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/.

Quando os Sistemas de Controlo Herdam a Mesma Incerteza que Deveriam Governar

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.

A governação de IA está atualmente a ser descrita e vendida através do que é mais fácil de visualizar, em vez do que é mais importante analisar.

A conversa é dominada por artefactos visíveis: painéis de políticas, diagramas de controlo em camadas, barreiras de proteção rotuladas e fluxos de execução passo a passo que sugerem que um pedido está a ser rigorosamente avaliado antes de qualquer ação ocorrer. Estas representações criam a impressão de controlo porque se assemelham à supervisão tradicional de sistemas. No entanto, obscurecem a questão mais consequente de que tipo de mecanismo está de facto a realizar a avaliação dentro dessas camadas.

Na maioria das arquiteturas atuais, um sistema de IA gera uma ação proposta, e essa proposta é encaminhada através de uma sequência de verificações que culminam noutro modelo a determinar se a ação deve prosseguir. Esta camada de avaliação pode ser descrita de formas diferentes consoante o fornecedor ou a estrutura. Pode ser apresentada como um motor de políticas, um mecanismo de pontuação de risco, ou um sistema de classificação de confiança complementado por regras. Apesar destas variações de terminologia, a estrutura subjacente mantém-se consistente. É atribuída a um modelo probabilístico a tarefa de avaliar a adequação de um resultado gerado por outro modelo probabilístico.

Esta escolha de design tem implicações significativas quando examinada sob a ótica do risco e da fiabilidade. Os sistemas probabilísticos, por definição, não produzem resultados determinísticos e repetíveis sob condições idênticas. Os seus resultados são moldados por padrões estatísticos derivados dos dados de treino, e as suas avaliações

são igualmente influenciadas pela inferência probabilística, em vez de critérios fixos. Quando um sistema deste tipo é usado para governar outro sistema do mesmo tipo, a função governativa não introduz uma forma fundamentalmente diferente de fiabilidade. Em vez disso, replica o mesmo perfil de incerteza numa camada de abstração superior.

Do ponto de vista da governação, isto cria uma estrutura circular. O sistema destinado a impor o controlo está sujeito à mesma variabilidade, instabilidade em casos limite, e sensibilidade ao contexto que o sistema que supervisiona. Como resultado, a presença de camadas, verificações ou envoltórios adicionais não se traduz necessariamente numa redução significativa do risco. Pode aumentar a complexidade, e pode criar a aparência de diligência, mas não altera o comportamento subjacente do sistema de uma forma que produza uma fronteira de controlo distinta.

Introduz recursão, não controlo.

Esta distinção torna-se mais clara quando comparada com o modo como a governação funciona noutros domínios de alto risco. Nos sistemas financeiros, por exemplo, os processos de aprovação de transações baseiam-se em regras predefinidas, registos de auditoria e critérios elaborados por humanos que podem ser rastreados e reconstruídos. Em contextos de engenharia crítica para a segurança, os sistemas de controlo são concebidos para operar com propriedades de fiabilidade diferentes das dos sistemas que monitorizam. O mecanismo governativo é intencionalmente separado tanto em função como em comportamento, para que possa fornecer um ponto de referência estável quando ocorre incerteza ou falha.

Os padrões atuais de governação de IA muitas vezes não estabelecem essa separação. Em vez disso, sobrepõem o julgamento probabilístico à geração probabilística e tratam a estrutura resultante como um sistema de controlo. Esta abordagem assume que adicionar mais etapas de avaliação dentro da mesma classe de fiabilidade produzirá resultados de governação. Na prática, introduz recursão em vez de controlo, porque cada camada depende do mesmo tipo de processo de raciocínio e está vulnerável aos mesmos tipos de falha.

A consequência não é meramente teórica. Quando um sistema produz um resultado que causa dano, a questão que se segue não é se havia múltiplas camadas envolvidas na decisão. A questão é se alguma dessas camadas funcionou como um controlo ver-

dadeiro, ou seja, se operou com um nível diferente de determinismo, rastreabilidade e responsabilização do que o sistema que governava. Se todas as camadas partilham o mesmo fundamento probabilístico, então o sistema carece de um ponto claro onde a responsabilidade possa ser ancorada ou as decisões possam ser fielmente reconstruídas.

> **O TESTE DECISIVO**
>
> A governação define-se pela introdução de um mecanismo que opera com um perfil de fiabilidade diferente do sistema a ser governado.

A posição que decorre desta análise é direta. A governação não se define pelo número de controlos presentes num sistema, nem pela complexidade visual da sua arquitetura. A governação define-se pela introdução de um mecanismo que opera com um perfil de fiabilidade diferente do sistema a ser governado. Sem essa distinção, as camadas adicionais não constituem controlo. Elas apenas estendem a mesma incerteza a mais componentes.

Quando uma camada de controlo se comporta da mesma forma que o sistema que deveria regular, a arquitetura não alcança a governação em sentido significativo. Produz uma estrutura recursiva em que a incerteza avalia a incerteza, e a fronteira entre ação e supervisão colapsa. É nesse colapso que reside o verdadeiro risco, independentemente de quão estruturado o sistema pareça à superfície.

> **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.

Como citar

Grizzle, H. M. (2026, April 18). Quando os Sistemas de Controlo Herdam a Mesma Incerteza que Deveriam Governar. Novara Consulting Group. https://www.novaracg.com/pt/2026/04/18/when-control-systems-inherit-the-same-uncertainty-they-are-meant-to-govern/