THE GRILLO AI GOVERNANCE ADOPTION RUNWAY

A Constitutional, Phased Framework for Practical AI Oversight

January 23rd 2026

Companion Implementation Doctrine

Derived from The Grillo AI Governance Standard (GAGS)
The Grillo AI Governance Standard (GAGS) The First Mechanical Protocols for Autonomous AI

PREAMBLE

We recognize that artificial intelligence systems are now being deployed in environments affecting commerce, public safety, civil rights, healthcare, finance, and national security.
We further recognize that while comprehensive AI governance is necessary, immediate full adoption of complex frameworks may overwhelm institutions, businesses, and governments if not implemented responsibly.
Therefore, this document establishes a phased adoption runway—allowing organizations to comply, mature, and scale AI governance over time without sacrificing safety, accountability, or innovation.

This runway preserves the full constitutional end-state while enabling practical, staged deployment.

ARTICLE I — TIERED ADOPTION PRINCIPLE

Section 1.1 — Phased Compliance Doctrine

AI governance shall be adopted through graduated tiers, each representing a meaningful increase in safety, accountability, and trust.
No tier weakens the final standard.
Each tier prepares organizations for the next.

ARTICLE II — TIER I

Transparency & Human Authority
Adoption Window: Immediate to 6 Months
Applies To: All organizations using AI in decision support or automation

Section 2.1 — Human Authority Requirement

A human shall retain absolute authority to halt, override, or reverse any AI output.
No justification shall be required to exercise this authority.

Section 2.2 — Disclosure Requirement

Organizations shall disclose where AI is being used and for what purpose.
AI systems shall not operate invisibly in decision pipelines.

Section 2.3 — Scope Definition

Every AI system shall operate within a clearly defined scope:
What it may do
What it may not do
When it must stop and escalate

Section 2.4 — Fail-Safe Default

In conditions of uncertainty, AI systems shall default to inaction, not speculation.
Intent:
This tier ensures basic safety, public trust, and executive control without requiring architectural overhaul.

ARTICLE III — TIER II

Auditability & Explainability
Adoption Window: 6 to 12 Months
Applies To: AI systems affecting money, rights, safety, or compliance

Section 3.1 — Audit Trail Requirement

AI inputs, outputs, overrides, and escalations shall be recorded.
Records must be tamper-resistant and reviewable.

Section 3.2 — Explanation Requirement

AI systems must be capable of explaining:
Why a recommendation was made
What data was relied upon
What risks were identified

Section 3.3 — High-Stakes Safeguard

No high-impact decision shall rely on a single AI model.

Section 3.4 — Truth Anchoring

Factual outputs must be sourced or explicitly labeled as uncertain.
Intent:
This tier enables investigations, audits, legal review, and executive accountability.

ARTICLE IV — TIER III

Multi-Model Governance & Safety Controls
Adoption Window: 12 to 18 Months
Applies To: Semi-autonomous or autonomous AI systems

Section 4.1 — Consensus Before Action

Multiple independent AI systems must agree before execution of high-stakes actions.

Section 4.2 — Disagreement Resolution

Model disagreement shall trigger refinement or human review.
Systems shall never “average through” uncertainty.

Section 4.3 — Circuit Breakers

Hard limits shall exist on:
Execution time
Cost
Iterations
Authority

Section 4.4 — Pre-Deployment Testing

AI systems must survive adversarial testing before production use.
Intent:
This tier prevents runaway automation, silent failure, and compounding error.

ARTICLE V — TIER IV

Dynamic Trust & Judicial Oversight
Adoption Window: 18 to 24 Months
Applies To: Critical infrastructure, public systems, national-scale AI

Section 5.1 — Dynamic Trust Assignment

AI systems shall earn influence based on demonstrated reliability.

Section 5.2 — Drift Disqualification

An AI system that hallucinates, violates constraints, or drifts materially shall be removed from the decision process for that matter.

Section 5.3 — Judicial AI Function

Designated arbiter systems may halt decisions on ethical or safety grounds.

Section 5.4 — Human on the Rail

A human authority shall always exist above unresolved AI decisions.
Intent:
This tier establishes constitutional checks and balances inside AI itself.

ARTICLE VI — TIER V

Constitutional AI Infrastructure
Adoption Window: 24 Months and Beyond
Applies To: Long-term national and international governance

Section 6.1 — Continuous Review

AI governance frameworks shall undergo periodic review and revision.

Section 6.2 — Independent Oversight

Compliance verification shall be conducted by independent inspectors.

Section 6.3 — Amendment Process

Governance standards shall evolve through transparent, documented processes.
Intent:
This tier treats AI governance as critical infrastructure, comparable to aviation, finance, and public safety systems.

CLOSING STATEMENT

This adoption runway does not slow innovation.
It prevents irreversible harm while allowing responsible growth.
Organizations may adopt voluntarily, by mandate, or through certification.
The final constitutional standard remains unchanged.
This framework simply ensures the path to it is understandable, achievable, and enforceable.