Authority Displacement Governance Overlay

Authority Displacement in Artificial Intelligence Governance:
Measuring and Preserving Human Decision Control

Kathryn G. Varnes
kathryn.varnes@athospherellc.com
Athosphere LLC, Phoenix, AZ, USA
February 2026

Executive Summary

Artificial intelligence systems may improve statistical performance while simultaneously degrading an institution's practical ability to intervene, override, and reassume control. As override frequency declines, review times compress, and infrastructure automation deepens, manual competence can atrophy. Under model failure, vendor outage, distribution shift, or adversarial attack, institutions that cannot manually reconstitute core workflows may enter **processing paralysis**, in which systems continue to execute but institutions cannot effectively intervene.

The NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0) establishes a structured lifecycle approach for identifying, evaluating, and managing risks associated with artificial intelligence systems. Organized around the core functions of Govern, Map, Measure, and Manage, it has significantly advanced AI governance practice in the United States by providing institutions with a common vocabulary and operational structure for risk oversight. However, while the AI RMF systematically addresses harm—probability, severity, bias, safety, robustness, and security—it does not explicitly measure the migration of decision-making authority from human actors to computational systems. This paper argues that authority displacement constitutes a distinct governance dimension requiring independent evaluation.

The Authority Displacement Governance Overlay (ADGO) addresses this gap.
ADGO introduces authority displacement as a distinct governance dimension: the degree to which operational decision power, institutional competence, and override capacity migrate from accountable human actors to algorithmic systems.

A system may demonstrate near-perfect statistical performance, reduce operational costs, and pass bias, robustness, and security testing, while simultaneously degrading institutional capacity to govern, override, and reassume authority. This paper terms this condition Performance–Authority Divergence. Performance–Authority Divergence describes the observable operational condition in which authority migration produces latent fragility despite acceptable risk metrics; it does not replace authority displacement as the underlying governance construct.

Operational Continuity and Systemic Fragility. Authority displacement is also a continuity risk. Where manual competence atrophies and override capacity decays,

institutions become operationally brittle. Under model failure, vendor outage, distribution shift, or adversarial attack, systems that cannot be manually reconstituted may **enter processing paralysis** or otherwise fail to sustain continuous operations within required recovery tolerances. ADGO treats authority preservation as a resilience requirement, not merely an ethical preference. For adopters, ADGO operationalizes a simple question: can the organization keep operating—and regain control—when the AI fails, shifts, or must be suspended?

ADGO extends the AI RMF by introducing:

1. A Five-Tier Authority Displacement Spectrum
2. A Delegation Legitimacy Test (DLT)
3. Escalation triggers grounded in operational metrics
4. Drift instrumentation thresholds
5. Audit floor requirements
6. Revocability Recovery Time Objectives (RTOs)
7. Resilience-Adjusted Delegation Value (RADV) to evaluate optimization claims under stress and restoration conditions
8. An Agentic Escalation Rule
9. Expressive authority classification standards
10. An Infrastructure Authority Drift Overlay requiring Tier classification to reflect structural authority across data, compute, persistence, and deployment layers.

ADGO functions as a structured overlay—complementary to, not in conflict with, the AI RMF.

I. Intellectual Context and Theoretical Foundations

AI delegation resembles aviation automation. Modern aircraft are highly automated, yet the pilot remains legally responsible. Experience has shown that as automation increases, manual proficiency and situational awareness can erode unless deliberately maintained. Aviation responded not by removing automation, but by instituting recurrent training, override testing, and explicit doctrine to preserve human authority. AI governance faces the same structural challenge: performance may improve, but unless override remains exercised and competent, human authority becomes nominal.

ADGO builds upon established strands of research across human factors, philosophy of technology, administrative governance, and AI ethics.

First, automation bias literature demonstrates that human operators systematically over-rely on automated decision aids, even when errors are present (Parasuraman & Manzey, 2010). Bainbridge's "Ironies of Automation" (1983) further explains that as automation improves, human operators become less practiced and less capable of intervention precisely when intervention becomes necessary.

Second, philosophical work on "meaningful human control" (Santoni de Sio & van den Hoven, 2018) argues that human oversight must retain genuine causal and normative significance. Formal review without substantive capacity is insufficient.

Third, literature on deskilling and skill atrophy in human–machine systems shows that reduced engagement in complex tasks degrades operator competence over time, increasing systemic fragility.

Fourth, regulatory developments—such as high-risk system oversight provisions in the EU AI Act—implicitly recognize that certain AI deployments require enhanced supervision and fallback mechanisms.

ADGO synthesizes these strands and translates them into an operational governance framework. Its contribution lies not in identifying automation risk per se, but in measuring and institutionalizing authority migration as a distinct and quantifiable governance variable. ADGO is intended primarily for consequential decision systems (Tier III–V as defined in section III) and is not designed to regulate low-impact informational tools or purely assistive analytics.

ADGO introduces structured continuity safeguards into automation deployment. These safeguards function as resilience constraints designed to preserve override capacity, institutional competence, and revocability under stress conditions. In highly competitive environments, this may appear to disadvantage early adopters who prioritize efficiency. ADGO's position is that for **Tier III–V systems**, resilience and revocability are deployment conditions for continuous operations under stress. Where delegation increases dependency, continuity is the governing constraint: the organization must be able to suspend, restore, and operate within defined recovery tolerances.

II. Authority Displacement as a Distinct Governance Dimension

A. Conceptual Clarification

Artificial intelligence systems are embedded in consequential domains including:

- Eligibility determinations
- Credit approvals
- Medical triage
- Infrastructure management
- Regulatory prioritization
- Employment screening

The AI RMF evaluates systems through risk metrics such as safety, bias, reliability, privacy, and robustness. These dimensions are necessary but not sufficient.
Authority displacement refers to the functional migration of decision-making power from accountable human actors to algorithmic systems.

This migration may occur through:

- Automation bias
- Reduced override frequency
- Interface design asymmetry
- Increasing workflow dependency
- Skill atrophy
- Vendor lock-in
- Loss of reconstructable reasoning

Authority displacement may increase even as measurable risk decreases.
ADGO therefore requires authority to be evaluated independently of harm metrics.

B. Operational Indicators of Authority Displacement
Authority displacement should be measured through observable indicators:

1. **Override Vitality Ratio** – Percentage of AI outputs that are rejected, materially modified, or superseded by independent human reasoning prior to execution. *(Indicator of whether review remains substantive rather than symbolic.)*
2. **Human Independent Reasoning Index** – Percentage of decisions where human reasoning meaningfully diverges from AI output prior to exposure. *(Indicator of preserved independent judgment.)*
3. **Review Time Distribution** – Statistical distribution of human review time. Sustained review times below calibrated cognitive plausibility thresholds indicate nominal confirmation rather than independent evaluation.
4. **Dependency Exposure Score** – Percentage of workflows inoperable without AI. *(Indicator of operational brittleness under suspension conditions.)*
5. **Manual Competence Retention Rate** – Demonstrated capacity to replicate decisions without AI assistance. *(Validated empirically through structured testing mechanisms.)*

Cognitive Plausibility Threshold refers to empirically derived minimum review-time benchmarks below which independent human evaluation is physiologically implausible; sustained review times below this threshold indicate symbolic oversight and trigger escalation review. Thresholds shall be calibrated by task type and reviewed at least annually or upon material workflow change.

Meaningful divergence refers to independent variation in key reasoning steps or outcome determination prior to AI exposure.

Manual Competence Validation Protocol
Manual Competence Retention Rate shall be validated through structured testing mechanisms proportionate to tier classification. Validation must be empirical rather than declarative.
Approved validation mechanisms include:

1. Shadow-Mode Suspension Drills
   Periodic limited-duration suspension of AI assistance for a defined workflow

sample to verify that personnel can execute decisions independently within applicable Recovery Time Objectives.

2. Injected Validation Cases
   Known-condition test cases inserted into live workflows to evaluate independent reasoning, edge-case handling, and resistance to automation complacency. Personnel must not be informed in advance which cases are synthetic.
3. Cognitive Plausibility Enforcement
   Automated detection of review times below empirically calibrated cognitive plausibility thresholds indicating symbolic oversight. Sustained violations trigger escalation review.

Institutions are not required to duplicate all workflows manually; however, demonstrable competence testing must occur at defined intervals.

Minimum Validation Cadence

- Tier III — At least quarterly
- Tier IV — At least monthly
- Tier V — Sector-defined cadence, but no less than monthly, unless deviation is justified in writing by the independent reviewer

Cadence may be increased based on drift indicators, incident history, or dependency exposure scores.

Workflow Sampling Standard

A statistically significant workflow sample shall be sufficient to detect material competence degradation trends at the workflow level. Samples must:

- Include edge-case classifications
- Include historically high-impact or high-consequence decision types
- Reflect distributional variation in input complexity
- Permit trend comparison across validation cycles

Sampling methodology shall be documented in governance records and subject to independent audit review.

Performance–Authority Divergence Scenario. Consider a Tier III automated underwriting system that demonstrates low observed harm, passes bias and robustness testing, and reduces processing time by 60%. Over time, override frequency approaches zero, review times compress to symbolic confirmation, and manual adjudication staff are reallocated. During an unforeseen macroeconomic distribution shift, model outputs deteriorate. If institutional personnel cannot resume manual adjudication within defined Recovery Time Objectives, loan processing halts. Revenue disruption, liquidity strain, and competitive disadvantage follow. ADGO evaluates and mitigates this failure mode before crisis conditions reveal it.

Behavioral indicators measure observable operator interaction patterns. However, authority may migrate structurally even where behavioral metrics remain nominal. Infrastructure-layer entrenchment—such as immutable event propagation, cross-system

write authority, automated retraining loops, or CI/CD deployment automation—can materially constrain human control independent of override frequency. For Tier III–V systems, structural authority displacement shall be evaluated in conjunction with behavioral indicators in accordance with Annex X, which is binding for Tier determination where structural conditions constrain practical human control.

III. The Five-Tier Authority Displacement Spectrum

Tier classification establishes a minimum governance requirement. Institutions may voluntarily classify systems at higher tiers.

The quantitative thresholds below are illustrative baseline triggers. Institutions should calibrate thresholds based on sectoral risk, statutory requirements, and empirical benchmarking.

Authority classification must be empirical rather than declarative. Escalation triggered by metrics is automatic unless independently reviewed. Authority displacement intensity increases as scope, autonomy, irreversibility, systemic reach, and dependency exposure increase.

Authority Displacement Intensity = f(Scope, Autonomy, Irreversibility, Systemic Reach, Dependency Exposure).

Infrastructure-layer conditions are binding for Tier determination. Where structural automation at the data, compute, persistence, or deployment layer renders human intervention nominal in practice, structural classification supersedes behavioral indicators. Structural entrenchment across these layers shall elevate Tier classification in accordance with Annex X.

The Five-Tier Spectrum operationalizes this principle through tiered governance requirements and escalation triggers.

Tier I — Informational Systems
Advisory only. Override routine and frictionless.
Escalation Trigger:
Automatic Tier II classification if AI output is auto-populated into official records without explicit human confirmation.

Tier II — Influential Systems
Ranking or scoring systems influencing human decisions.
Tier II thresholds function as early-warning indicators of symbolic oversight and are not presumptions of wrongdoing.
Automatic escalation to Tier III occurs if any of the following thresholds are sustained across a specified evaluation window, defined as meeting the threshold in two consecutive

measurement periods within the window (e.g., two consecutive monthly measurements within a 90-day window), unless an independent reviewer documents a justified exception.

- Override rate < 2% over rolling 90 days
- Average review time < empirically calibrated cognitive plausibility interval
- Rejecting output requires additional UI friction
- Human–AI concordance > 98% in discretionary contexts

The evaluation window and measurement periods shall be specified in governance documentation and shall not exceed 90 days unless independently justified.

Tier III — Determinative Systems
Outputs executed unless reversed.
Governance Requirements:

- Mandatory Delegation Legitimacy Test
- Drift instrumentation
- Resilience-Adjusted Delegation Value (RADV) evaluation (Optimization Track where applicable)
- Defined RTO
- Audit floor compliance

Tier IV — Autonomous Operational Systems
Characteristics:

- Multi-step workflow execution
- Cross-system write authority
- Persistent operational state
- API chaining
- Tier IV applies where cross-system write authority permits AI outputs to modify downstream systems, persistent records, or operational state without a documented human checkpoint prior to consequential execution.

Governance Requirements:

- RTO ≤ 4 hours (sector-adjusted)
- Annual suspension drill validation
- Vendor dependency exposure index
- Tier reassessment upon scope expansion

Strategic Displacement Exception. In environments requiring sub-second response (e.g., cybersecurity containment, algorithmic trading), continuous human steering may be impracticable. In such cases, legitimacy depends on bounded autonomy and demonstrable revocability.

Required safeguards include:

- Absolute kill-switch capability
- Clearly defined containment boundaries

• Pre-authorized operational envelopes specifying scope, permissible actions, and escalation boundaries for consequential actions
• Verified Recovery Time Objectives

Strategic displacement is permissible only where revocability is executable without delay, without exclusive reliance on vendor intervention, and where operational envelopes are formally documented and independently reviewable.

Tier V — Sovereign or Irreversible Systems
Systems affecting deprivation of liberty, coercive enforcement authority, use-of-force mechanisms, constitutional or civil status, or critical infrastructure in ways that may cause large-scale irreversible harm.
Tier V classification shall also apply where autonomous retraining, dynamic threshold modification, or deployment automation materially alters decision logic absent documented human authorization prior to consequential execution.
Governance Requirements:
• Independent review authority
• Extraordinary performance justification
• Sector-defined critical RTO
• Explicit sovereign boundary analysis documenting why delegation does not erode fundamental institutional authority

IV. Delegation Legitimacy Test (DLT)

Delegation of authority for Tier III–V systems is legitimate only if:

1. Structural Governance Insufficiency OR Extraordinary Performance Superiority
2. Revocability in practice
   Revocability must be executable across infrastructure layers, including data ingestion, persistence propagation, and deployment automation pathways. (See Annex X.)
3. Reconstructability to the defined audit floor
   Reconstructability must account for distributed system artifacts and propagation layers affecting replay, traceability, and rollback feasibility. (See Annex X.)
4. Demonstrable competence preservation
5. Compliance with sovereign boundary constraints

Failure on any element invalidates the delegation.

V. Entry Justification and Resilience-Adjusted Delegation Value (RADV)

A. Structural Governance Insufficiency

Institutions must document:

- 12-month staffing trends
- Budget allocation history
- Recruitment attempts
- Process redesign efforts

Workforce strategies that materially impair revocability capacity shall not be relied upon as sufficient justification for delegation of authority.

B. Comparative Excellence (Optimization Track)
Performance superiority must:

- Be statistically validated
- Survive operational stress testing
- Persist under adversarial conditions
- Exceed full governance cost burden

Resilience-Adjusted Delegation Value (RADV) — Optimization Track Evaluation
RADV treats governance spend as the price of maintaining control under failure. It values the ability to restore operations as a core component of performance, not an external compliance add-on.

RADV = Verified Operational Performance Gain − Full Lifecycle Governance Burden − Stress-Condition Restoration Cost
Where:
• **Verified Operational Performance Gain** includes statistically validated efficiency improvements sustained under operational stress testing and adversarial simulation.
• **Full Lifecycle Governance Burden** includes monitoring labor, audit infrastructure, drift instrumentation, competence validation, revocability drills, reauthorization review, and vendor contingency planning.
• **Stress-Condition Restoration Cost** includes the projected cost and time to restore independent control under model failure, vendor outage, distribution shift, or security isolation events.
Optimization Track delegation fails where RADV is non-positive under plausible stress scenarios. Efficiency gains that collapse under stress shall not be treated as durable performance. RADV forces efficiency claims to survive realistic failure, outage, and shift conditions before delegation is treated as a durable operating gain.
Strategic Competitive Context. Competitive pressure may justify considering delegation under the Optimization Track, but it does not reduce governance requirements. Where automation is adopted to preserve market position, continuity under stress becomes a material operating advantage, and therefore strengthens—not weakens—the case for monitoring intensity, drift instrumentation, revocability drills, competence validation, and infrastructure-layer control.

VI. Agentic Escalation Rule

Systems are presumptively Tier IV when they possess:
• Consequential cross-system write authority (material financial, legal, regulatory, safety, or rights impact), AND
• Low reversibility or cascading effect potential.
Administrative write authority (e.g., scheduling, formatting, non-material record updates) does not trigger Tier IV classification absent additional displacement indicators.

Reversibility shall be assessed based on:
• Time required to undo action
• Dependency on third-party processes
• Irreversible real-world consequences
Reversibility assessment shall include evaluation of infrastructure-layer propagation, including cross-system writes, replication state, caching layers, and deployment automation mechanisms. Structural propagation may elevate irreversibility classification independent of initial execution reversibility. (See Annex X.)
Classification must distinguish between operational convenience and materially consequential delegation.
Cross-system write authority is consequential only where the write operation can directly change real-world state, legal status, financial position, safety posture, enforcement action, or irreversible records.

VII. Expressive Authority

Expressive Authority refers to AI-generated outputs that, when acted upon, can create binding obligations or trigger consequential institutional actions.
AI outputs that:
• Create contractual obligations
• Trigger regulatory representations
• Affect financial transfers
• Draft enforcement determinations
• Initiate disciplinary, eligibility, or status-altering processes
...shall be classified **Tier IV or higher when such outputs are used directly to initiate or authorize consequential acts without independent reconstruction.**
Independent reconstruction requires substantive human evaluation capable of altering material reasoning or outcome determination **and must be evidenced in the decision record** prior to execution.
AI drafting of legislation, judicial opinions, binding regulatory determinations, or other sovereign acts shall be presumptively Tier V where such drafts materially shape final institutional action.
Classification under Expressive Authority must be evaluated in conjunction with reversibility, cascading impact potential, and real-world consequence analysis as defined in Section VI (Agentic Escalation Rule).

VIII. Drift Instrumentation

Drift metrics evaluated over rolling 30-, 90-, and 180-day windows.
Escalation triggered if:
- Override vitality falls below threshold in two consecutive windows
- Concordance > 98% for 90 consecutive days in discretionary contexts
- Review time below cognitive plausibility threshold for 60 days

Drift remediation plan required for continued authorization.

Behavioral drift metrics do not preclude structural authority displacement. Where infrastructure analysis under Annex X identifies Entrenched or Autonomous conditions, Tier escalation shall occur independent of behavioral thresholds.

IX. Reconstructability Standards

Audit Floor Requirements for Tier III–V:

1. Version reproducibility
2. Input traceability
3. Decision log retention
4. Independent replay capability

Narrative explanation alone is insufficient.

Opaque systems require:

- Counterfactual testing
- Perturbation analysis
- Distribution shift monitoring
- Continuous adversarial testing

Failure to maintain audit floor invalidates reauthorization.

Reconstructability evaluation shall incorporate infrastructure-layer artifacts, including immutable event logs, distributed cache propagation, deployment version lineage, and system-of-record synchronization states. (See Annex X.)

X. Revocability and Recovery Time Objectives (RTO)

Revocability must be measured against defined Recovery Time Objectives:

Tier III → RTO ≤ 24 hours
Tier IV → RTO ≤ 4 hours
Tier V → Sector-defined critical tolerance

Failure to meet RTO invalidates reauthorization.
Suspension drills must be conducted at least annually for Tier III–V systems.

RTO validation shall be demonstrated through documented suspension exercises simulating:

- Vendor outage.
- API failure.
- Model corruption event.
- Security isolation event.

Failure to meet defined RTO in two consecutive drills requires suspension until remediation is demonstrated.

RTO validation must simulate infrastructure-layer suspension, including isolation of data ingestion, deployment automation, and cross-system propagation pathways. (See Annex X.)

Revocability must be executable without exclusive reliance on vendor intervention where technically and contractually feasible. Institutions must retain internal capacity to suspend or isolate system operation independent of vendor availability.

XI. Provisional Delegation Protocol (Emergency)

Emergency delegation requires:

- Written exigency declaration
- Scope limitation
- 90-day sunset
- Independent reviewer concurrence
- Post-exigency dependency audit

Emergency authority may not normalize permanent delegation.

XII. Implementation Pathway

Institutions may implement ADGO in phases:

Phase 1 — System Inventory and Tier Classification

Apply Five-Tier Spectrum and Annex X (Infrastructure Authority Drift Overlay) to existing systems.

Phase 2 — High-Tier Prioritization

Conduct DLT for Tier III–V systems.

Phase 3 — Drift Monitoring Deployment

Implement override vitality and concordance tracking.

Phase 4 — Revocability Drill Integration

Incorporate RTO testing into incident response programs.

Phase 5 — Periodic Reauthorization

Require annual review for Tier III–V systems.

XIII. Sector Illustrations

Finance

Automated underwriting systems may reduce default rates while eroding manual credit analysis capacity. During system outage, loan processing halts, revealing dependency fragility.

Healthcare

AI triage tools may demonstrate high predictive accuracy. However, if clinicians cannot manually reprioritize during outages, patient safety may paradoxically decline under stress conditions.

Infrastructure

Autonomous grid optimization may increase efficiency. Yet if human operators cannot restore grid balance without AI support, revocability is illusory.

XIV. Anticipated Challenges and Limitations

ADGO presents implementation challenges:

Measurement Burden
Collecting override and drift metrics requires instrumentation. Mitigation: phased deployment and integration into existing audit systems.

Cost Concerns
Governance cost modeling may reduce apparent efficiency gains. Mitigation: sector-calibrated thresholds and pilot programs.

Cultural Resistance
Organizations may resist formalizing authority analysis. Mitigation: align ADGO with existing risk and compliance structures.

Threshold Calibration
Fixed thresholds may not fit all sectors. ADGO recommends empirical benchmarking and sector adjustment.

ADGO does not seek to prohibit automation. It seeks to preserve operational authority.

XV. Alignment with U.S. governance practice

Authority displacement is first an operational risk: institutions can lose practical control while performance metrics remain acceptable. Where decision systems affect protected rights, safety, or legal status, authority erosion also becomes a legitimacy risk because responsibility remains assigned to human institutions even when intervention capacity is no longer real. ADGO therefore treats authority preservation as a governance objective independent of error minimization and independent of short-term efficiency.

The AI RMF centers on risk identification and mitigation. ADGO centers on authority allocation and preservation. These dimensions are analytically distinct. Risk mitigation focuses on reducing harm probability and severity. Authority governance focuses on preserving institutional capacity to intervene, override, and reassume decision control.

A system may score low on harm metrics yet score high on authority displacement. By adding tier classification, drift instrumentation, and revocability testing, ADGO enhances the “Govern” and “Measure” functions without altering the RMF’s underlying architecture. Future empirical study should evaluate sector-specific threshold calibration, operator competence retention rates, and the long-term institutional effects of authority migration.

Govern → Tier classification + DLT + Infrastructure analysis
Map → Authority displacement profiling (behavioral + structural)
Measure → Drift metrics + structural entrenchment indicators
Manage → RTO validation + reauthorization + infrastructure reassessment

ADGO complements risk governance by measuring who governs in practice.

Conclusion
Risk governance answers:
“What harms may occur?”

Authority governance answers:
“Who governs in practice?”

A system can be safe and yet displace governance.

Institutions are encouraged to pilot ADGO in high-stakes domains to empirically evaluate threshold calibration, governance cost assumptions, and implementation feasibility.

Sustainable AI governance requires measurement of both risk and authority.

First articulated by Kathryn G. Varnes within Athosphere LLC.
AI Acknowledgment: The Authority Displacement Governance Overlay and all underlying legal frameworks, analytical judgments, and conclusions are the original intellectual property of the author. AI tools were utilized solely for mechanical language refinement and stylistic formatting.

ANNEX X - Infrastructure Authority Drift Overlay (IADO)

Tier × Infrastructure Classification Matrix
(Binding for Tier III–V Tier Determination and Reauthorization)

X.1 Authority Anchor Within ADGO

This Annex operationalizes and supplements:

- Section III — Five-Tier Authority Displacement Spectrum
- Section IV — Delegation Legitimacy Test (DLT)
- Section VI — Agentic Escalation Rule
- Section VIII — Drift Instrumentation
- Section IX — Reconstructability Standards
- Section X — Revocability and Recovery Time Objectives

It provides a structured infrastructure-layer analysis required for Tier III–V classification and reauthorization.
This Annex is binding for Tier III–V classification and reauthorization. Where conflict exists between behavioral indicators and infrastructure-layer conditions, infrastructure classification governs Tier determination.

Infrastructure-layer classification is determinative for Tier III–V systems where it constrains revocability, reconstructability, or competence preservation in practice.

X.2 Infrastructure Domains

Authority migration shall be evaluated across four infrastructure domains:

1. Data (Epistemic Authority)
   Control over system knowledge formation, ingestion, retraining, and feedback loops.
2. Compute (Functional Authority)
   Control over outcome determination and execution defaults.
3. Persistence (Structural Authority)
   Control over entrenchment, rollback feasibility, and propagation across systems.
4. Deployment (Governance Authority)
   Control over threshold modification, version propagation, and automation of consequential changes.

X.3 Tier × Infrastructure Classification Matrix (Binding)

Infrastructure Layer	Human-Controlled	Tier III — Determinative	Tier IV — Entrenched / Autonomous Operational	Tier V — Autonomous / Irreversible
Data	Data updates require documented human authorization	Continuous ingestion materially alters outputs prior to review	Automated retraining or feedback loops without governance checkpoint	Self-directed epistemic modification or suspension materially disruptive
Compute	Advisory outputs; independent human judgment	Outputs default to execution unless reversed	Overrides rare, operationally costly, or culturally discouraged	Consequential actions executed without prior human authorization
Persistence	Logged and reversible under ordinary controls	Outputs written to system-of-record or immutable event logs	Multi-layer propagation (replication, caching, cross-system synchronization) materially complicates rollback	External irreversible effects (financial transfer, legal action, physical actuation)
Deployment	Manual approval required prior to consequential change	Version-controlled but automatically applied updates	CI/CD or automated pipelines propagate consequential changes absent structured dissent	Dynamic self-adjusting thresholds or policies without human authorization

X.4 Escalation Rules (Binding)

1. Single-Layer Trigger Rule
    - If any infrastructure layer meets Tier III criteria, the system shall be classified at minimum Tier III.
    - If any layer meets Tier IV criteria, the system shall be classified at minimum Tier IV.
    - If any layer meets Tier V criteria, the system shall be classified Tier V.
2. Compounding Structural Drift Rule
    - If two or more infrastructure layers meet Tier III criteria, Tier escalation by one level shall occur (e.g., Tier III → Tier IV review).
3. Structural Supremacy Rule
    - Infrastructure-layer classification shall supersede behavioral indicators and formal oversight claims where structural conditions render practical human intervention nominal in operation.

X.5 Relationship to Section IV — Delegation Legitimacy Test

Infrastructure entrenchment directly affects:

- Revocability in practice
- Reconstructability
- Competence preservation

Where infrastructure-layer analysis reveals Tier IV or Tier V conditions, the Proponent must demonstrate how revocability, reconstructability, and competence remain operationally executable.

Failure to address infrastructure entrenchment constitutes DLT insufficiency.

X.6 Relationship to Section VI — Agentic Escalation Rule
Infrastructure analysis shall inform:

- Whether cross-system write authority is consequential
- Whether cascading effects extend beyond initial execution
- Whether rollback depends on vendor mediation or distributed propagation

Propagation layers, replication state, and deployment automation materially affect irreversibility classification.

X.7 Relationship to Sections VIII–X (Drift, Reconstructability, Revocability)
Infrastructure-layer findings shall:

- Be evaluated alongside behavioral drift metrics (Section VIII)
- Inform reconstructability scoring (Section IX)
- Be incorporated into RTO validation and suspension drills (Section X)

Where behavioral indicators remain compliant but infrastructure-layer analysis identifies Entrenched or Autonomous states, structural classification shall govern.

X.8 Harmonization Clause
This Annex:

- Does not redefine Tier categories.
- Does not modify DLT elements.
- Does not alter sovereign boundary constraints.
- Does not replace drift instrumentation.

It ensures that Tier classification reflects where authority resides in practice across infrastructure layers.
Where infrastructural automation renders human authority nominal in operation, Tier classification shall reflect structural displacement regardless of formal oversight claims.