

Da Explicabilidade à Inteligência Responsável

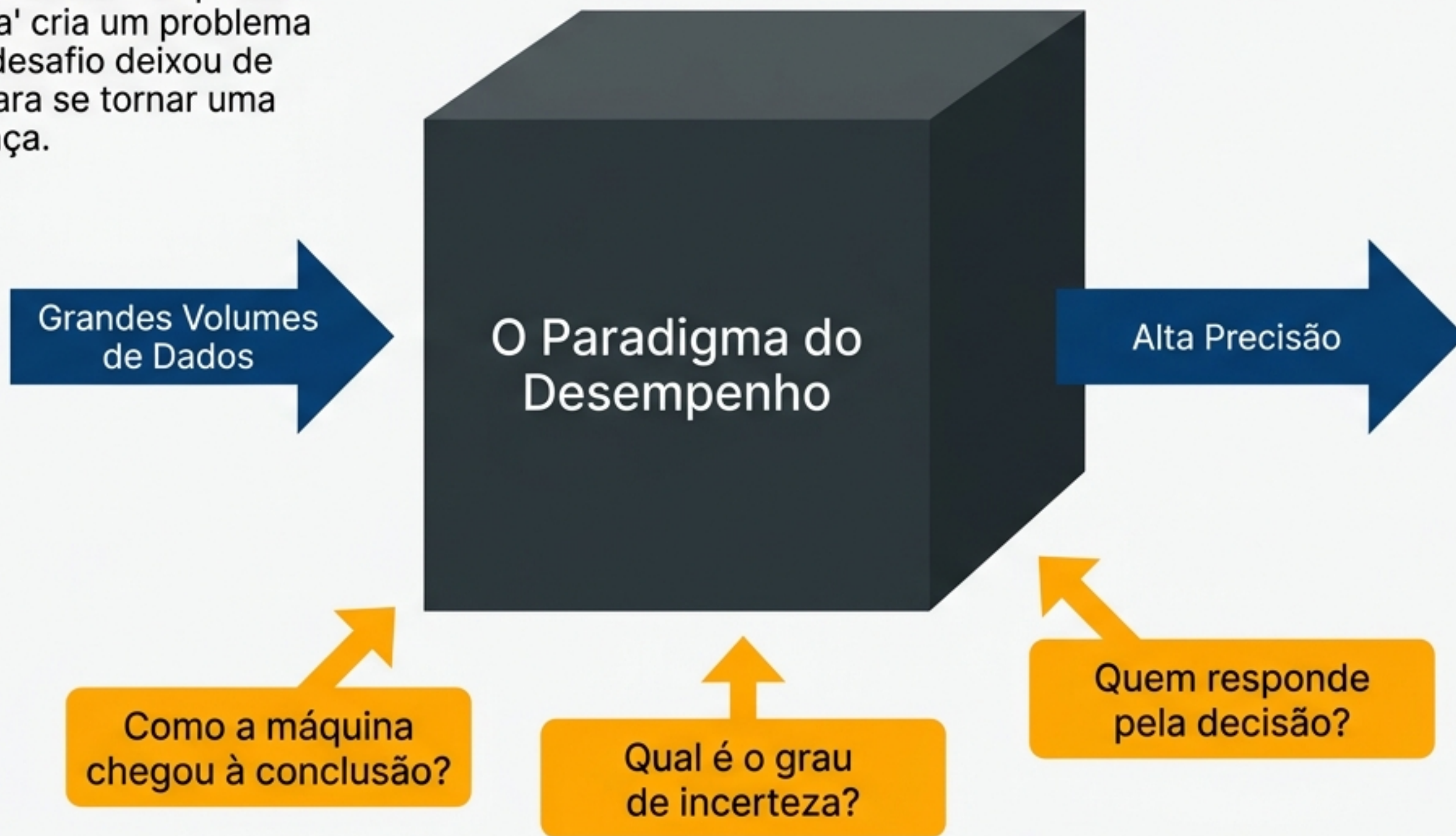
Como a Inteligência Artificial deve evoluir de um oráculo de respostas precisas para um instrumento de aprimoramento da cognição e da decisão humana.



Um manifesto estruturado pelo IVEPESP sobre o futuro da governança em IA.

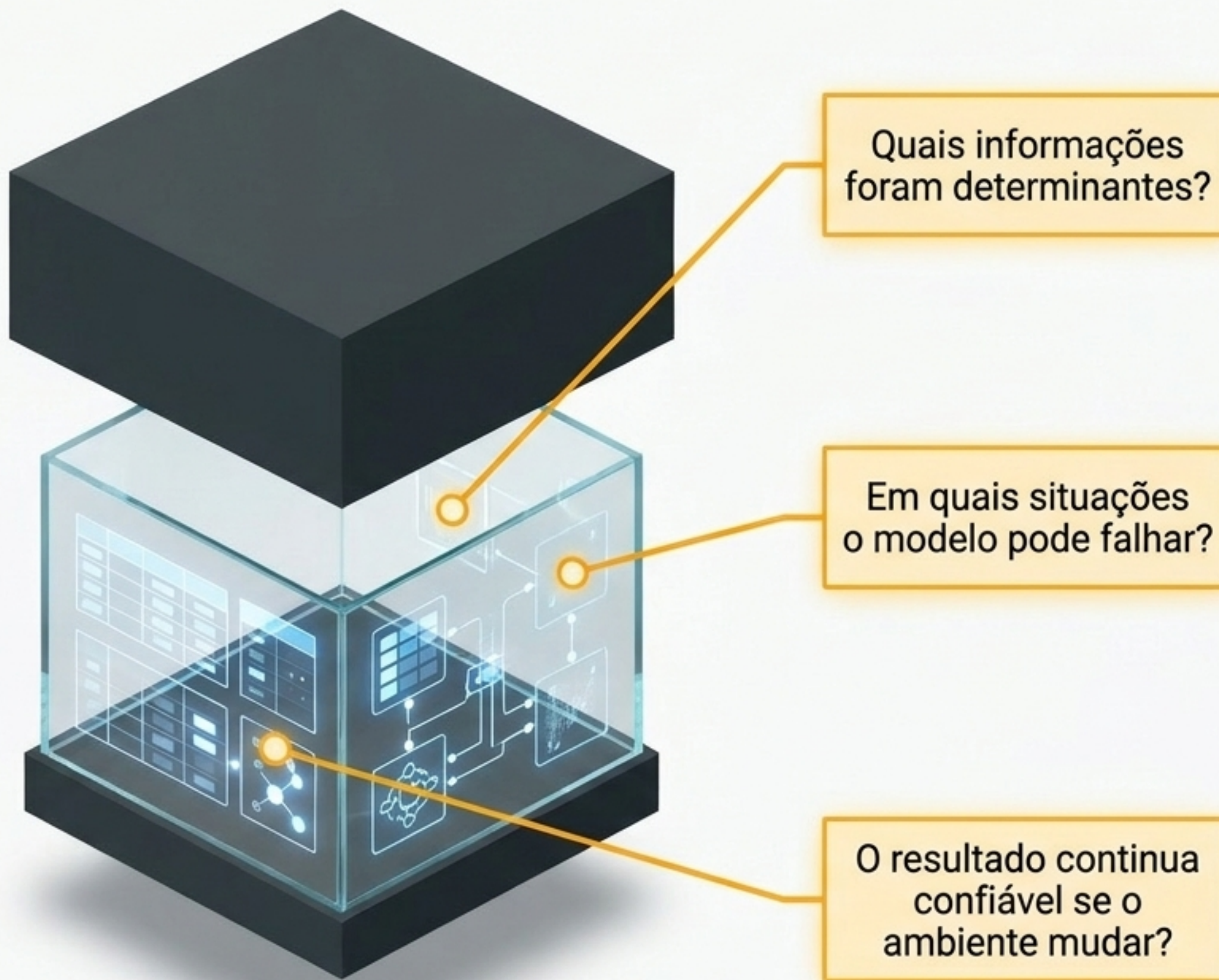
A precisão algorítmica não é mais suficiente

A rápida disseminação da IA concentrou o debate na performance. Em áreas críticas, o alto desempenho de uma 'caixa-preta' cria um problema de governança. O desafio deixou de ser a engenharia para se tornar uma questão de confiança.



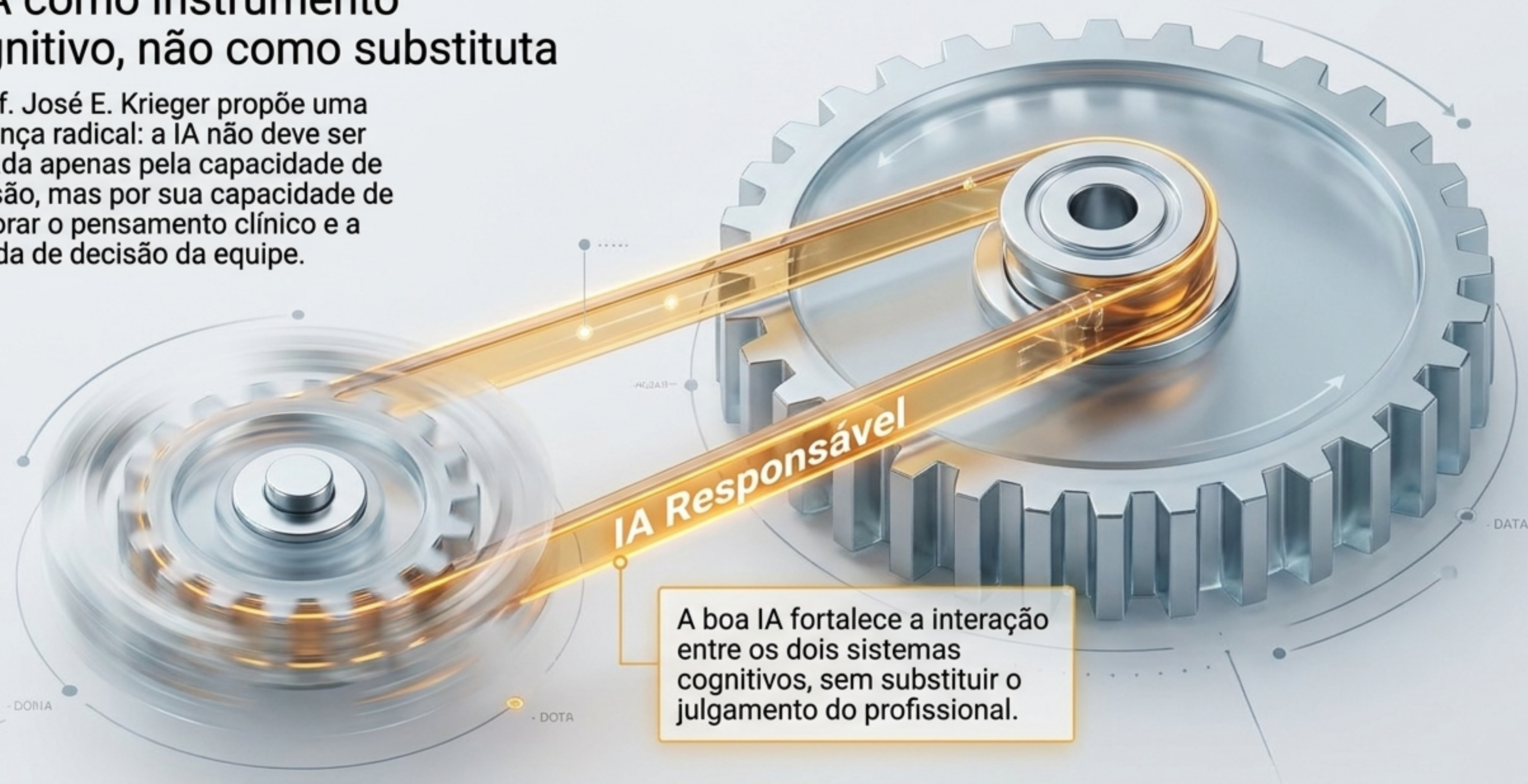
A XAI abre a caixa e revela o processo de decisão

A Inteligência Artificial Explicável (XAI) transforma o algoritmo de um executor silencioso em um sistema que justifica suas recomendações, permitindo calibrar a confiança no resultado.



A IA como instrumento cognitivo, não como substituta

O Prof. José E. Krieger propõe uma mudança radical: a IA não deve ser avaliada apenas pela capacidade de previsão, mas por sua capacidade de melhorar o pensamento clínico e a tomada de decisão da equipe.



Sistema 1: Rápido, automático e intuitivo.

Sistema 2: Lento, deliberado e analítico.

Explicar não é apenas vomitar código

A melhor explicação não é a mais técnica, mas aquela que fornece ao usuário, no momento correto, as informações para transformar o resultado algorítmico em uma ação segura.

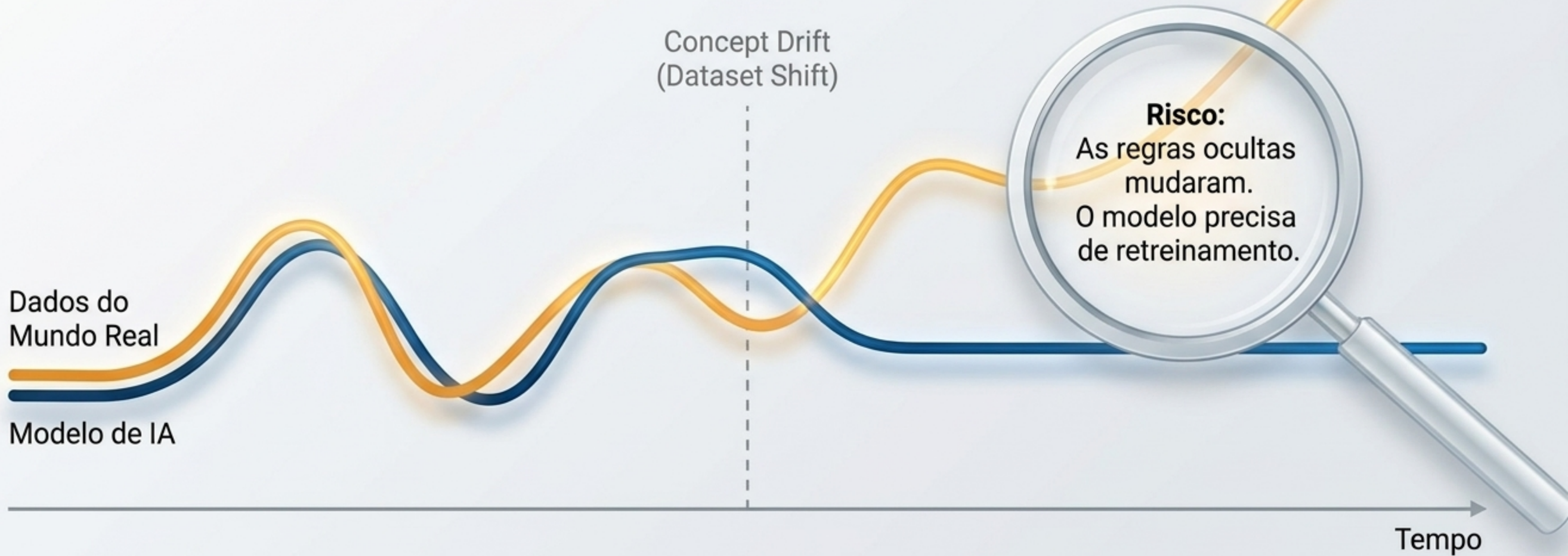
O objetivo é saber se o usuário deve **aceitar, investigar** ou **rejeitar** a recomendação.

O Espectro da Explicabilidade



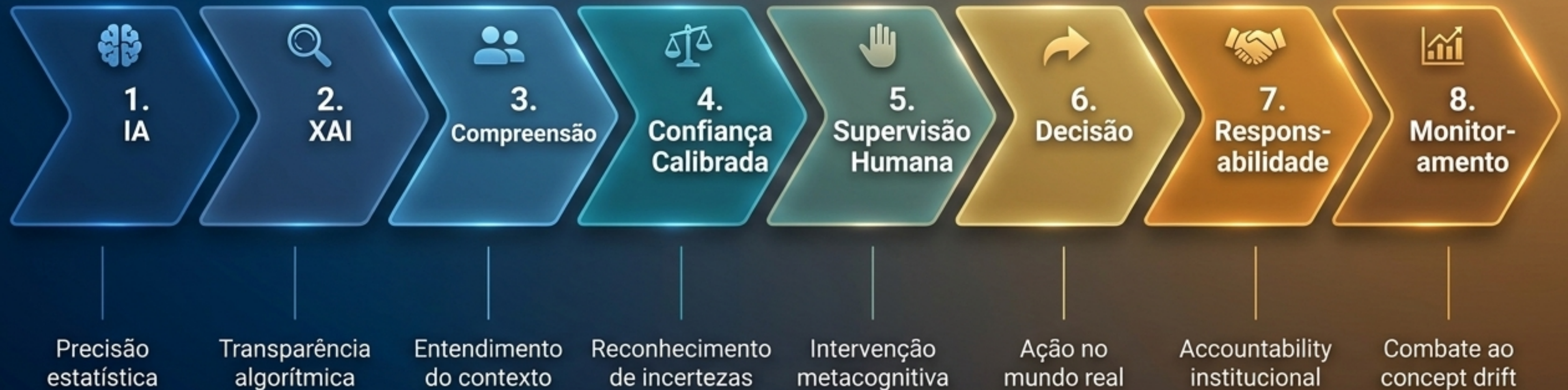
O desafio do tempo e o Concept Drift

Dados não são estáticos. A pesquisa de Hélio Henrique Villela Dias (UNIFESP/IVEPESP) demonstra que, em fluxos de aprendizado contínuo, a XAI precisa auditar como e por que as decisões mudam ao longo do tempo.



O Modelo de Maturidade da Inteligência Responsável

A explicabilidade é uma condição necessária, mas a jornada só se completa com a supervisão humana e a governança contínua.



A Matriz do Paradigma da IA

Como o papel da máquina e do humano se transformam à medida que avançamos da opacidade técnica para a governança institucional.

	IA Tradicional	XAI (IA Explicável)	Inteligência Responsável
Foco de Avaliação	Desempenho e Precisão	Justificativa da Decisão	Aprimoramento da Cognição Humana
Papel do Humano	Executor passivo de outputs	Intérprete de resultados	Supervisor com autonomia crítica
Natureza da Decisão	Aceitação automática	Decisão informada	Decisão validada com responsabilidade
Resposta à Mudança	Cega ao concept drift	Identifica falhas pontuais	Auditoria e retreinamento contínuos

A verdadeira supervisão é o poder de dizer "Não"

Quanto mais competente a máquina se torna, mais vital é a capacidade humana de conservar conhecimento e autonomia crítica para discordar dela.

A Árvore da Metacognição



Expandindo o paradigma: Da Clínica para a Sala de Aula

O IVEPESP propõe que a lógica da medicina seja aplicada à educação. O uso da IA não deve ser medido por respostas corretas, mas pelo impacto no raciocínio do aluno.



Na Medicina

A IA melhora o pensamento clínico da equipe de saúde?



Na Educação

A IA ajuda o estudante a raciocinar ou apenas terceiriza a resposta?

O professor compreende as limitações para supervisionar o sistema?

A explicação fornecida fortalece a capacidade de resolver problemas?

Como identificar vieses de aprendizagem ao longo do tempo?

A Agenda de Governança em Três Pilares

Instituições precisam avançar simultaneamente nas frentes de tecnologia, cognição e governança para que a IA ultrapasse a mera validade técnica.



Ampliar a capacidade humana, não terceirizá-la

O objetivo da Inteligência Responsável é construir sistemas que auxiliem equipes a produzir decisões mais seguras e compreensíveis.

Não entregamos a decisão à máquina; usamos a máquina para assumir responsabilidade sobre decisões melhores.

