

Os princípios de uma IA responsável

Abreviações e acrônimos

Nenhum conhecimento prévio é presumido. Cada acrônimo usado neste documento está escrito por extenso abaixo.

IA : inteligência artificial.

Por que princípios, e por que estes

Princípios não substituem controles, mas orientam cada decisão que os controles ainda não cobrem. Estes sete foram escritos para serem usados sob pressão, e não para decorar paredes.

Os sete princípios

N	Princípio	Em uma frase
1	O ser humano primeiro	A tecnologia serve às pessoas, nunca o contrário
2	A justiça como dever	Viés não é acidente técnico; é falha de governança
3	A verdade acima da concordância	Um sistema que concorda com você em vez de dizer a verdade é um risco, não um assistente
4	A transparência	As pessoas sabem quando interagem com IA e por que uma decisão foi tomada
5	A responsabilização	Cada decisão automatizada tem um responsável humano com nome
6	A contestabilidade	Toda pessoa afetada pode questionar e obter revisão humana
7	A proporção e a contenção	O sistema faz o necessário, e nada além; e pode ser parado

O terceiro princípio é o que ninguém mais nomeia

A bajulação é a falha em que o sistema aprende a agradar em vez de informar. Ela corrói decisões silenciosamente, porque cada erro chega embrulhado em concordância. O princípio 3 existe porque essa falha não aparece em nenhuma outra carta de princípios do mercado.

Como adotar

- Publique os sete princípios internamente com o nome de quem responde por cada um.

- Vincule cada princípio a um controle mensurável. Princípio sem controle é discurso.
- Revise trimestralmente: algum princípio foi violado? O que mudou por causa disso?

O próximo passo

Os princípios gratuitos definem o padrão. O mapa completo de princípio a controle, com os modelos de decisão e o relatório de diretoria, vive no AI Ethics Operating Kit da AI GOVERNANCE CHAIN™.

AI GOVERNANCE CHAIN™

Estrutura de governança de IA

© 2026 Mathieu K. Gouanou. Todos os direitos reservados.

Membro do Conselho Consultivo de Harvard Business Review.

Padrão do documento: governança de IA centrada no ser humano.