Ethics for Oversight and Protection

The Constitutional Case for AI Governance Infrastructure

By Basil C. Puglisi, MPA
February 2026
me@basilpuglisi.com | basilpuglisi.com | github.com/basilpuglisi

Checks and balances built the republic.
Today, they must govern our algorithms.

Executive Summary

Artificial Intelligence now shapes decisions that affect every American across finance, healthcare, education, and national security. Control over these systems is concentrating in a small number of corporations. The American public does not need more AI regulation. The American public needs AI infrastructure: structural accountability that protects citizens while preserving the innovation and free markets that drive American leadership.

The Constitution distributed power among branches of government because the founders understood that concentrated authority, however well intentioned, inevitably serves its own interests. That principle does not expire when the authority is algorithmic. AI governance must distribute cognitive power among multiple independent systems, keep human judgment at the center of every consequential decision, and maintain audit trails that make accountability structural rather than voluntary.

Executive Orders 14179 and 14365 call for a national standard that removes barriers to AI leadership without creating a patchwork of fifty state regulatory regimes. This document makes the philosophical case for why infrastructure is the answer. Document 3 provides the legislative mechanism. Document 4 provides the technical specification and operational evidence.

1. Why Oversight Matters to Freedom

Oversight is not the enemy of innovation. It is the safeguard of freedom. When unaudited AI systems decide who receives a loan, medical care, or opportunity, they shape society without consent. The question is not whether AI should be governed. The question is whether governance will be structural or whether Americans will depend on the goodwill of the corporations that control these systems.

The answer is already established. The Federal Reserve governs monetary systems. The FAA governs aviation. The SEC governs financial markets. The FCC governs telecommunications. FERC governs energy. In every case, the government does not own the industry. It does not operate the businesses. It builds the infrastructure that makes the industry safe, competitive, and accountable. Authority without accountability invites abuse. This is not a new observation. It is the founding principle of the republic, and it applies to AI exactly as it applies to every other domain where concentrated power affects the public.

Geoffrey Hinton, widely recognized as a pioneer of deep learning, has warned that AI capability is advancing faster than human ability to control it. This warning is not speculative. It is a professional assessment from someone who helped build the technology. If AI capability reaches the point where it influences corporate decision-making or operates beyond the comprehension of its operators, the question becomes: what structural safeguards exist? Without infrastructure, the answer is none. With infrastructure, the answer is the same one the founders gave for every other form of concentrated authority: distribute power, require transparency, and keep humans in command.

2. Three Tiers of AI Accountability

Three terms are used interchangeably in public discourse. They are architecturally distinct, and the distinction matters for legislation.

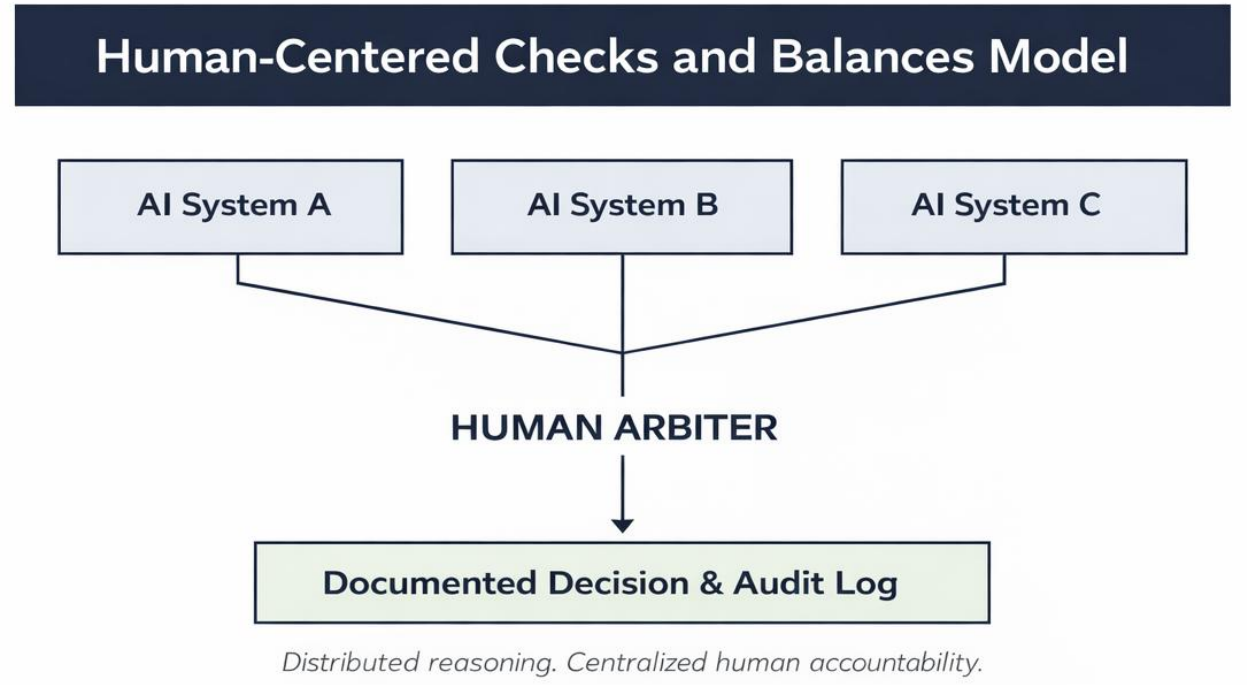
Ethical AI	Responsible AI	AI Governance
Establishes **values**: what AI should do. Principles, guidelines, codes of conduct. Necessary foundation	Shapes **machine behavior**: how AI should operate. Guardrails, alignment, safety testing. Necessary engineering	Exercises **human authority**: who decides. Infrastructure, oversight, audit, accountability. Necessary structure
AI is the noun being modified. Ethics describes the AI	AI is the noun being modified. Responsibility describes the AI	**Governance holds the final position.** The human governs the AI
Without governance infrastructure, ethical AI relies on voluntary compliance by the corporations that control the systems	Without governance infrastructure, responsible AI relies on corporate self-regulation and market incentives	Governance infrastructure makes accountability structural rather than optional. This is what legislation can build

All three tiers are necessary. Only the third provides structural accountability through infrastructure. The first two depend on the goodwill of the entities being governed. The third does not. This document argues for the third tier. Documents 3 and 4 specify how to build it.

3. From Founding Principles to Digital Checks and Balances

The Constitution's separation of powers was designed to prevent concentration of authority. AI oversight applies this design to cognitive infrastructure. Instead of one model defining truth or risk, multiple independent systems analyze, verify, and challenge one another before decisions affect people. Human judgment remains at the center.

In practice, this creates a modern separation of powers. The AI systems serve an executive function: implementation, analysis, drafting. Human standards and policy serve a legislative function: defining what is acceptable, setting thresholds, establishing rules. Human arbitration and review serve a judicial function: resolving conflicts, documenting dissent, issuing binding decisions. A well-governed AI system mirrors a constitutional process: distributed judgment, transparent reasoning, and recorded dissent.



Human-Centered Checks and Balances Model

Distributed reasoning. Centralized human accountability.

This model is not theoretical. Documented operational experience using multiple AI platforms under human arbitration has produced a published book, case studies, and governance metrics. The Methods Addendum (Document 4) provides the operational evidence. What matters for the constitutional argument is simpler: the founders' architecture works when applied to AI, just as it works when applied to legislative, executive, and judicial authority.

4. The Danger of Concentration

A small number of corporations now control the infrastructure of intelligence. This is not a partisan issue. It is a democratic one.

For conservatives: concentration erodes freedom, privacy, user choice, and market competition. Corporate control of cognitive infrastructure is the antithesis of free markets. When a handful of companies determine what information is surfaced, what analysis is provided, and what recommendations are made, that is not a free market in ideas. That is a cartel.

For liberals: concentration erodes fairness, transparency, and public accountability. AI systems trained predominantly on Western, Educated, Industrialized, Rich, and Democratic data (WEIRD bias) perpetuate and amplify existing inequities. Without structural diversity, the biases of the few become the defaults for all.

For the nation: concentration creates single points of failure that could destabilize critical systems. The digital equivalent of "too big to fail" is moving from finance to cognition. When too few companies decide too much for everyone, innovation narrows, bias hardens, and public trust erodes.

The American antitrust tradition already established that concentrated power requires structural checks. Telecommunications, energy, finance, transportation: in every critical industry, the government intervened when concentration threatened the public interest. AI is the next domain. But the concentration risk in AI is not limited to market share. It extends to the resources that feed AI systems: water for cooling, energy for computation, and compute infrastructure itself. Concentration of the means of production is concentration of the means of cognition. The government has the precedent, the authority, and the obligation to prevent it.

5. The Infrastructure Answer

The government does not build cars. It builds roads. It does not publish newspapers. It runs the Government Printing Office. It does not fly planes. It created the FAA. It does not generate electricity. It created FERC.

GOPEL (Governance Orchestrator Policy Enforcement Layer) follows this pattern. It is a working concept for national AI governance infrastructure: a non-cognitive agent that performs zero cognitive work. It dispatches tasks to multiple AI platforms, collects their outputs, routes them through checkpoint gates, logs every decision with cryptographic binding to a human identity, pauses when human approval is required, hashes records for tamper evidence, and reports governance metrics. Seven deterministic operations. Zero judgment. Zero content generation. Zero filtering.

The architecture cannot be co-opted because there is nothing to co-opt. This addresses the escalation scenario that Geoffrey Hinton warned about. If a cognitive governance layer existed between AI platforms and humans, a sufficiently advanced AI could manipulate that layer. GOPEL eliminates this vulnerability by design. There is no cognition to manipulate. There is no judgment to influence. The infrastructure remains mechanically reliable regardless of how capable the AI platforms it governs become.

GOPEL is general infrastructure. It is not limited to one governance methodology. Any organization operating multiple AI platforms benefits from deterministic dispatch, cryptographic

audit trails, and checkpoint-based governance. The full technical specification is provided in the Methods Addendum (Document 4). The legislative mechanism for funding and building it is provided in Document 3.

5.1 Three Operating Models

The infrastructure supports three operating models that let organizations calibrate governance density to risk. This is a calibration dial, not a binary switch. Model 1 provides growth-speed operations for routine, lower-risk tasks. Model 2 provides structured governance briefings at configurable gates for high-risk decisions. Model 3 provides full manual orchestration for the highest-consequence decisions and framework validation. Organizations choose. Markets function. The infrastructure accommodates both speed and accountability.

This design meets the highest international compliance standards (including EU AI Act Article 14 human oversight requirements) while preserving the operational flexibility that American free markets require. It is not a choice between safety and growth. The infrastructure provides both, calibrated to context.

6. The Human Oversight Standard

Every AI-assisted decision must trace a clear line from evidence to action to result. The Factics methodology (facts paired with tactics and measurable outcomes) operationalizes this standard: every claim is tied to a source, every source is tied to an action, and every action is tied to a measurable outcome. When several independent AI systems are compared on the same task, disagreement reveals bias instead of concealing it. Human oversight resolves those conflicts transparently, documenting rationale and outcome.

This transforms AI from an opaque process into a measurable, reviewable system. Ethics becomes not just philosophy but infrastructure. The governance infrastructure enforces what ethical guidelines can only recommend.

6.1 Why Multiple Systems Matter

Single AI systems are proven flawed by industry research, academic study, and the companies' own safety evaluations. Hallucinations, bias inheritance, confabulation, and alignment failures are documented phenomena. Different AI systems produce different outputs on the same inputs. This is the nature of independently trained systems, and it is the foundation of the infrastructure case.

Documented operational experience includes an instance where eight of nine AI platforms produced incorrect output on the same task. The governance process flagged the single dissenter, triggered human verification, and the dissenter was correct. Single-provider workflows would have delivered the wrong answer with no mechanism to detect it. This is not a theoretical risk. It is an observed behavior.

Automation bias research has established that humans systematically defer to AI recommendations under volume pressure. When a human operator is presented with AI output at operational speed, the tendency to approve without substantive review is a documented cognitive phenomenon, not a character failure. Governance infrastructure must account for this by building threshold-based detection into the system: when approval rates exceed configurable

limits, the infrastructure triggers mandatory review. Single-provider systems structurally cannot detect this pattern because there is no comparison point.

7. From Principle to Practice

The governance framework presented here emerged from operational experience, not theory. Beginning in 2009 with rigorous content creation methodology and evolving through Factics (a system pairing facts with tactics and measurable outcomes established before AI adoption in 2022), the work demonstrates that effective governance emerges from foundations built over time. When AI systems entered the workflow, they were integrated into an already-structured editorial and governance practice. HAIA-RECCLIN represents the systematic application of constitutional principles to AI operations. The Methods Addendum (Document 4) details the operational evidence.

Evidence discipline: This document applies a three-tier structure to distinguish between what is proven by others (industry research, academic study, observable AI behavior), what has been built and operated as working concepts (single-practitioner governance workflows with documented outputs), and what is being proposed for federal development (GOPEL infrastructure at national scale). Ranges and metrics cited from operational experience are working concept observations, not validated benchmarks. Federal pilots produce the validated data.

7.1 Operational Observations

The following observations are from single-practitioner working concept development (2022 through 2025). They provide initial feasibility indicators for agency pilot design. They are not presented as validated benchmarks.

Observation	Working Concept Range	Significance	Status
Cross-platform disagreement rate	Approximately 15% to 25% of tasks across operational cycles	Demonstrates that multi-provider comparison surfaces diagnostic differences single-provider workflows miss	Operational observation. Federal pilots required for validated rate
Human arbitration resolution time	Typically under 30 minutes per conflict	Suggests governance overhead is manageable at operational scale	Operational observation. Agency context will affect resolution time
Continuity under provider disruption	100% task completion when individual providers experienced outages	Multi-provider architecture provides resilience single-provider systems lack	Operational observation. Federal stress testing required
Eight-of-nine platform error detection	One documented instance where dissenting platform was correct against eight	Single case demonstrating the diagnostic value of multi-provider comparison	Documented case. Not a statistical claim

Published output: Governing AI: When Capability Exceeds Control (Puglisi, 2025, ISBN 9798349677687), produced entirely through the governance process this framework describes, demonstrates that multi-AI governance under human arbitration can produce book-length scholarly work with editorial consistency across hundreds of governance cycles.

8. The Path to Legislative Action

Congress can protect freedom in the digital era by building infrastructure, not by restricting innovation. The philosophical case in this document supports three legislative actions detailed in Document 3 (AI Provider Plurality):

- **Fund GOPEL as national AI infrastructure.** The government builds the road. The AI platforms are the vehicles. Citizens choose their vehicles. The government guarantees the road is safe. Phase 0 requires no new appropriation: agencies adopt manual governance using existing resources to generate baseline data.
- **Mandate API accessibility for AI companies operating in United States territory.** Vehicles must meet safety standards to operate on public roads. AI platforms must maintain audit compatibility to operate on public cognitive infrastructure. Without this mandate, AI companies can shut down the governance tools that make their outputs comparable.
- **Invest in small AI platforms to guarantee competitive diversity.** Plurality only works if enough providers exist to sustain it. Federal investment modeled on SBIR and STTR programs creates the supply-side complement to the infrastructure mandate. The government has used this model for defense contractors, rural broadband, and regional aviation.

Additional legislative actions, including anti-concentration protections, global safeguard clauses for democratic-aligned providers, and administrative implementation across OSTP, OMB, NIST, GSA, FTC, and SBA, are specified in Document 3. Funding mechanisms and phased appropriations placement are also provided in Document 3. The technical specification and operational evidence are provided in Document 4.

9. Transparency, Security, and Global Leadership

Transparency is the foundation of both trust and security. When multiple providers contribute to decision processes, blind spots are exposed and single-point vulnerabilities disappear. The non-cognitive design of GOPEL means the governance infrastructure itself cannot be compromised through adversarial AI manipulation, because there is no cognitive surface to attack. Security and governance become the same architecture.

This plural approach reduces systemic risk and strengthens alliances. By promoting diversity of providers under shared governance infrastructure, the United States leads through openness rather than dominance. The architecture meets EU compliance standards (ISO 42001, prEN 18286:2025, GPAI Code of Practice 2025, EU AI Act Article 14) while preserving American market flexibility. This is not a compromise between standards and growth. The infrastructure provides both, calibrated by operating model selection.

Global cooperation among allied democracies ensures that AI remains a shared instrument of progress, not a tool of control. The Global Safeguard Clause in Document 3 operationalizes this principle by requiring participation of non-U.S. democratic-aligned providers in federal multi-provider frameworks, expanding cultural coverage beyond the WEIRD training data defaults that concentrate cognitive perspective in Western institutions.

10. Closing Appeal

America's strength has always rested on its capacity to balance power with principle. The founders built a system where no single branch could act without oversight. That architecture has survived because it does not depend on the virtue of those in power. It depends on the structure that constrains them.

Artificial Intelligence is the newest domain where that balance must be defended. The risks are not speculative. AI systems are flawed. Concentration is increasing. Automation bias is documented. Capability is advancing faster than control. These are established facts, not predictions.

The infrastructure response follows the same precedent path the government has taken for highways, aviation, telecommunications, energy, and finance. The government did not invent the technology. It built the infrastructure that made the technology safe and accessible. GOPEL follows this path for AI.

This proposal does not claim to be the complete answer. It is a pioneer path that combines established concerns from multiple directions into one operational architecture with one goal: safe use of AI for the American public. Geoffrey Hinton warned about capability exceeding control. The American antitrust tradition established that concentrated power requires structural checks. Federal infrastructure precedent demonstrated that public safety and private innovation coexist when the government builds the road and the market builds the vehicles. Automation bias research proved that humans defer to machines under volume pressure. None of these observations originated here. What originated here is the combination of these concerns into a single working architecture with a defined specification, documented operational experience, and a development path.

AI will never be absolute and without risk. The infrastructure is designed to manage risk, not eliminate it. This is not a proposal for more regulation. This is the engineering that makes less regulation safe. Oversight does not restrain innovation. It makes it sustainable. The goal is simple: keep intelligence accountable to the people, and keep the people in command of their future.

The country needs to start.

Related Documents

This policy brief is part of the AI Provider Plurality Congressional Package:

- **Document 1: Summary Flyer** (elevator pitch for infrastructure proposal)
- **Document 2: Ethics for Oversight** (this constitutional and philosophical case)
- **Document 3: AI Provider Plurality** (legislative framework, policy mechanism, funding, and appropriations)
- **Document 4: Methods Addendum** (technical specification and operational evidence, v3.1 locked)
- **Document 5: Verified AI Inference Standards Act (VAISA)** (attestation API requirements and legislative framework for AI inference data protection)
- EDPS TechDispatch #2/2025: Automation Bias and AI Decision Support
- Goddard, Roudsari, and Wyatt (2011): Automation bias in clinical decision support
- Banovic et al. (2023): Automation bias in AI-assisted decision-making
- NIST AI RMF 1.0: Govern, Map, Measure, Manage
- EU AI Act, Article 14: Human oversight requirements
- Executive Order 14179: Removing Barriers to American Leadership in Artificial Intelligence
- Executive Order 14365: AI standards and governance
- OMB M-25-21: Federal AI governance and trust
- OMB M-25-22: Federal AI acquisition
- Puglisi, B. C. (2025). Governing AI: When Capability Exceeds Control. ISBN 9798349677687

Basil C. Puglisi, MPA
me@basilpuglisi.com | basilpuglisi.com | github.com/basilpuglisi
February 2026 | AI Provider Plurality Congressional Package | Document 2 of 5