This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2026/04/18/when-control-systems-inherit-the-same-uncertainty-they-are-meant-to-govern/.

Cuando los sistemas de control heredan la misma incertidumbre que se supone deben gobernar

Heather Grizzle · April 18, 2026

When AI guardrails are themselves probabilistic models, governance inherits the uncertainty it is meant to control. What real control requires.

La gobernanza de la IA se está describiendo y vendiendo actualmente en función de lo que resulta más fácil de visualizar, en lugar de lo que es más importante analizar.

La conversación está dominada por artefactos visibles: paneles de políticas, diagramas de controles superpuestos, barreras de protección etiquetadas y flujos de ejecución paso a paso que sugieren que una solicitud se está evaluando rigurosamente antes de que ocurra cualquier acción. Estas representaciones crean la impresión de control porque se asemejan a la supervisión tradicional de sistemas. Sin embargo, ocultan la pregunta más importante de qué tipo de mecanismo está realmente realizando la evaluación dentro de esas capas.

En la mayoría de las arquitecturas actuales, un sistema de IA genera una acción propuesta, y esa propuesta se enruta a través de una secuencia de comprobaciones que culminan con otro modelo determinando si la acción debe proceder. Esta capa de evaluación puede describirse de diferentes maneras según el proveedor o el marco de trabajo. Puede presentarse como un motor de políticas, un mecanismo de puntuación de riesgo o un sistema de clasificación por confianza complementado con reglas. A pesar de estas variaciones en la terminología, la estructura subyacente se mantiene constante. Se encarga a un modelo probabilístico la tarea de evaluar la idoneidad de un resultado generado por otro modelo probabilístico.

Esta decisión de diseño tiene implicaciones significativas cuando se examina desde una perspectiva de riesgo y fiabilidad. Los sistemas probabilísticos, por definición, no

producen resultados deterministas y repetibles bajo condiciones idénticas. Sus resultados están determinados por patrones estadísticos derivados de los datos de entrenamiento, y sus evaluaciones están igualmente influenciadas por la inferencia probabilística en lugar de criterios fijos. Cuando se utiliza un sistema de este tipo para gobernar otro sistema del mismo tipo, la función de gobernanza no introduce una forma de fiabilidad fundamentalmente diferente. En cambio, replica el mismo perfil de incertidumbre en una capa de abstracción superior.

Desde el punto de vista de la gobernanza, esto crea una estructura circular. El sistema que se supone debe imponer el control está sujeto a la misma variabilidad, inestabilidad ante casos límite y sensibilidad al contexto que el sistema que supervisa. Como resultado, la presencia de capas, comprobaciones o envoltorios adicionales no se traduce necesariamente en una reducción significativa del riesgo. Puede aumentar la complejidad y puede crear la apariencia de diligencia, pero no cambia el comportamiento subyacente del sistema de una manera que produzca un límite de control diferenciado.

Esta distinción se vuelve más clara al compararla con la forma en que funciona la gobernanza en otros ámbitos de alto riesgo. En los sistemas financieros, por ejemplo, los procesos de aprobación de transacciones se basan en reglas predefinidas, registros de auditoría y criterios elaborados por humanos que pueden rastrearse y reconstruirse. En contextos de ingeniería críticos para la seguridad, los sistemas de control están diseñados para operar con propiedades de fiabilidad distintas a las de los sistemas que supervisan. El mecanismo de gobernanza se separa intencionalmente tanto en función como en comportamiento, de modo que pueda proporcionar un punto de referencia estable cuando surja incertidumbre o falla.

> Introduce recursión en lugar de control.

Los patrones actuales de gobernanza de la IA a menudo no establecen esa separación. En cambio, superponen el juicio probabilístico sobre la generación probabilística y tratan la estructura resultante como un sistema de control. Este enfoque asume que añadir más pasos de evaluación dentro de la misma clase de fiabilidad producirá resultados de gobernanza. En la práctica, introduce recursión en lugar de control, porque cada capa depende del mismo tipo de proceso de razonamiento y es vulnerable a los mismos tipos de fallos.

La consecuencia no es simplemente teórica. Cuando un sistema produce un resultado que causa daño, la pregunta que sigue no es si hubo múltiples capas involucradas en la decisión. La pregunta es si alguna de esas capas funcionó como un control verdadero, es decir, si operó con un nivel diferente de determinismo, trazabilidad y responsabilidad que el sistema que gobernaba. Si todas las capas comparten el mismo fundamento probabilístico, entonces el sistema carece de un punto claro en el que se pueda anclar la responsabilidad o en el que las decisiones puedan reconstruirse de manera fiable.

> **LA PRUEBA DEFINITORIA**
>
> La gobernanza se define por la introducción de un mecanismo que opera con un perfil de fiabilidad diferente al del sistema que está siendo gobernado.

La postura que se deriva de este análisis es directa. La gobernanza no se define por la cantidad de controles presentes en un sistema, ni por la complejidad visual de su arquitectura. La gobernanza se define por la introducción de un mecanismo que opera con un perfil de fiabilidad diferente al del sistema que está siendo gobernado. Sin esa distinción, las capas adicionales no constituyen control. Extienden la misma incertidumbre a más componentes.

Cuando una capa de control se comporta de la misma manera que el sistema que se supone debe regular, la arquitectura no logra una gobernanza significativa. Produce una estructura recursiva en la que la incertidumbre evalúa a la incertidumbre, y el límite entre la acción y la supervisión colapsa. Ese colapso es donde reside el verdadero riesgo, sin importar cuán estructurado parezca el sistema en la superficie.

> **TIPS**
>
> Know something about a sign language AI system that the public should know? Send it to tips@novaracg.com. Every tip is read by me and kept confidential. If you work in the UK, writing to us is not by itself a protected disclosure under the Public Interest Disclosure Act 1998, so before you share employer information, check your options with Protect (protect-advice.org.uk), the free whistleblowing advice line.

Cómo citar

Grizzle, H. M. (2026, April 18). Cuando los sistemas de control heredan la misma incertidumbre que se supone deben gobernar. Novara Consulting Group. https://www.novaracg.com/es/2026/04/18/when-control-systems-inherit-the-same-uncertainty-they-are-meant-to-govern/