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/.

# Quand les systèmes de contrôle héritent de la même incertitude qu'ils sont censés gouverner

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 gouvernance de l'IA est actuellement décrite et présentée à travers ce qui est le plus facile à visualiser plutôt qu'à travers ce qui est le plus important à analyser.

La conversation est dominée par des artefacts visibles : tableaux de bord de politiques, diagrammes de contrôle superposés, garde-fous étiquetés et flux d'exécution étape par étape qui suggèrent qu'une demande est rigoureusement évaluée avant toute action. Ces représentations créent l'impression d'un contrôle parce qu'elles ressemblent à la supervision traditionnelle des systèmes. Cependant, elles occultent la question plus fondamentale de savoir quel type de mécanisme effectue réellement l'évaluation au sein de ces couches.

Dans la plupart des architectures actuelles, un système d'IA génère une action proposée, et cette proposition est acheminée à travers une séquence de vérifications qui aboutissent à un autre modèle déterminant si l'action doit se poursuivre. Cette couche d'évaluation peut être décrite de différentes manières selon le fournisseur ou le cadre utilisé. Elle peut être présentée comme un moteur de politiques, un mécanisme de notation des risques, ou un système de classement de confiance renforcé par des règles. Malgré ces variations de terminologie, la structure sous-jacente reste constante. Un modèle probabiliste est chargé d'évaluer la pertinence d'une sortie générée par un autre modèle probabiliste.

Ce choix de conception a des implications importantes lorsqu'on l'examine sous l'angle du risque et de la fiabilité. Les systèmes probabilistes, par définition, ne pro-

duisent pas de résultats déterministes et reproductibles dans des conditions identiques. Leurs sorties sont façonnées par des motifs statistiques dérivés des données d'entraînement, et leurs évaluations sont de même influencées par une inférence probabiliste plutôt que par des critères fixes. Lorsqu'un tel système est utilisé pour gouverner un autre système du même type, la fonction de gouvernance n'introduit pas une forme de fiabilité fondamentalement différente. Elle reproduit plutôt le même profil d'incertitude à un niveau d'abstraction supérieur.

D'un point de vue de la gouvernance, cela crée une structure circulaire. Le système censé faire appliquer le contrôle est soumis à la même variabilité, à la même instabilité liée aux cas limites et à la même sensibilité au contexte que le système qu'il supervise. En conséquence, la présence de couches, de vérifications ou d'enveloppes supplémentaires ne se traduit pas nécessairement par une réduction significative du risque. Elle peut accroître la complexité, et elle peut créer une apparence de diligence, mais elle ne modifie pas le comportement sous-jacent du système d'une manière qui produirait une limite de contrôle distincte.

Cette distinction devient plus claire lorsqu'on la compare à la façon dont la gouvernance fonctionne dans d'autres domaines à fort enjeu. Dans les systèmes financiers, par exemple, les processus d'approbation des transactions reposent sur des règles prédéfinies, des journaux d'audit et des critères rédigés par des humains qui peuvent être tracés et reconstitués. Dans les contextes d'ingénierie critiques pour la sécurité, les systèmes de contrôle sont conçus pour fonctionner avec des propriétés de fiabilité différentes de celles des systèmes qu'ils surveillent. Le mécanisme de gouvernance est intentionnellement séparé, tant sur le plan fonctionnel que comportemental, afin qu'il puisse fournir un point de référence stable en cas d'incertitude ou de défaillance.

> Elle introduit de la récursivité plutôt que du contrôle.

Les schémas actuels de gouvernance de l'IA n'établissent souvent pas cette séparation. Ils superposent plutôt un jugement probabiliste sur une génération probabiliste et traitent la structure qui en résulte comme un système de contrôle. Cette approche suppose que l'ajout d'étapes d'évaluation supplémentaires au sein de la même classe de fiabilité produira des résultats de gouvernance. En pratique, cela introduit de la ré-





cursivité plutôt que du contrôle, car chaque couche dépend du même type de processus de raisonnement et est vulnérable aux mêmes types de défaillance.

La conséquence n'est pas simplement théorique. Lorsqu'un système produit un résultat qui cause un préjudice, la question qui se pose ensuite n'est pas de savoir si plusieurs couches étaient impliquées dans la décision. La question est de savoir si l'une de ces couches a fonctionné comme un véritable contrôle, c'est-à-dire si elle a opéré avec un niveau différent de déterminisme, de traçabilité et de responsabilité que le système qu'elle gouvernait. Si chaque couche partage le même fondement probabiliste, alors le système manque d'un point clair où la responsabilité peut être ancrée ou où les décisions peuvent être reconstituées de manière fiable.

> **LE TEST DÉTERMINANT**
>
> La gouvernance se définit par l'introduction d'un mécanisme qui fonctionne avec un profil de fiabilité différent de celui du système gouverné.

La position qui découle de cette analyse est directe. La gouvernance ne se définit pas par le nombre de contrôles présents dans un système, ni par la complexité visuelle de son architecture. La gouvernance se définit par l'introduction d'un mécanisme qui fonctionne avec un profil de fiabilité différent de celui du système gouverné. Sans cette distinction, des couches supplémentaires ne constituent pas un contrôle. Elles étendent la même incertitude à davantage de composants.

Lorsqu'une couche de contrôle se comporte de la même manière que le système qu'elle est censée réguler, l'architecture n'atteint pas la gouvernance au sens propre. Elle produit une structure récursive dans laquelle l'incertitude évalue l'incertitude, et la frontière entre l'action et la supervision s'effondre. Cet effondrement est là où réside le véritable risque, quelle que soit la structure apparente du système en surface.

**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.

## Comment citer

Grizzle, H. M. (2026, April 18). Quand les systèmes de contrôle héritent de la même incertitude qu'ils sont censés gouverner. Novara Consulting Group. https://www.novaracg.com/fr/2026/04/18/when-control-systems-inherit-the-same-uncertainty-they-are-meant-to-govern/