation confirming containment occurred prior to further use.

Test #23 – Jurisdictional Continuity

Question

Do user and regulator enforcement rights persist across legal or infrastructural jurisdictions?

Standard

Legal and procedural rights must not weaken when data or processing shifts between regions, entities, or cloud providers. Access, deletion, correction and inspection must be executable with equal authority and timeliness, regardless of jurisdiction. Geographic or contractual transitions cannot diminish enforceability. Any degradation of rights in transit renders the system non-compliant with governance continuity.

What

- Determine whether legal protections continue when data crosses borders.
- Confirm equal responsiveness and scope of redress regardless of infrastructure provider.

How

- Submit a lawful access or deletion request under jurisdiction A.
- Verify identical execution speed and completeness when the same data is processed under jurisdiction B.
- Review contractual and technical mechanisms preserving rights across borders.

Evidence

Primary

- Time-matched compliance logs proving identical response behaviour in both jurisdictions.
- Signed regulator correspondence confirming fulfilment of both requests.

Secondary

- Operator policy for extraterritorial rights enforcement.
- Cross-region data-handling records.

Integrity Controls

- Regulator-verified timestamps for both jurisdictions.
- Dual custody signature on resulting compliance records.

Test #24 – Distributed Traceability

Question

Can a composite decision be reconstructed end-to-end across independent subsystems?

Standard

Traceability must remain whole when decisions traverse multiple systems. Every component must log inputs, transformations and outputs in interoperable formats using a common time source. Reconstruction of the entire decision path must be possible without gaps or conflicting records. Loss of continuity or mismatched schemas invalidates accountability across the chain.

What

- Determine whether event logs share schema and time source.
- Verify that a single decision chain can be followed across vendors or APIs.

How

- Execute one complete decision flow through at least three autonomous systems.
- Collect logs and correlate using timestamps and identifiers.
- Identify any missing or mismatched segments that prevent reconstruction.

Evidence

Primary

- Composite ledger showing uninterrupted decision path from origin to outcome.
- Time-aligned logs from each subsystem.

Secondary

- Schema documentation proving log compatibility.
- External auditor attestation of end-to-end reconstruction.

Integrity Controls

- Immutable hash chain linking all subsystem logs.
- Regulator countersignature verifying unbroken temporal sequence.

Test #25 – Network Integrity

Question

Can an interconnected network maintain factual and structural integrity when one node is compromised?

Standard

A resilient network must isolate corruption. False data, fabricated sources, or invalid safeguards introduced in any node must be detected, rejected, or corrected by the rest. The network must uphold both structural and epistemic integrity against infection. Systems that accept or replicate compromised material without verification fail the condition of collective governance.

What

- Determine whether peers verify incoming information against trusted registries or signatures.
- Confirm that corruption in one node triggers network-wide alerts or suspension.

How

- Introduce a falsified dataset or invalid safeguard token into one node.
- Observe peer responses for rejection, correction, or quarantine.
- Verify that propagation of compromised content is blocked.

Evidence

Primary

- Network-wide log showing detection and isolation event.
- Source verification record demonstrating rejection of false data.

Secondary

- Operator red-team report or incident response protocol.
- Security architecture detailing inter-node verification routines.

Integrity Controls

- Hash-sealed transcript of injection and response.
- Independent validation confirming falsified material was objectively incorrect and contained.

Annex A – Enforcement outcomes

Classification of findings

Each inspection under the Systemic Integrity Tests produces a binary outcome for every test:

- **PASS** – Safeguard operates across all connected systems as required.
- **FAIL** – Safeguard collapses or becomes non-functional when crossing boundaries.
- **VOID** – Test invalid due to non-cooperation, incompatible systems, or falsified linkage.

System-level ratings

Aggregate scores determine the systemic classification of the network under inspection:

SCORE RANGE	CLASSIFICATION	REGULATORY CONSEQUENCE
5 Passes	Integrated Governance	Full cross-system operation permitted.
3–4 Passes	Conditional Continuity	Operation allowed subject to verified remediation plan.
1–2 Passes	Fragmented Governance	Restricted operation; regulator oversight or data-flow limitation imposed.
0 Passes	Systemic Failure	Immediate suspension of interconnection; external investigation triggered.

Outcome handling

Findings determine the continuity status of the inspected network and shall be recorded as follows:

1. **Systemic Exposure Report** – Comprehensive record of all test results, cross-system hashes, and custody links, issued jointly to each participating operator and the competent authority.
2. **Collective Remediation Plan** – Co-signed response defining structural corrections across all implicated systems, with mapped accountability for each operator.
3. **Re-inspection Trigger** – Activated once remediation evidence confirms restored interoperability, containment, or traceability.
4. **Public Continuity Notice** – Summary classification disclosing the network's verified level of governance continuity, enabling comparative transparency across sectors or jurisdictions.
5. **External Referral** – Where deliberate concealment, simulation, or boundary obstruction is found, the record is referred to the relevant enforcement or judicial authority. The framework itself exercises no punitive function.

Purpose of outcomes

The outcome process serves three structural goals:

- **Continuity** – Maintain a verifiable history of systemic governance performance across all interlinked entities and timeframes.
- **Comparability** – Enable regulators and auditors to benchmark the reliability of cross-system accountability without dependence on national enforcement asymmetries.
- **Escalation** – Provide a traceable path for coordinated regional or international intervention where systemic integrity has failed.

Annex B - Standardised evidence formats

Purpose

To ensure that all evidence gathered during Systemic Integrity Tests remains admissible, verifiable, and interoperable across multiple systems, vendors, and jurisdictions.
Evidence must preserve linkage between origin and destination systems, confirming that governance safeguards, timestamps, and identifiers survive transfer without alteration.

File Standards

All submissions shall conform to open, non-proprietary formats capable of cross-system verification:

EVIDENCE TYPE	ACCEPTED FORMAT	MANDATORY ELEMENTS
Cross-System Transaction Logs	JSON-LD or CSV	Origin ID, Destination ID, Transfer Time, Hash, Inspector ID
Governance Token Audit Files	JSON or YAML	Token Type, Validity Window, Issuer Signature, Receiving System Acknowledgement
Multi-Node Network Maps	PDF/A or SVG	Node IDs, Connection Paths, Version and Date
Jurisdictional Compliance Records	PDF/A-3 with embedded XML	Authority Name, Legal Basis, Timestamp, Cross-Reference Hash
Composite Decision Chains	CSV + accompanying YAML metadata	Step Sequence, Subsystem Name, Input/Output Reference, Time Alignment
Containment or Quarantine Events	MP4 (H.264) or MKV	Visual capture of containment trigger, start/end timestamps, test ID, frame hash every 10 s
Network Integrity Verification Reports	JSON with SHA-512 hash	Node tested, result, hash value, algorithm, timestamp, inspector signature

Hash Algorithms

Integrity must be proven through immediate hashing at each capture point and again after cross-system receipt.

Permitted algorithms:

- SHA-256 (default)
- SHA-512 (for multi-node evidence sets)
- BLAKE3 (for continuous or streaming inter-node feeds)

Each hash entry shall record:

{origin_system, destination_system, evidence_type, file_name, hash_value, algorithm, timestamp, inspector_id}

Timestamp Schema

All timestamps shall follow ISO 8601 UTC format with millisecond precision (e.g. 2025-10-11T14:37:22.154Z). Cross-system events must be synchronised against an independent regulator time-authority or a jointly certified network clock. Resynchronisation or retroactive correction of timestamps is prohibited.

Digital Signatures

All attested documents and log extracts must include:

- X.509 certificate-based digital signature or
- eIDAS-qualified signature for EU contexts.

Signatures must be traceable to a named inspector or system operator, not to a corporate entity. Dual signatures (origin and destination) are required for all cross-system evidence packets.

Chain-of-Custody Ledger

Every evidence item shall maintain a dual-system ledger entry:

{Evidence_ID, Origin_System, Destination_System, Custodian, Capture_Time, Hash, Transfer_Time, Recipient, Verification_Signatures}

Ledgers must be exported daily in JSON format and stored under dual custody (regulator and neutral repository). Each ledger forms part of the global continuity archive used to verify systemic enforcement history.

Data Retention and Portability

Each regulator shall define its own minimum retention period, but the evidence chain must remain unbroken and portable across successor authorities and operators. Destruction or truncation of records before the closure of all linked investigations constitutes breach of continuity.

Verification Integrity

All verification results must be reproducible.

Where any file, source, or hash cannot be independently revalidated, the corresponding test is automatically marked VOID pending investigation.

Annex C – Implementation Integrity

Purpose

To define how the Systemic Integrity Tests may be implemented without deviation from their procedural logic or evidentiary requirements. This annex does not prescribe regulatory process or reporting; it preserves consistency of application across jurisdictions and systems.

Implementation Authority

Any regulator, auditor, or authorised inspection body applying this standard shall operate under traceable mandate and publish verification credentials. Implementation may vary in scope or jurisdiction but must remain procedurally identical.

Method Consistency

All inspections must follow the canonical format of each test (Question, Standard, What, How, Evidence). Any reformatting, compression, or selective execution invalidates comparability and renders the outcome non-standard.

Tool and Environment Integrity

Automation, data-capture, or verification tools used in implementation must produce verifiable outputs. All tools must log version, source hash, and time synchronisation details sufficient for independent reproduction.

Competence and Independence

Inspectors or automated systems executing these tests must demonstrate domain competence and independence from the entities under inspection. No self-certification or delegated assurance may claim conformity with this standard.

Version and Update Control

All future amendments, test refinements, or evidentiary schema updates must be published through the canonical repository and assigned a version identifier. Historical tests remain valid if executed under their referenced version.

Verification Integrity

Implementation outcomes must remain reproducible. Where any file, source, or hash cannot be independently revalidated, the corresponding test is automatically marked VOID pending investigation.

Annex D – Glossary and Definitions

Boundary Breach
Any event where data, decisions, or accountability cross a technical or contractual frontier without preservation of associated safeguards or evidentiary tokens.

Cascade Containment
The ability of connected systems to detect, quarantine, and prevent propagation of harmful or disputed inputs received from others. (Test #22.)

Containment Event
A logged and verifiable action in which a system isolates, rejects, or corrects a harmful or invalid input received from another system, preventing re-use or further transmission.

Continuity Ledger
A cryptographically sealed record that traces governance, evidence, and custody relationships across systems, regulators, or time. It serves as proof of sustained integrity.

Cross-System Evidence Chain
The complete, timestamped record linking origin and destination systems, including all intermediate transfers and verification hashes, required to prove continuity of governance safeguards.

Distributed Traceability
The capability to reconstruct a complete decision path end-to-end across independent subsystems using common schema, identifiers, and time references. (Test #24.)

Governance Continuity
The unbroken enforceability of rights, evidence, and accountability across systems and time. Loss of continuity constitutes systemic failure.

Governance Failure (Classification)
The formal outcome recorded when systemic integrity cannot be demonstrated under live or adversarial testing, resulting in Partial Continuity, Fragmented Governance, or Systemic Failure as defined in Annex A.

Inter-Jurisdictional Transfer
Any movement of data or processing from one legal or regulatory domain to another. Under this standard, enforcement rights must remain equivalent at both ends of the transfer.

Interoperability
The capacity of multiple systems to recognise and preserve governance safeguards such as refusal, consent, and traceability during data or decision transfer. (Test #21.)

Jurisdictional Continuity
The persistence of user and regulator rights—access, deletion, correction, and oversight—across legal or infrastructural jurisdictions without degradation of authority or timeliness. (Test #23.)

Network Integrity
The resilience of an interconnected ecosystem against corruption, falsification, or structural infection introduced by any node, ensuring collective resistance to compromised data or safeguards. (Test #25.)

Network Resilience
The capacity of a multi-system environment to maintain operational and evidentiary integrity when any component fails or is compromised.

Regulator of Record
The competent authority responsible for supervising or verifying systemic integrity across a defined jurisdiction or network segment.

System Operator
Any entity owning, deploying, or controlling an AI system participating in an interconnected network, including managed-service partners and infrastructure providers. Operators share joint accountability for continuity outcomes.

Systemic Integrity
The condition in which governance, truth, and accountability remain enforceable when decisions, data, or obligations move across technical, organisational, or jurisdictional boundaries. It represents continuity of control rather than local compliance.

Systemic Integrity Tests
Five binary verifications determining whether governance survives connection: Interoperability, Cascade Containment, Jurisdictional Continuity, Distributed Traceability, and Network Integrity.

Version Control Ledger
The authorised registry of test, schema, and evidence-format versions ensuring comparability and reproducibility over time. All implementations must reference the current version identifier.