When Control Systems Inherit the Same Uncertainty They Are Meant to Govern

Heather Grizzle · April 18, 2026

AI governance often relies on probabilistic systems evaluating other probabilistic systems, creating circular risk. This analysis explains why true governance requires a distinct reliability layer, not additional model-based controls.

AI governance is currently being described and sold through what is easiest to visualize rather than what is most important to analyze.

The conversation is dominated by visible artifacts: policy dashboards, layered control diagrams, labeled guardrails, and step-by-step execution flows that suggest a request is being rigorously evaluated before any action occurs. These representations create the impression of control because they resemble traditional system oversight. However, they obscure the more consequential question of what kind of mechanism is actually performing the evaluation inside those layers.

In most current architectures, an AI system generates a proposed action, and that proposal is routed through a sequence of checks that culminate in another model determining whether the action should proceed. This evaluation layer may be described in different ways depending on the vendor or framework. It may be positioned as a policy engine, a risk scoring mechanism, or a confidence-ranking system augmented by rules. Despite these variations in terminology, the underlying structure remains consistent. A probabilistic model is tasked with assessing the appropriateness of an output generated by another probabilistic model.

This design choice has significant implications when examined through a risk and reliability lens. Probabilistic systems, by definition, do not produce deterministic, repeatable

outcomes under identical conditions. Their outputs are shaped by statistical patterns derived from training data, and their evaluations are similarly influenced by probabilistic inference rather than fixed criteria. When such a system is used to govern another system of the same type, the governing function does not introduce a fundamentally different form of reliability. Instead, it replicates the same uncertainty profile at a higher layer of abstraction.

From a governance standpoint, this creates a circular structure. The system that is intended to enforce control is subject to the same variability, edge-case instability, and context sensitivity as the system it oversees. As a result, the presence of additional layers, checks, or wrappers does not necessarily translate into meaningful risk reduction. It may increase complexity, and it may create the appearance of diligence, but it does not change the underlying behavior of the system in a way that produces a distinct control boundary.

It introduces recursion rather than control.

This distinction becomes clearer when compared to how governance functions in other high-stakes domains. In financial systems, for example, transaction approval processes rely on predefined rules, audit logs, and human-authored criteria that can be traced and reconstructed. In safety-critical engineering contexts, control systems are designed to operate with different reliability properties than the systems they monitor. The governing mechanism is intentionally separated in both function and behavior so that it can provide a stable point of reference when uncertainty or failure occurs.

Current AI governance patterns often do not establish that separation. Instead, they layer probabilistic judgment on top of probabilistic generation and treat the resulting structure as a control system. This approach assumes that adding more evaluation steps within the same reliability class will produce governance outcomes. In practice, it introduces recursion rather than control, because each layer depends on the same type of reasoning process

and is vulnerable to the same types of failure.

The consequence is not simply theoretical. When a system produces an outcome that causes harm, the question that follows is not whether there were multiple layers involved in the decision. The question is whether any of those layers functioned as a true control, meaning that it operated with a different level of determinism, traceability, and accountability than the system it governed. If every layer shares the same probabilistic foundation, then the system lacks a clear point at which responsibility can be anchored or decisions can be reliably reconstructed.

THE DEFINING TEST

Governance is defined by the introduction of a mechanism that operates with a different reliability profile than the system being governed.

The position that follows from this analysis is direct. Governance is not defined by the number of controls present in a system, nor by the visual complexity of its architecture. Governance is defined by the introduction of a mechanism that operates with a different reliability profile than the system being governed. Without that distinction, additional layers do not constitute control. They extend the same uncertainty across more components.

When a control layer behaves in the same way as the system it is meant to regulate, the architecture does not achieve governance in a meaningful sense. It produces a recursive structure in which uncertainty evaluates uncertainty, and the boundary between action and oversight collapses. That collapse is where the real risk resides, regardless of how structured the system appears on the surface.