# QUEM CONTROLA NÃO ASSINA, QUEM ASSINA NÃO CONTROLA?

## Opacidade de comando, assinatura terminal e pulverização sociotécnica da responsabilidade na inteligência artificial

*Jandislei Antonio Genova*[1]


### Resumo

O artigo examina uma assimetria possível em sistemas decisórios assistidos por inteligência artificial: atores a montante podem concentrar parcela relevante do controle sobre arquitetura, implantação e registros, enquanto a pessoa no extremo operacional recebe o resultado e assume, por assinatura ou validação, a aparência de autoria integral. Adota-se método teórico-documental, com revisão narrativa rastreável e triangulação de estudos experimentais, casos institucionais e fontes jurídicas. Propõe-se a **opacidade de comando e imputação**, definida como o desalinhamento verificável entre controle material, custódia da evidência e exposição formal do agente terminal. Como instrumento de reconstrução factual, apresenta-se a **Matriz de Correspondência Material da Responsabilidade (MCMR)**, aplicada demonstrativamente aos casos Tempe e Horizon e contrastada com um cenário analítico de ausência de opacidade. A matriz não atribui culpa nem cria regime de responsabilidade. Distingue os elementos substantivos — conduta, dever, contribuição causal e normativa, previsibilidade e capacidade de prevenção — das consequências processuais associadas à custódia da prova. Conclui-se que a assinatura humana é juridicamente relevante, mas não demonstra, isoladamente, autoria decisória material; sua força imputativa depende das condições efetivas de conhecimento, autoridade, recusa e reconstrução do processo.


**Palavras-chave:** inteligência artificial; responsabilidade algorítmica; supervisão humana; opacidade de comando e imputação; autonomia decisória.


### Abstract

**THOSE WHO CONTROL DO NOT SIGN; THOSE WHO SIGN DO NOT CONTROL? Command-and-attribution opacity, terminal signature, and the sociotechnical dispersion of responsibility in artificial intelligence**

This article examines a possible asymmetry in AI-assisted decision systems: upstream actors may concentrate relevant control over architecture, deployment, and records, while the person at the operational endpoint receives the output and, by signing or validating it, assumes the appearance of full authorship. It adopts a theoretical-documentary method, combining a traceable narrative review with experimental studies, institutional cases, and legal sources. The article proposes **command-and-attribution opacity**, defined as a verifiable misalignment between material control, evidentiary custody, and the terminal actor's formal exposure. As a factual reconstruction instrument, it introduces the **Material Correspondence of Responsibility Matrix (MCRM)**, demonstratively applied to the Tempe and Horizon cases and contrasted with an analytical scenario in which opacity is absent. The matrix neither assigns fault nor creates a liability regime. It distinguishes substantive elements—conduct, legal duty, causal and normative contribution, foreseeability, and preventive capacity—from procedural consequences associated with evidentiary custody. The article concludes that human signature is legally relevant but does not, by itself, establish material decisional authorship; its attributive force depends on effective conditions of knowledge, authority, refusal, and reconstruction of the decision-making process.


**Keywords:** artificial intelligence; algorithmic accountability; human oversight; command-and-attribution opacity; decision-making autonomy.

---

1 Advogado, escritor e pesquisador independente em Direito, inteligência artificial e governança institucional. Autor de *Autonomia Decisória na Era da IA: formação da vontade, influência cognitiva estrutural e responsabilidade jurídica em ambientes algorítmicos* (Editora Dialética, 2026).

## 1 INTRODUÇÃO

Uma decisão pode chegar ao mundo com nome, matrícula funcional, registro profissional e assinatura humana. Esses elementos identificam quem praticou o ato terminal, mas não demonstram, isoladamente, quem definiu as premissas, selecionou os dados, escolheu o modelo, ajustou os limiares, configurou a interface, alterou a versão, restringiu as alternativas ou conservou os registros necessários para reconstituir o processo. Em ambientes assistidos por inteligência artificial (IA), a visibilidade documental do signatário pode coexistir com a invisibilidade material daqueles que organizaram as condições da decisão.

O problema não está apenas em que "muitas mãos" participam de sistemas complexos. Essa dificuldade acompanha burocracias, corporações, infraestruturas e tecnologias muito anteriores à IA (Thompson, 1980; Nissenbaum, 1996). A questão específica examinada neste artigo surge quando a distribuição do poder e da informação não corresponde à distribuição da imputação. O controle relevante permanece a montante ou lateralmente na cadeia sociotécnica; o ato juridicamente reconhecível concentra-se a jusante. Em sua forma mais aguda, quem controla não aparece no documento final, enquanto quem aparece não controla plenamente aquilo que valida.

O título é deliberadamente interrogativo. Não se afirma que todo fornecedor escape à responsabilidade, que todo profissional seja um mero executor ou que a assinatura humana tenha perdido valor jurídico. A hipótese é mais restrita e verificável: **a assinatura deixa de ser indício suficiente de autoria decisória material quando o signatário não dispõe das condições informacionais, cognitivas, temporais e institucionais necessárias para revisar, rejeitar e reconstruir o resultado**.

Essa hipótese interessa simultaneamente à autonomia decisória e à responsabilidade. O agente terminal pode conservar a competência formal e, ainda assim, perder parcela das condições materiais de seu exercício. A captura não exige sua destituição: basta que o ambiente defina previamente o conjunto de informações, opções e justificativas que chegarão ao momento da escolha (Genova, 2026). Depois do dano, a mesma arquitetura pode produzir movimento inverso. O poder exercido a montante desaparece do campo imediato da imputação, enquanto o último agente humano permanece disponível para explicar, assinar e responder.

O artigo propõe denominar essa configuração **opacidade de comando e imputação**. A expressão não descreve simples incompreensibilidade matemática. Tampouco exige

conspiração, dolo ou ocultação deliberada. Designa uma assimetria institucional observável entre três planos:

1. **comando**, relativo a quem define arquitetura, finalidade, dados, parâmetros, integração e condições de uso;
2. **prova**, relativo a quem produz, conserva, interpreta e pode revelar registros sobre o funcionamento;
3. **imputação**, relativo a quem aparece como autor, revisor, executor ou responsável perante terceiros.

A tese central é:

> Em cadeias decisórias assistidas por IA, a assinatura terminal não basta para qualificar responsabilidade. A imputação jurídico-substantiva deve ser reconstruída, conforme o regime aplicável, a partir da conduta, do dever jurídico, da contribuição causal e normativa, da previsibilidade e da capacidade de prevenção. Já o domínio das evidências fundamenta deveres de preservação, exibição, cooperação e eventual redistribuição do ônus probatório; não constitui, isoladamente, fundamento de responsabilidade.

A contribuição pretendida é de síntese e operacionalização, não de ruptura *ex nihilo*.[2] A literatura já descreveu a opacidade técnica (Burrell, 2016), os limites da transparência (Ananny; Crawford, 2018), o problema das muitas mãos (Thompson, 1980), as lacunas de responsabilidade (Matthias, 2004; Santoni de Sio; Mecacci, 2021), as zonas de deformação moral (Elish, 2019), a lavagem de agência (Rubel; Castro; Pham, 2019), a supervisão humana meramente simbólica (Green, 2022; Crootof; Kaminski; Price II, 2023), o controle humano significativo (Santoni de Sio; Van den Hoven, 2018) e os horizontes limitados de responsabilização em cadeias algorítmicas (Cobbe; Veale; Singh, 2023). O conceito proposto articula esses aportes ao problema jurídico da **concentração terminal da autoria aparente**, especialmente quando a custódia da prova permanece com atores dotados de maior controle material.

Para testar a plausibilidade do argumento, o texto triangula: a) estudos experimentais sobre a influência de recomendações algorítmicas no julgamento profissional; b) pesquisas sobre atribuição de responsabilidade em decisões assistidas; c) o acidente com veículo de testes da Uber em Tempe, investigado pelo *National Transportation Safety Board* (NTSB); e d) o sistema Horizon do serviço postal britânico, utilizado como comparação institucional não fundada em aprendizado de máquina. O objetivo não é afirmar que tais situações são idênticas, mas identificar mecanismos recorrentes e delimitar seus alcances.

2 “Categoria proposta” significa síntese conceitual de trabalho submetida à crítica e à validação empírica. Não se reivindica primazia absoluta sobre toda formulação anterior acerca de comando, imputação ou cadeias sociotécnicas.

Ao final, formula-se a **Matriz de Correspondência Material da Responsabilidade (MCMR)**, instrumento heurístico-diagnóstico para confrontar o responsável formal com os centros de controle, prevenção, benefício e evidência. A matriz não substitui regras de imputação civil, administrativa, penal, trabalhista ou profissional. Sua função é impedir que a última assinatura seja tomada como explicação completa da cadeia e indicar quais fatos e documentos ainda precisam ser demonstrados.

## 2 MÉTODO, CORPUS E LIMITES DA INVESTIGAÇÃO

A pesquisa assume a forma de **ensaio teórico-documental de reconstrução conceitual**, apoiado em revisão narrativa orientada por problema e triangulação empírica e jurídica. O levantamento foi encerrado em 28 de julho de 2026. Não se trata de revisão sistemática, metanálise ou estudo de prevalência; por isso, o artigo não reivindica exaustividade bibliométrica nem generalização estatística para todos os domínios de uso da IA.

O corpus foi organizado em quatro conjuntos. O primeiro reúne literatura sobre opacidade, sistemas sociotécnicos, automação e prestação de contas. O segundo contém estudos empíricos sobre decisões profissionais assistidas por algoritmos e sobre percepções de responsabilidade. O terceiro abrange documentos oficiais relativos aos casos Uber/Tempe e Horizon. O quarto reúne fontes jurídicas brasileiras e europeias referentes a revisão automatizada, registros, prova, responsabilidade, supervisão humana e governança.

Adotaram-se três critérios de inclusão: a) relação direta com controle, supervisão, decisão, evidência ou imputação; b) possibilidade de identificação da fonte e do método; e c) capacidade de sustentar, contradizer ou limitar a hipótese. Artigos revisados por pares foram priorizados para os mecanismos comportamentais; documentos oficiais, para fatos institucionais e conteúdo normativo; obras teóricas, para distinções conceituais. Notícias e comentários públicos não foram utilizados como fundamento probatório principal.

Como trilha de auditoria da revisão narrativa, foram combinados descritores em português e inglês relativos a *algorithmic accountability*, *human oversight*, *meaningful human control*, *responsibility gaps*, *moral crumple zones*, *automation bias*, cadeias algorítmicas, decisão automatizada, prova e responsabilidade civil. A identificação partiu de obras centrais, referências regressivas e pesquisa dirigida de trabalhos posteriores que as citam. Afirmações quantitativas foram conferidas na publicação primária; fatos institucionais e conteúdo normativo, em decisões, relatórios ou repositórios oficiais.

**Quadro 1 – Parâmetros de rastreabilidade da revisão narrativa**

| Conjunto documental | Repositórios e fontes consultados | Estratégia de verificação |
|---|---|---|
| **Literatura teórica e empírica** | Páginas editoriais dos periódicos, registros DOI e listas regressivas de referências. | Combinações dos descritores indicados; conferência de método, amostra, limitações e dados na publicação primária. |
| **Casos institucionais** | NTSB; *Courts and Tribunals Judiciary*; CCRC; documentos oficiais do *Post Office Horizon IT Inquiry*. | Busca pelo evento, sistema, atores e número do processo; prioridade para relatório técnico e decisões judiciais. |
| **Direito brasileiro** | Planalto; CNJ; STJ; páginas de periódicos jurídicos com identificação editorial. | Conferência do texto vigente, do precedente qualificado e da pertinência dogmática das fontes secundárias. |
| **Direito europeu** | EUR-Lex; CURIA; EDPB e documentos do antigo Article 29 Data Protection Working Party. | Conferência de versão, autoria institucional, âmbito de aplicação e cronograma normativo. |

Fonte: elaboração do autor (2026).

Não houve protocolo PRISMA, dupla triagem independente, registro prévio ou contagem exaustiva de resultados. Por isso, o quadro oferece **rastreabilidade**, e não reprodutibilidade equivalente à de uma revisão sistemática. A opção é compatível com o caráter reconstrutivo do ensaio, mas limita qualquer pretensão de mapear a totalidade da literatura.

A triangulação obedece a uma hierarquia epistêmica. Experimentos demonstram influência sob condições controladas, mas não estabelecem automaticamente causalidade em organizações reais. Pesquisas de percepção revelam como participantes distribuem responsabilidade, mas não determinam responsabilidade jurídica. Relatórios de acidente identificam causas e fatores contribuintes em um evento específico; não constituem amostra universal. Casos históricos oferecem comparações mecanísticas, não equivalências tecnológicas. Textos normativos indicam deveres e modelos regulatórios, mas sua aplicação depende do contexto, do âmbito territorial e do regime jurídico correspondente.

Para evitar ambiguidade vocabular, o artigo distingue cinco planos. **Atribuição causal** descreve contribuição factual para um resultado; **censura moral** avalia reprovação; **prestação de contas institucional** exige explicar, justificar e corrigir; **responsabilidade jurídica** impõe consequências previstas pelo regime aplicável; e **consequências probatórias** disciplinam preservação, exibição, distribuição do ônus e valoração da falta de cooperação. A MCMR fornece um diagnóstico factual anterior a esses juízos e organiza elementos úteis a cada um deles, sem estabelecer causalidade, culpa, solidariedade ou presunção automática.

A categoria proposta é, portanto, uma **hipótese diagnóstica**. Para que exista opacidade de comando e imputação, não basta haver tecnologia complexa, divisão do trabalho ou assinatura humana. Devem ser identificáveis, em conjunto ou com intensidade relevante: a) controle material distribuído ou concentrado fora do signatário; b) assimetria de informação e de autoridade; c) custódia desigual de registros; e d) concentração da autoria ou exposição formal no extremo terminal.

O artigo tampouco presume a inocência do profissional que assina. Competência, conhecimento, previsibilidade, liberdade de recusa e cumprimento de deveres próprios continuam relevantes. O que se rejeita é a passagem automática de uma premissa documental — "assinou" — para uma conclusão totalizante — "controlou e causou sozinho".[3]

## 3 DA OPACIDADE TÉCNICA À CADEIA DE RESPONSABILIDADE

### 3.1 Três opacidades e uma quarta pergunta

Burrell (2016) distingue três fontes de opacidade em sistemas de aprendizado de máquina: segredo corporativo ou estatal; insuficiência de letramento técnico; e complexidade decorrente do modo como modelos operam em escala e alta dimensionalidade. A distinção é decisiva porque cada problema exige resposta diferente. Segredo pode demandar acesso regulatório ou auditoria independente; déficit de competência requer formação e tradução; complexidade intrínseca exige testes, documentação e avaliação de comportamento, e não apenas leitura de código.

Ananny e Crawford (2018) mostram, contudo, que tornar componentes visíveis não equivale a conhecer o sistema. Transparência pode revelar fragmentos sem explicar relações organizacionais, escolhas de finalidade ou efeitos. A abertura de uma caixa não garante inteligibilidade do conjunto, assim como uma explicação local da saída não revela quem escolheu o objetivo que o modelo otimiza.

Selbst et al. (2019) advertem que abstrair um artefato de seu contexto sociotécnico pode ocultar instituições, práticas e relações de poder que participam do resultado. Para este artigo, a unidade de análise não é apenas o modelo, mas a cadeia formada por fornecedores, integradores, organizações, gestores, profissionais, regras, interfaces e registros.

A hipótese deste artigo acrescenta uma pergunta institucional às tipologias técnicas: **opaco para quem, mas controlável por quem?** Uma arquitetura pode ser indecifrável em

3 Controle, no artigo, é graduado e específico à tarefa. Não equivale a domínio total, nem constitui, isoladamente, critério suficiente de causalidade ou responsabilidade jurídica.

todos os seus detalhes e, ainda assim, permanecer governável em aspectos fundamentais. A organização pode escolher se usa o sistema, em qual domínio, com quais dados, limiares, contratos, interfaces, métricas, testes, alertas e possibilidades de interrupção. Confundir impossibilidade de explicar cada cálculo com ausência de controle sobre a implantação dissolve decisões organizacionais em fatalidade técnica.

Do mesmo modo, a informação pode estar distribuída de forma assimétrica. O usuário terminal talvez não conheça a versão do modelo, os incidentes anteriores ou as restrições do conjunto de treinamento. O fornecedor, embora não possa antecipar cada saída, pode conhecer alterações, falhas recorrentes e métricas agregadas. O implantador pode dominar a integração e os incentivos de produtividade. O profissional pode conhecer o caso concreto. A cadeia contém conhecimentos diferentes; responsabilidade adequada exige mapear quais eram relevantes para prevenir o dano.

### 3.2 Muitas mãos, lacunas e pulverização

Thompson (1980) formulou o problema das muitas mãos na responsabilidade de agentes públicos: resultados coletivos surgem de inúmeras contribuições, dificultando localizar responsabilidade individual sem abandonar a exigência de prestação de contas. Nissenbaum (1996) transportou a preocupação para sociedades computadorizadas e identificou barreiras como a multiplicidade de participantes, a normalização dos erros, a tendência de culpar o computador e a proteção da propriedade de software sem correspondente responsabilidade.

Matthias (2004) descreveu a lacuna de responsabilidade associada a autômatos capazes de aprender, cujo comportamento pode escapar ao controle e à previsibilidade tradicionalmente exigidos para censura moral. Santoni de Sio e Mecacci (2021) demonstraram que não existe uma única lacuna, mas ao menos quatro: de culpabilidade, de responsabilização moral, de prestação pública de contas e de responsabilidade ativa. A pluralização evita que reparação, culpa, explicação e dever preventivo sejam tratados como se fossem a mesma questão.

A pulverização sociotécnica da responsabilidade ocorre quando cada participante consegue apontar para outro fragmento da cadeia: o desenvolvedor não escolheu o caso concreto; o fornecedor não definiu o fluxo de trabalho; o implantador seguiu a documentação; o gestor delegou ao especialista; o especialista recebeu a recomendação; e a máquina, personificada, “decidiu”. A existência de contribuições distintas é real. O problema aparece

quando a divisão funcional se converte em redução sistemática da prestação de contas sem que diminua o poder conjunto de produzir o resultado.

Campolo e Crawford (2020) denominam *determinismo encantado* a combinação entre pretensões de capacidade extraordinária da IA e negação de responsabilidade por seus efeitos. O sistema aparece poderoso ao justificar investimento e adoção, mas limitado ou autônomo quando se exige explicação. O discurso pode, assim, concentrar autoridade e dispersar deveres.

### 3.3 Cadeias algorítmicas e horizonte de responsabilização

Sistemas de IA raramente são produtos unitários. Modelos de base, conjuntos de dados, APIs, nuvens, interfaces, ajustes, filtros, aplicações e organizações usuárias compõem cadeias de fornecimento. Cobbe, Veale e Singh (2023) mostram que atores nessas cadeias possuem visibilidade limitada para além de seu **horizonte de responsabilização**. Contratos, abstrações técnicas e separações organizacionais restringem o que cada agente conhece e pode demonstrar.

Esse diagnóstico impede respostas ingênuas. Não basta ordenar que todos conheçam tudo. A complexidade real exige divisão de tarefas, confiança e especialização. A governança deve, porém, assegurar que a fragmentação não destrua a capacidade coletiva de reconstruir decisões. Revisabilidade significa organizar o sistema para permitir contestação, investigação e correção, mediante registros, interfaces e relações institucionais adequadas (Cobbe; Lee; Singh, 2021).

A opacidade de comando e imputação localiza um padrão dentro dessas cadeias: a limitação de visibilidade não é neutra quando os agentes que retêm controle e registros permanecem menos expostos que aquele que realiza o ato final. A categoria não substitui o horizonte de responsabilização; descreve a **assimetria direcional** entre o horizonte de quem controla e o horizonte de quem responde.

### 3.4 Delimitação em relação aos conceitos próximos

**Quadro 2 – Delimitação da categoria em relação aos conceitos próximos**

| Conceito anterior | Problema que descreve | Distinção da categoria proposta |
|---|---|---|
| **Problema das muitas mãos** | Dificuldade de atribuir responsabilidade em ações coletivas e organizacionais. | A opacidade de comando e imputação exige, além da pluralidade, desalinhamento entre controle, prova e exposição terminal. |

| Conceito anterior | Problema que descreve | Distinção da categoria proposta |
|---|---|---|
| **Lacuna de responsabilidade** | Insuficiência de categorias de culpa, prestação de contas ou dever preventivo diante da autonomia e imprevisibilidade. | A categoria proposta pode existir mesmo quando há responsáveis juridicamente identificáveis; seu foco é onde o poder e a autoria aparente são posicionados. |
| **Zona de deformação moral** | Operador humano absorve culpa por falha de sistema complexo ou automatizado. | A assinatura terminal amplia o fenômeno ao plano documental e probatório, sem pressupor que o operador seja inocente. |
| **Lavagem de agência** | Agente atribui à tecnologia uma decisão própria para reduzir sua responsabilidade. | Pressupõe desvio estratégico de agência; a opacidade proposta pode emergir sem intenção, por contratos, desenho e rotinas. |
| **Horizonte de responsabilização** | Visibilidade limitada entre atores de uma cadeia algorítmica. | A contribuição está em confrontar o horizonte técnico com a concentração formal da imputação e com a custódia da evidência. |
| **Captura sem destituição** | Preservação da competência formal com deslocamento das condições materiais da decisão. | Funciona como tese mais ampla; a opacidade de comando e imputação descreve um de seus mecanismos jurídico-institucionais. |

Fonte: elaboração do autor (2026).

A diferenciação evita duas formas de inflação conceitual. A primeira seria renomear fenômenos consolidados. A segunda seria reunir problemas distintos sob uma expressão tão ampla que nada pudesse refutá-la. A categoria proposta somente agrega valor se permitir identificar uma relação que os termos anteriores, isoladamente, não tornam suficientemente visível: **quanto mais a autoria formal se concentra no extremo da cadeia, mais necessário se torna verificar se controle e evidência seguiram o mesmo caminho**.

## 4 OPACIDADE DE COMANDO E IMPUTAÇÃO

### 4.1 Definição

Define-se **opacidade de comando e imputação** como:

> a assimetria verificável pela qual decisões relevantes sobre finalidade, desenho, dados, critérios, integração, atualização, condições de operação e preservação de registros permanecem em níveis superiores ou laterais de uma cadeia sociotécnica, enquanto a pessoa situada no ponto terminal recebe o resultado e assume, por assinatura, validação ou execução, a aparência de autoria integral.

"Comando" não significa controle absoluto de cada saída. Significa capacidade material de configurar condições que alteram probabilidades, opções, incentivos e possibilidades de correção. "Imputação" é utilizada em dois níveis: **funcional ou institucional**, quando designa autoria documental, censura organizacional, exposição pública e posição processual; e **jurídico *stricto sensu***, quando depende dos pressupostos do regime civil, administrativo, disciplinar, trabalhista ou penal aplicável. A primeira não predetermina a

segunda. “Opacidade” não exige segredo completo. Pode decorrer de contratos fragmentados, interfaces simplificadas, falta de interoperabilidade, retenção de logs, segredo industrial, mudanças de versão ou desconhecimento do próprio usuário sobre quais componentes participaram do resultado.

A definição contém três planos.

No **plano de comando**, pergunta-se quem escolheu o problema, o modelo, os dados, os limiares, as métricas, a integração, o ritmo de trabalho e o mecanismo de recusa. Quem pode atualizar, suspender ou substituir o sistema possui uma forma relevante de poder, ainda que não determine individualmente cada saída.

No **plano probatório**, pergunta-se quem detém logs, documentação de versão, testes, relatórios de incidente, trilhas de entrada e saída, registros de modificação e informações sobre falhas conhecidas. Aquele que controla a evidência controla, em parte, a reconstrução futura da causalidade.

No **plano de imputação**, identifica-se quem assina, comunica, executa, atende a vítima, responde ao empregador ou aparece no processo. É comum que esse agente seja o mais visível e o menos capaz de explicar toda a cadeia.

### 4.2 Concentração privada da capacidade probatória

A opacidade torna-se especialmente grave quando a organização afetada depende do fornecedor para produzir evidências contra o próprio fornecedor. Logs incompletos, prazos curtos de retenção, formatos proprietários, cláusulas de confidencialidade e atualizações silenciosas podem impedir a reprodução do estado do sistema no momento do evento. A prova deixa de ser simples registro do passado e passa a integrar a arquitetura de poder.

Chama-se **concentração privada da capacidade probatória** a custódia exclusiva ou predominante, por atores privados participantes da cadeia sociotécnica, dos elementos indispensáveis à reconstrução de uma decisão socialmente relevante. A categoria descreve uma posição factual, não presume ilicitude, inversão do ônus ou dever geral de publicidade. Sigilo médico, segredo de justiça, proteção de dados, segurança e propriedade intelectual impõem limites legítimos. Quando houver base jurídica, a resposta adequada é preservação, acessibilidade condicionada e auditabilidade perante sujeitos autorizados, não exposição indiscriminada.

Sem essas salvaguardas, surge um paradoxo: o signatário é plenamente identificável, mas o sistema que estruturou sua decisão não é reconstituível. A individualização documental ocupa o lugar da explicação causal.

### 4.3 Imputação terminal e a metáfora do fusível humano

Elish (2019) emprega a imagem da *moral crumple zone* para descrever situações em que o componente humano de um sistema automatizado absorve o peso moral e jurídico de uma falha, de modo semelhante à área de deformação de um veículo. A expressão **fusível humano**, utilizada aqui apenas como metáfora explicativa e não como categoria autônoma, destaca função próxima: o trabalhador permanece no circuito para interromper a corrente em caso de falha e pode tornar-se o componente mais facilmente substituível após o incidente.

A **imputação terminal** é o efeito documental dessa arquitetura. O agente final não é necessariamente escolhido como bode expiatório por decisão consciente. Sua assinatura, registro de acesso, matrícula ou presença física o tornam o ponto de menor custo para a atribuição inicial. O fornecedor é remoto; o modelo é complexo; o contrato é confidencial; os dados foram agregados; a versão mudou. O indivíduo, ao contrário, tem nome e dever profissional.

Há, contudo, uma fronteira indispensável. Reconhecer a imputação terminal não autoriza apagar deveres próprios. Um médico, magistrado, servidor, engenheiro ou analista que compreende o risco, possui tempo, acesso e poder de recusa e, ainda assim, valida resultado manifestamente inadequado pode responder por sua contribuição. O argumento é de **correspondência**, não de exoneração: a responsabilidade do terminal deve ser proporcional ao seu controle e não utilizada para encobrir contribuições superiores.

### 4.4 Uma regra diagnóstica

A hipótese pode ser condensada em quatro movimentos:

> O poder é concentrado antes do dano; a operação é distribuída durante o uso; a responsabilidade é pulverizada depois do dano; e a prova permanece com quem controlava a infraestrutura.

Essa formulação não é uma lei empírica universal. É um padrão a ser investigado. Ele estará ausente quando o agente terminal possuir acesso relevante, competência, tempo, autoridade de rejeição, alternativa operacional e registros suficientes; quando contratos distribuírem deveres de modo verificável; e quando fornecedores, implantadores e gestores permanecerem expostos de acordo com suas contribuições.

## 5 A ASSINATURA HUMANA ENTRE AUTORIA FORMAL E CONTROLE MATERIAL

### 5.1 A assinatura não é vazia — mas pode ser insuficiente

Assinar produz efeitos. A assinatura autentica, vincula, encerra procedimentos e permite identificar uma pessoa. Em contextos profissionais, pode representar juízo técnico próprio e acionamento de deveres de diligência. Seria equivocado tratá-la como simples ornamento sempre que houver IA.

O erro oposto é considerá-la prova autossuficiente de controle. A autoria formal responde à pergunta "quem exteriorizou o ato?". A expressão **autoria decisória material** é empregada, neste artigo, em sentido analítico: indica participação efetiva na formação, configuração ou controle das condições da decisão, sem converter fornecedores, gestores ou projetistas em coautores jurídicos do ato terminal. Ela exige perguntas adicionais: quem definiu as premissas, quais alternativas estavam disponíveis, que informação não chegou ao signatário, que recomendação foi apresentada primeiro, quanto tempo havia para contestá-la, qual era o custo de rejeição e quem podia modificar o sistema?

Crootof, Kaminski e Price II (2023) criticam a armadilha regulatória de inserir genericamente um humano no circuito como resposta a qualquer limitação algorítmica. Sistemas humano-máquina possuem propriedades próprias; o ser humano não corrige automaticamente a máquina, assim como a máquina não corrige automaticamente o ser humano. Green (2022), ao analisar políticas de supervisão em algoritmos governamentais, adverte que a exigência abstrata de controle humano pode legitimar sistemas problemáticos e deslocar a responsabilidade para operadores de linha de frente.

Santoni de Sio e Van den Hoven (2018) oferecem parâmetro mais exigente ao formularem o **controle humano significativo**. Sua abordagem requer, de um lado, que o sistema acompanhe razões humanas relevantes (*tracking*) e, de outro, que seus resultados possam ser reconduzidos a agentes humanos capazes de compreender seu papel e responder por ele (*tracing*). A opacidade de comando e imputação não substitui essa formulação. Acrescenta-lhe uma pergunta probatória e institucional: os agentes aos quais o resultado pode ser reconduzido também detêm poder, informação e acesso às evidências proporcionais à responsabilidade que lhes será atribuída?

A supervisão é real apenas quando pode alterar o curso da decisão. Exigir uma assinatura após organizar o fluxo para que discordar seja informacionalmente impossível, economicamente punido ou temporalmente inviável transforma controle em cerimônia.

### 5.2 Oito condições da assinatura materialmente significativa

Uma assinatura ou revisão assistida por IA possui maior valor como indício de autoria decisória material quando estão presentes, de forma proporcional ao risco, as seguintes condições:

1. **objeto e proveniência conhecidos** — o revisor sabe quais sistemas, versões e fontes participaram do resultado;
2. **acesso aos elementos relevantes** — pode consultar dados, documentos, sinais de incerteza e justificativas necessárias;
3. **conhecimento de capacidades e limites** — recebe informação atualizada sobre desempenho, falhas conhecidas e contexto de validação;
4. **competência e treinamento** — possui domínio profissional e preparação para avaliar a ferramenta;
5. **tempo e condições cognitivas adequadas** — a carga de trabalho não converte revisão em confirmação reflexa;
6. **autoridade efetiva** — pode ignorar, alterar, rejeitar, interromper ou escalar o resultado sem retaliação indevida;
7. **rota alternativa** — existe procedimento independente para decidir quando o sistema é inadequado, indisponível ou contestado;
8. **rastreabilidade posterior** — entradas, saídas, versões, intervenções e eventos relevantes são preservados e acessíveis.

Nenhuma condição é absoluta. Seu conteúdo varia entre triagem administrativa, diagnóstico, concessão de crédito, sentença, manutenção industrial e moderação de conteúdo. O ponto comum é que **responsabilidade sem capacidade correspondente degenera em imputação ritual**.

### 5.3 Influência cognitiva estrutural e assinatura

A supervisão pode falhar antes mesmo do clique. A recomendação algorítmica define uma âncora, seleciona informação, ordena hipóteses e oferece linguagem pronta. O agente conserva a faculdade formal de discordar, mas precisa produzir esforço adicional para sair do caminho sugerido. Sob pressão de tempo, autoridade percebida e repetição, o resultado automatizado pode tornar-se o padrão, e a revisão humana, a exceção.

Esse mecanismo se relaciona à **influência cognitiva estrutural**: o poder de organizar as condições nas quais a vontade se forma, sem eliminar o ato consciente de escolha (Genova,

2026). A assinatura final não demonstra que a influência foi irresistível, mas tampouco a neutraliza. Sua relevância jurídica depende de quanto o signatário pôde conhecer e contestar.

Parasuraman e Riley (1997) mostraram que o uso de automação envolve não apenas utilização adequada, mas abuso, desuso e uso indevido. Bainbridge (1983) já havia identificado uma ironia: quanto mais confiável a automação, menos o operador pratica as habilidades necessárias para intervir justamente nas situações raras e críticas. A presença humana pode, portanto, ser nominalmente preservada enquanto sua capacidade de recuperação se deteriora.

## 6 EVIDÊNCIA EMPÍRICA: O QUE A REVISÃO HUMANA CONSEGUE FILTRAR?

### 6.1 Recomendações incorretas em diagnóstico clínico

Gaube et al. (2021) realizaram experimento com 265 médicos: 138 radiologistas e 127 profissionais de medicina interna ou emergência. Cada participante avaliou oito radiografias de tórax acompanhadas de conselho diagnóstico correto ou incorreto. Todo o aconselhamento havia sido produzido por especialistas humanos, embora parte fosse apresentada como proveniente de IA. A precisão final piorou significativamente diante de conselhos incorretos, independentemente da fonte atribuída.

A análise individual é particularmente relevante. Entre os radiologistas, 27,54% seguiram todos os conselhos incorretos que receberam; entre profissionais de medicina interna ou emergência, a proporção chegou a 41,73%. Apenas 28,26% dos radiologistas e 17,32% do outro grupo rejeitaram todos os conselhos incorretos. O estudo não demonstra que médicos sejam incapazes de supervisionar IA. Mostra que competência profissional e atribuição formal da decisão não garantem, sozinhas, filtragem independente.

Rezazade Mehrizi et al. (2023) examinaram 2.760 decisões de 92 radiologistas em quinze exames de mamografia. Quando os profissionais consultaram sugestões corretas, a probabilidade de diagnóstico correto foi de 78%; quando consultaram sugestões incorretas, a probabilidade de decisão incorreta foi de 72%, isto é, a acurácia caiu para 28%. Variações nas informações de explicabilidade e na preparação atitudinal em relação à IA apresentaram efeito limitado para superar a influência das sugestões.

O resultado é empiricamente expressivo, mas deve ser lido com cautela. O experimento foi realizado fora do ambiente clínico, com monitores e condições não uniformes; a amostra combinou participantes europeus e iranianos; e quinze casos não abrangem a diversidade da prática. Ainda assim, o desenho revela que “explicar” a

recomendação e lembrar o profissional de manter atitude crítica podem não bastar quando a própria sugestão organiza o percurso analítico.

Morey, Rayo e Woods (2025) reforçaram a necessidade de avaliar o sistema conjunto. Em estudo com 450 estudantes de enfermagem e doze enfermeiros licenciados, participantes analisaram dez casos históricos com e sem recomendações e explicações algorítmicas. O desempenho melhorou quando a recomendação era correta e piorou quando ela era enganosa. Os autores propõem dois requisitos mínimos: medir empiricamente o desempenho de pessoas e IA em conjunto e incluir casos difíceis nos quais o sistema apresente desempenho forte, mediano e fraco.

Os três estudos convergem em ponto limitado, porém robusto: **não é metodologicamente aceitável inferir segurança a partir da acurácia isolada do modelo nem da simples presença de um profissional no circuito**.

### 6.2 A responsabilidade percebida também se desloca

Edwards et al. (2025) aplicaram pesquisa a 649 profissionais de radiografia, radioterapia, medicina nuclear e ultrassonografia. Os participantes geralmente aceitaram responsabilidade moral por decisões assistidas, mas 33% também atribuíram responsabilidade compartilhada a desenvolvedores e instituições. Resultados adversos, conhecimento sobre ética de IA e disciplina profissional alteraram as avaliações. Trata-se de percepção moral, não de decisão judicial; ainda assim, o estudo mostra que a atribuição social não é fixa no usuário terminal.

Tsumura e Yamada (2026), em experimento com 588 adultos japoneses, verificaram que conhecimento prévio sobre o agente e importância do tema modificaram a distribuição de responsabilidade entre usuário, agente e desenvolvedor ou fornecedor. Em situações percebidas como mais graves, aumentou a atribuição aos responsáveis pelo sistema e diminuiu a dirigida ao usuário. O desenho se baseou em observação de cenários e possui limites culturais e ecológicos. Seu valor está em demonstrar que responsabilidade percebida é sensível à apresentação da tecnologia e ao contexto, não simples reflexo da autoria formal.

Essa evidência também alerta para o movimento oposto: a personificação da IA pode transformá-la em alvo moral autônomo e reduzir a atenção aos agentes humanos. A máquina passa a “errar”, “escolher” ou “recusar”, enquanto decisões de design e implantação deixam o primeiro plano. O deslocamento pode convertê-la em bode expiatório algorítmico sem

patrimônio, deveres ou capacidade real de reparar; a expressão é descritiva e não constitui categoria autônoma adicional.

### 6.3 Tempe: falha terminal e falha organizacional podem coexistir

Em 18 de março de 2018, um veículo de testes da Uber, operando com sistema automatizado e uma operadora de segurança, atropelou e matou uma pedestre em Tempe, Arizona. O *National Transportation Safety Board* concluiu que a causa provável incluiu a falha da operadora em monitorar o ambiente e o sistema, em razão de distração com telefone celular. O mesmo relatório identificou fatores contribuintes a montante: procedimentos inadequados de avaliação de risco da Uber Advanced Technologies Group, supervisão ineficaz dos operadores, ausência de mecanismos adequados contra complacência de automação e cultura de segurança insuficiente, além de fiscalização estatal deficiente (National Transportation Safety Board, 2019).

O caso é importante porque recusa a falsa escolha entre culpa individual e falha sistêmica. A conduta da operadora importou; a arquitetura organizacional também. A presença de um ser humano preparado para intervir não compensou um sistema que dependia de vigilância contínua em condições propícias à complacência. Responsabilizar apenas a pessoa visível teria produzido explicação causal incompleta; absolver automaticamente essa pessoa seria igualmente inadequado.

Tempe fornece, assim, evidência institucional para a regra de correspondência: responsabilidade pode ser cumulativa, diferenciada e proporcional às contribuições. O último agente não precisa ser eliminado da análise para que os atores a montante ingressem nela.

### 6.4 Horizon: caso de contraste tecnológico

O sistema Horizon, utilizado pelo *Post Office* britânico, não era uma IA generativa nem constitui exemplo equivalente de aprendizado de máquina.[4] Sua utilidade é a de um **caso de contraste tecnológico**: permite verificar se o mecanismo institucional descrito pelo artigo depende ou não de características próprias da IA.

No julgamento *Bates and Others v Post Office Ltd (No. 6: Horizon Issues)*, a *High Court* concluiu que erros, defeitos e *bugs* tinham potencial para causar discrepâncias em contas de agências e que isso efetivamente ocorrera. Também reconheceu que a Fujitsu podia inserir, editar ou excluir remotamente dados transacionais sem que o subagente soubesse,

4 A comparação com Horizon é funcional, não tecnológica. Sua inclusão serve para isolar a distribuição institucional de controle, contestação e prova, evitando atribuir à IA um problema que pode existir em sistemas informatizados convencionais.

contrariando posições anteriormente sustentadas pelo *Post Office* (Reino Unido, 2019). A importância probatória do caso não reside em afirmar que toda diferença era sistêmica, mas em demonstrar que o operador local não controlava integralmente nem a infraestrutura nem os elementos necessários para contestar sua aparente infalibilidade.

Em *Hamilton and Others v Post Office Limited*, a *Court of Appeal* anulou 39 das 42 condenações examinadas. O tribunal registrou falhas de investigação e divulgação relacionadas à confiabilidade do Horizon e considerou, nos casos pertinentes, que os acusados haviam sido processados sobre fundamento técnico no qual não se podia confiar (Reino Unido, 2021). A *Criminal Cases Review Commission*, responsável pelas remessas, também destaca que os condenados não tiveram acesso adequado a informações sobre falhas do sistema (Criminal Cases Review Commission, [s.d.]).

O caso revela mecanismo anterior à IA: a organização e seu fornecedor detêm infraestrutura, registros e capacidade superior de investigação; o operador aparece como responsável pela diferença contábil; e a contestação depende de material que não está sob seu domínio. Horizon, portanto, não "confirma" por si só a categoria proposta. Ele controla sua **especificidade tecnológica**: se opacidade de comando e imputação pode surgir sem aprendizado de máquina, a IA não é causa necessária. Sistemas de IA podem ampliar o risco por escala, variabilidade, complexidade, atualizações remotas e dependência de fornecedores, mas a estrutura básica é organizacional.

## 6.5 Síntese e alcance

**Quadro 3 – Síntese das evidências empíricas e institucionais**

| Fonte e contexto | Achado relevante | Alcance e limite |
|---|---|---|
| **Gaube et al. (2021), 265 médicos** | Conselhos incorretos degradaram diagnósticos; parcelas relevantes dos profissionais seguiram todos os conselhos incorretos recebidos. | Experimento web com oito casos; demonstra suscetibilidade, não responsabilidade jurídica nem prática universal. |
| **Rezazade Mehrizi et al. (2023), 92 radiologistas e 2.760 decisões** | 78% de acerto com sugestões corretas; 72% de erro com sugestões incorretas consultadas. | Estudo exploratório fora do ambiente clínico; forte evidência de influência, sem generalização irrestrita. |
| **Morey, Rayo e Woods (2025), 450 estudantes e 12 enfermeiros** | IA correta melhorou e IA enganosa degradou o desempenho; explicações não garantiram segurança conjunta. | Predomínio de estudantes e conjunto limitado de casos; sustenta avaliação do sistema humano-IA. |
| **Edwards et al. (2025) e Tsumura e Yamada (2026)** | Resultado, conhecimento e importância do contexto alteraram a atribuição de responsabilidade. | Mede percepção moral em cenários, não imputação legal. |

| Fonte e contexto | Achado relevante | Alcance e limite |
|---|---|---|
| **NTSB (2019), acidente de Tempe** | Falha da operadora coexistiu com deficiências de risco, supervisão, cultura e fiscalização. | Estudo oficial de caso; não representa todo sistema automatizado. |
| ***Bates* (2019), *Hamilton* (2021) e CCRC, Horizon** | Decisões judiciais reconheceram falhas técnicas, assimetria informacional e deficiências de investigação e divulgação; 39 condenações foram anuladas no julgamento conjunto. | Comparação institucional não baseada em IA; controla a especificidade tecnológica, sem provar prevalência em cadeias de IA. |

Fonte: elaboração do autor (2026).

A triangulação não "prova" a categoria como lei geral. Ela sustenta seus mecanismos componentes: influência de recomendações, insuficiência da supervisão nominal, variabilidade da atribuição, coexistência de falhas individuais e sistêmicas e importância do acesso à prova. A validação direta do conceito exigirá estudos de campo que mapeiem contratos, logs, autoridade, incentivos e decisões em cadeias reais.

## 7 RESPONSABILIDADE E PROVA NO DIREITO BRASILEIRO

### 7.1 O ordenamento já contém critérios além da assinatura

O direito brasileiro não precisa esperar por uma teoria unitária da IA para reconhecer que causalidade, dever, risco, controle e prova podem estar distribuídos. Os arts. 186, 187 e 927 do Código Civil articulam ato ilícito, abuso de direito e dever de reparar (Brasil, 2002). Sua aplicação depende do regime incidente, da conduta, do nexo, do dano e, quando exigida, da culpa. A mera participação em uma cadeia não produz responsabilidade automática; por outro lado, a ausência de assinatura no documento final não elimina deveres próprios.

A doutrina civilista brasileira recomenda preservar essa decomposição analítica. Tepedino e Silva (2019) examinam os efeitos da IA sobre cada pressuposto tradicional do dever de indenizar e rejeitam soluções indiferentes à dogmática da responsabilidade civil. Mulholland (2020) enfrenta autonomia, imputabilidade e reparação sem transformar o sistema em sujeito que absorveria as escolhas de desenvolvedores, fornecedores e usuários. Em chave legislativa, Melo (2024) sustenta que as cláusulas gerais do Código Civil e do Código de Defesa do Consumidor já oferecem estrutura relevante, embora sua aplicação exija identificar a participação concreta de cada agente. Esses aportes reforçam o limite da tese: a categoria sociotécnica organiza a reconstrução dos fatos, mas não substitui ilicitude, defeito, risco, nexo ou culpa quando o regime os exigir.

Nas relações de consumo, quando caracterizadas, os arts. 12 e 14 do Código de Defesa do Consumidor disciplinam responsabilidade por defeitos de produtos e serviços, e o art. 25, § 1º, preserva solidariedade quando houver mais de um responsável pela causação do dano. O art. 6º, VIII, permite facilitação da defesa e inversão do ônus da prova nas condições legais (Brasil, 1990). Essas regras não formam um "direito geral da IA" e não se aplicam indistintamente a qualquer relação, mas demonstram que a arquitetura jurídica pode alcançar múltiplos participantes.

No processo civil, o art. 373, § 1º, do Código de Processo Civil autoriza distribuição dinâmica do ônus da prova quando as peculiaridades da causa tornarem impossível ou excessivamente difícil o encargo ordinário ou quando a parte contrária tiver maior facilidade de produzir a prova. Os arts. 396 a 400 disciplinam exibição de documento ou coisa (Brasil, 2015). Em litígios algorítmicos, esses mecanismos tornam-se centrais quando logs, documentação de versão e relatórios permanecem exclusivamente com fornecedor ou implantador. A facilidade probatória, contudo, altera a disciplina da demonstração; não converte custódia de dados em autoria do dano nem dispensa decisão fundamentada sobre os pressupostos do art. 373, § 1º.

O desafio não é inventar responsabilidade pelo simples uso de IA. É impedir que a fragmentação técnica produza impossibilidade probatória. A parte não pode ser onerada a demonstrar aquilo que o outro participante decidiu não registrar, não conservar ou não disponibilizar.

### 7.2 LGPD: decisão automatizada, registros e prestação de contas

A Lei Geral de Proteção de Dados Pessoais oferece elementos relevantes quando a cadeia envolve tratamento de dados pessoais. O art. 6º consagra transparência, prevenção e responsabilização e prestação de contas. O art. 37 exige registros das operações; o art. 38 admite relatório de impacto; o art. 42 disciplina reparação por danos causados em violação à legislação, com hipóteses de solidariedade; e o art. 46 impõe medidas de segurança desde a concepção até a execução (Brasil, 2018).

O art. 20 assegura ao titular o direito de solicitar revisão de decisões tomadas **unicamente** com base em tratamento automatizado que afetem seus interesses e obriga o controlador, mediante solicitação, a fornecer informações claras e adequadas sobre critérios e procedimentos, observados os segredos comercial e industrial. Se a informação for negada com fundamento no segredo, a Autoridade Nacional de Proteção de Dados pode realizar auditoria para verificar aspectos discriminatórios.

A redação vigente não exige expressamente que a revisão seja realizada por pessoa natural.[5] Além disso, seu campo literal alcança decisões unicamente automatizadas. Surge uma questão interpretativa: intervenção humana meramente cerimonial seria suficiente para retirar a decisão do alcance do art. 20? Uma leitura material, coerente com transparência, prevenção e prestação de contas, deve examinar se a pessoa possuía autoridade, informação e possibilidade real de modificar o resultado. Caso contrário, a presença humana pode funcionar como artifício de desautomação formal sem autonomia substantiva.

Essa conclusão não amplia o texto legal por retórica. Almada e Maranhão (2023) demonstram que o tratamento brasileiro das decisões automatizadas possui contornos próprios e não admite transplante automático das salvaguardas europeias. A tensão permanece entre uma classificação binária — automatizada ou humana — e sistemas híbridos nos quais a influência se distribui gradualmente. A solução exige análise do caso, da finalidade, do fluxo e dos deveres dos agentes de tratamento.

O precedente repetitivo do *credit scoring* oferece analogia doméstica anterior à LGPD. No REsp nº 1.419.697/RS, Tema 710, o Superior Tribunal de Justiça reconheceu a licitude do método, preservou o segredo da fórmula e, simultaneamente, assegurou ao consumidor informação clara sobre os dados utilizados, admitindo reparação quando houver abuso e dano nos termos fixados pelo regime aplicável (Brasil, 2014). O acórdão não resolve a responsabilidade por IA, mas demonstra que segredo empresarial e poder informacional podem ser compatibilizados com acesso aos elementos necessários à contestação.

### 7.3 Poder Judiciário: da autoridade final ao controle efetivo

A Resolução CNJ nº 615/2025, em texto compilado com as alterações da Resolução CNJ nº 674/2026, estabelece diretrizes para desenvolvimento, uso e governança de IA no Poder Judiciário. Entre seus elementos estão supervisão humana efetiva e proporcional ao risco, capacitação contínua, registros de fontes automatizadas e do grau de supervisão, logs de uso, auditoria, monitoramento, comunicação de eventos adversos e possibilidade de modificação dos produtos pelo magistrado (Brasil, 2025).

O Anexo classifica como baixo risco a produção de textos de apoio à elaboração de atos judiciais quando a supervisão e a versão final sejam realizadas pelo magistrado e com base em suas instruções. O modelo é normativamente importante, mas a categoria proposta neste artigo acrescenta uma cautela: **versão final e assinatura não demonstram, por si,**

5 A redação original do § 3º do art. 20, que mencionava revisão por pessoa natural, foi vetada. A redação vigente assegura solicitação de revisão, sem impor expressamente a identidade humana do revisor. A razão formal do veto consta da Mensagem nº 288, de 8 de julho de 2019 (Brasil, 2019).

**supervisão efetiva**. É necessário que o magistrado possa identificar fontes, conferir citações, compreender limitações, rejeitar a saída, registrar o uso e decidir por via alternativa.

No campo judicial, a distância entre controle formal e material possui gravidade particular. A fundamentação pertence ao julgador; a IA não pode substituir valorações sobre prova, mérito ou direitos. Porém, a obrigação individual do magistrado não exime tribunal, gestor e fornecedor de assegurar ambiente que torne possível o cumprimento. Governança institucional inadequada não se cura com a última assinatura.

### 7.4 Função dogmática limitada da MCMR

A MCMR atua antes da conclusão jurídica: organiza fatos, poderes, deveres e fontes de prova que serão qualificados pelo regime incidente. Em responsabilidade subjetiva, controle, conhecimento e possibilidade de intervenção podem informar previsibilidade, diligência e evitabilidade, mas não dispensam culpa, dano e nexo quando exigidos. Em regimes objetivos, a matriz pode auxiliar a localizar defeito, risco, atividade, cadeia de fornecimento e causas excludentes, sem criar solidariedade fora das hipóteses legais. No processo, a assimetria de informação pode justificar exibição ou distribuição dinâmica do ônus, mas não autoriza presumir automaticamente o mérito da pretensão.

No domínio médico, Nogaroli (2026) mostra que a mediação algorítmica pode elevar deveres profissionais de cuidado, informação e supervisão. Essa conclusão é compatível com o argumento deste artigo desde que não seja convertida em responsabilidade exclusiva do médico: o dever terminal permanece, enquanto fornecedor e organização respondem pelos deveres próprios que o direito lhes atribuir. A MCMR serve precisamente para manter essas perguntas separadas.

Essa limitação preserva a diferença entre **diagnóstico sociotécnico** e **imputação jurídica**. A contribuição do artigo não é formular uma cláusula geral de responsabilidade por IA, e sim impedir que a aplicação das cláusulas existentes parta de uma reconstrução factual truncada pela visibilidade do último ato.

## 8 DIREITO EUROPEU: SUPERVISÃO, DECISÃO AUTOMATIZADA E PROVA

### 8.1 Supervisão como capacidade no Regulamento de Inteligência Artificial

O Regulamento (UE) 2024/1689 oferece parâmetro comparativo. Para sistemas de alto risco, o art. 14 exige supervisão humana efetiva, proporcional ao risco, ao nível de autonomia e ao contexto. As pessoas encarregadas devem poder compreender capacidades e limitações,

permanecer conscientes da tendência de confiar automaticamente na saída, interpretar resultados, ignorá-los, substituí-los, revertê-los ou interromper o sistema. O art. 26, nº 2, exige que o implantador atribua a supervisão a pessoas com competência, formação, autoridade e apoio necessários (União Europeia, 2024a).

Esses requisitos deslocam a supervisão do plano da presença para o da capacidade. Um ser humano está materialmente no circuito quando possui meios para mudar o resultado, não apenas quando seu nome consta ao final. O Regulamento constitui referência comparativa; não é regra geral aplicável a cadeias brasileiras fora de seu âmbito territorial.[6]

Em 24 de julho de 2026, foi publicado no *Jornal Oficial da União Europeia* o Regulamento (UE) 2026/1744, em vigor desde 27 de julho de 2026, que alterou o Regulamento (UE) 2024/1689 e outros atos setoriais para simplificar a aplicação das regras harmonizadas de inteligência artificial. Entre as alterações, a aplicação do regime de alto risco foi fixada em 2 de dezembro de 2027 para os sistemas abrangidos pelo art. 6º, nº 2, e pelo Anexo III, e em 2 de agosto de 2028 para os sistemas abrangidos pelo art. 6º, nº 1, incorporados a produtos regulados do Anexo I (União Europeia, 2026). Entrada em vigor não equivale à aplicação imediata de todas as obrigações. A precisão temporal importa: os arts. 14 e 26 são empregados aqui como parâmetros normativos comparados, respeitados o âmbito material, o âmbito territorial e o cronograma escalonado.

O parâmetro europeu também enfrenta limites. Competência, autoridade e apoio podem ser declarados em política e negados na prática por metas, desenho de interface, dependência contratual ou medo de punição. A conformidade documental precisa ser confrontada com dados de uso: frequência de rejeição, tempo de revisão, incidentes, escalamentos e capacidade de operar sem a ferramenta.

### 8.2 Intervenção humana significativa e o caso SCHUFA

As diretrizes europeias sobre decisões individuais automatizadas esclarecem que a participação humana não pode ser apenas simbólica: deve ser exercida por pessoa com autoridade e competência para alterar a decisão, mediante consideração efetiva dos dados disponíveis, e não como confirmação rotineira da saída (Article 29 Data Protection Working Party, 2018). Embora elaboradas pelo antigo Article 29 Data Protection Working Party e posteriormente endossadas pelo European Data Protection Board, elas oferecem critério funcional útil para distinguir revisão substantiva de assinatura terminal.

6 O Regulamento (UE) 2024/1689 aplica-se conforme seus critérios territoriais e materiais e possui cronograma escalonado, alterado pelo Regulamento (UE) 2026/1744 para determinadas regras de alto risco. Neste artigo, seus arts. 14 e 26 são utilizados como referência comparativa de desenho regulatório.

O Tribunal de Justiça da União Europeia aprofundou essa leitura no caso *SCHUFA*. Ao interpretar o art. 22 do Regulamento Geral sobre a Proteção de Dados, considerou que a produção automatizada de uma pontuação de solvabilidade pode constituir decisão automatizada quando terceiros lhe atribuem papel determinante na celebração, execução ou cessação de uma relação contratual (União Europeia, 2023). O dado relevante para este artigo não é transpor automaticamente o precedente ao Brasil, mas registrar sua recusa a uma classificação puramente formal: a decisão não deixa de ser materialmente automatizada apenas porque um terceiro pratica o ato final.

Esse raciocínio converge com a hipótese da opacidade de comando e imputação. Entre a saída e a assinatura, é necessário perguntar se houve juízo independente ou se a arquitetura anterior já delimitara, de maneira determinante, o resultado disponível. A convergência é analítica; os pressupostos, direitos e consequências permanecem dependentes de cada regime jurídico.

### 8.3 Acesso à prova e presunções na responsabilidade por produtos

A Diretiva (UE) 2024/2853, relativa à responsabilidade por produtos defeituosos, enfrenta de modo expresso a assimetria probatória em litígios tecnicamente complexos. Seus arts. 9 e 10 preveem, sob requisitos próprios, divulgação de elementos probatórios relevantes e presunções refutáveis quando a complexidade técnica ou científica tornar excessivamente difícil demonstrar defeito ou nexo causal (União Europeia, 2024b).

A Diretiva não cria regra universal de responsabilidade para todo uso de IA, não elimina o exame causal e depende de transposição pelos Estados-membros. Sua importância comparativa é mais precisa: o direito europeu reconhece que a parte que controla informação técnica pode tornar impraticável a prova da parte lesada e admite respostas processuais proporcionais. Isso fornece suporte normativo externo à noção de **concentração privada da capacidade probatória**, sem converter custódia de dados, isoladamente, em culpa ou responsabilidade.

## 9 MATRIZ DE CORRESPONDÊNCIA MATERIAL DA RESPONSABILIDADE

### 9.1 Finalidade

A **Matriz de Correspondência Material da Responsabilidade (MCMR)** é um instrumento heurístico-diagnóstico destinado a verificar se a exposição formal de cada participante corresponde aos poderes, capacidades e deveres efetivamente presentes na cadeia. A mudança de “teste” para “matriz” é deliberada: o instrumento não passou por validação

externa, avaliação de confiabilidade entre avaliadores ou aplicação comparativa a uma amostra de casos.[7] Pode organizar avaliação de impacto, revisão contratual, auditoria, investigação interna, instrução processual e desenho regulatório, mas não produz conclusão jurídica.

A matriz não atribui culpa por soma mecânica de pontos. A responsabilidade jurídica depende do regime aplicável e de requisitos próprios. A MCMR organiza perguntas, participantes e documentos para evitar que o último ato visível encerre prematuramente a investigação e para explicitar onde a informação continua insuficiente.

## 9.2 Dimensões

**Quadro 4 – Dimensões da Matriz de Correspondência Material da Responsabilidade**

| Dimensão | Pergunta de controle | Evidência documental esperada |
|---|---|---|
| **Arquitetura e finalidade** | Quem escolheu o problema, o modelo, os dados, métricas, limiares e usos permitidos? | Avaliação de impacto; contrato; documentação técnica; atas de aprovação; registros de configuração. |
| **Operação e interrupção** | Quem podia rejeitar, alterar, suspender, escalar ou substituir o sistema no caso concreto? | Matriz de alçadas; registros de *override*; procedimento de contingência; política de não retaliação. |
| **Informação e competência** | Quem conhecia versão, limitações, falhas, incerteza e contexto de validação? | Model cards; instruções; histórico de versões; alertas; treinamentos; testes de competência. |
| **Prevenção e mitigação** | Quem tinha capacidade razoável de antecipar, detectar e reduzir o risco? | Testes conjuntos humano-IA; monitoramento; relatórios de incidente; auditorias; planos de correção. |
| **Benefício e externalidade** | Quem capturou ganho econômico, institucional ou operacional e quem suportou custos e riscos? | Modelo de negócio; incentivos; metas; contratos; indicadores de produtividade e impacto. O benefício é indicador posicional, não fundamento autônomo de responsabilidade. |
| **Prova e reconstrução** | Quem produziu, reteve e podia revelar os registros necessários para reconstituir a decisão? | Logs de entrada e saída; versão do modelo; trilha de intervenção; retenção; cadeia de custódia; acesso de auditor. |
| **Imputação terminal** | Quem assinou, comunicou ou executou o resultado, e em quais condições materiais? | Documento final; tempo de revisão; telas apresentadas; alternativas disponíveis; justificativa de concordância ou recusa. |

Fonte: elaboração do autor (2026).

A aplicação requer cinco etapas:

1. **delimitar a unidade de análise** — identificar o evento, o sistema, o período e o regime jurídico relevante;

7 A MCMR não passou por validação externa, teste de confiabilidade entre avaliadores ou aplicação comparativa em amostra de casos. Benefício econômico ou institucional é indicador de posição na cadeia, não fundamento autônomo de responsabilidade. O mesmo vale para a custódia da prova: sua relevância depende dos deveres jurídicos e das circunstâncias do caso.

2. **reconstruir a sequência temporal** — distinguir decisões anteriores à implantação, intervenções durante o uso e condutas posteriores ao evento;
3. **mapear os participantes** — fornecedor, integrador, organização implantadora, gestores, agente terminal, autoridades e terceiros relevantes;
4. **classificar a posição de cada participante em cada dimensão** — vínculo ou domínio direto e predominante (**D**), vínculo compartilhado ou condicionado (**C**), ausência de relação relevante (**A**) ou informação insuficiente (**I**);
5. **confrontar exposição formal e controle material** — registrar divergências sem produzir pontuação agregada, preservando o exame jurídico específico.

As letras não traduzem graus de culpa. Elas tornam visíveis lacunas informacionais e impedem que uma conclusão normativa seja escondida dentro de um número. Para uso institucional auditável, cada classificação deve indicar a fonte que a sustenta, o grau de incerteza e eventual divergência entre avaliadores. A concordância entre avaliadores é requisito de validação futura, não atributo já demonstrado da matriz. Concluída a classificação, confrontam-se quatro correspondências:

1. **controle–dever**: quem controlava possuía dever proporcional de prevenção?
2. **controle–informação**: quem podia agir tinha acesso ao conhecimento necessário?
3. **imputação–capacidade**: quem responde podia materialmente evitar, rejeitar ou corrigir?
4. **prova–exposição**: quem está mais exposto também possui acesso aos registros necessários para se defender e explicar, e quais consequências processuais podem decorrer da custódia assimétrica?

A matriz indica possível opacidade quando as respostas se tornam sistematicamente negativas: o signatário não controla; quem controla não registra; quem registra não revela; e a exposição terminal permanece maior que a capacidade material. Essa constatação abre a investigação jurídica; não a encerra.

### 9.3 Aplicação demonstrativa ao caso Horizon

A aplicação utiliza os achados judiciais de *Bates* e *Hamilton* e a documentação institucional da CCRC. Seu objetivo é testar se a matriz pode reconstruir um sistema informatizado convencional, sem confundi-lo com IA e sem reexaminar a responsabilidade individual dos subagentes.

**Quadro 5 – Aplicação demonstrativa da MCMR ao caso Horizon**

| Dimensão | Achado oficial relevante | Leitura diagnóstica |
|---|---|---|
| **Arquitetura e finalidade** | O *Post Office* contratou e implantou o Horizon, desenvolvido e mantido com participação da Fujitsu, como infraestrutura contábil das agências. | *Post Office* e Fujitsu: controle arquitetural e institucional relevante (**D**); subagentes: ausência de controle sobre o desenho central (**A**). |
| **Operação e interrupção** | Os subagentes operavam a unidade e registravam transações, mas não podiam modificar a infraestrutura central nem impedir intervenções remotas. | Subagentes: controle direto sobre rotinas locais e condicionado sobre discrepâncias (**C**); atores centrais: controle predominante sobre sistema e suporte (**D**). |
| **Informação e competência** | A *High Court* reconheceu que existiam falhas conhecidas, efeitos em contas e capacidades de acesso remoto que não eram plenamente conhecidas pelos subagentes. | *Post Office* e Fujitsu: informação técnica predominante (**D**); subagentes: conhecimento operacional localizado e informação técnica incompleta (**C**). |
| **Prevenção e mitigação** | *Bates* registrou erros capazes de afetar contas; *Hamilton* apontou falhas graves de investigação e divulgação antes das acusações. | Atores centrais: capacidade superior de investigar, corrigir e revelar (**D**); subagentes: capacidade limitada de detectar a origem sistêmica (**C**). |
| **Benefício e externalidade** | A infraestrutura apoiava a gestão central da rede; subagentes e suas famílias suportaram consequências financeiras, profissionais e penais das discrepâncias. | *Post Office*: benefício institucional direto (**D**); a distribuição do benefício apenas localiza posições e não determina responsabilidade. |
| **Prova e reconstrução** | Registros técnicos, históricos de falha, acesso remoto e documentos de suporte estavam sob domínio predominante do *Post Office* e da Fujitsu; a divulgação foi considerada inadequada em casos examinados. | Atores centrais: domínio probatório predominante (**D**); subagentes: ausência de controle sobre os registros centrais (**A**) e acesso apenas parcial aos registros locais (**C**). |
| **Imputação terminal** | Diferenças contábeis foram atribuídas a subagentes e serviram de fundamento a acusações; 39 condenações foram anuladas em *Hamilton*. | Subagentes: exposição terminal direta (**D**); atores centrais: exposição menos imediata apesar do controle e da prova (**C**). |

Fonte: elaboração do autor com base em Reino Unido (2019; 2021) e Criminal Cases Review Commission ([s.d.]).

A matriz torna visível a correspondência interrompida entre operação local e controle técnico central, sobretudo no eixo probatório. Ela não permite concluir que cada déficit decorreu do Horizon nem que todo subagente estivesse isento de deveres próprios. Permite concluir algo mais limitado: a aparência contábil e documental do agente terminal não podia ser tomada como explicação completa sem acesso às falhas, intervenções e registros do sistema.

## 9.4 Aplicação demonstrativa ao caso Tempe

A aplicação abaixo utiliza apenas achados expressos no relatório oficial do NTSB e possui finalidade demonstrativa.[8] Não reabre a investigação, não substitui o regime jurídico aplicável nem converte o relatório técnico em sentença de responsabilidade.

**Quadro 6 – Aplicação demonstrativa da MCMR ao caso Tempe**

| Dimensão | Achado oficial relevante | Leitura diagnóstica |
|---|---|---|
| **Arquitetura e finalidade** | A Uber ATG desenvolvia e testava o sistema automatizado; o desenho desativava recursos originais de alerta e frenagem de emergência do veículo durante a operação automatizada e não os substituía por mitigação equivalente. | Uber ATG: controle direto (**D**); operadora: ausência de controle arquitetural relevante (**A**). |
| **Operação e interrupção** | A operadora podia assumir o controle, mas o procedimento dependia de vigilância contínua; a empresa havia retirado o segundo operador anteriormente utilizado. | Operadora: controle condicionado sobre a intervenção imediata (**C**); Uber ATG: controle direto sobre procedimentos, redundância e carga de supervisão (**D**). |
| **Informação e competência** | O NTSB não identificou falta de habilitação, experiência ou conhecimento da operadora como fator causal; a empresa conhecia o projeto, o comportamento do sistema e o programa de testes. | Uber ATG: conhecimento arquitetural direto (**D**); operadora: conhecimento operacional parcial (**C**). Não há base para presumir incompetência individual. |
| **Prevenção e mitigação** | A distração da operadora foi causa provável; também contribuíram avaliação inadequada de risco, supervisão ineficaz, ausência de medidas contra complacência, cultura de segurança insuficiente e fiscalização estatal deficiente. | Operadora: capacidade imediata condicionada (**C**); Uber ATG: controle preventivo organizacional direto (**D**); Estado do Arizona: controle direto sobre sua esfera fiscalizatória (**D**). |
| **Benefício e externalidade** | O teste integrava o programa de desenvolvimento da Uber ATG, enquanto pedestres e a operadora estavam expostos aos riscos imediatos da operação em via pública. | Uber ATG: benefício institucional direto (**D**); vítima e público: ausência de controle sobre o programa (**A**). O benefício não constitui fundamento autônomo de responsabilidade. |
| **Prova e reconstrução** | A infraestrutura da empresa registrava dados do sistema e imagens; a Uber dispunha de meios para monitorar retrospectivamente o comportamento dos operadores, utilizados de forma limitada. | Uber ATG: domínio probatório direto (**D**); operadora: ausência de controle sobre a infraestrutura de registro (**A**). |
| **Imputação terminal** | A operadora era a pessoa imediatamente identificável e deixou de monitorar adequadamente; o relatório, contudo, registrou fatores contribuintes organizacionais e regulatórios. | Operadora: exposição terminal direta (**D**); demais atores: exposição menos imediata (**C**). A visibilidade terminal não esgota a cadeia. |

8 O NTSB identifica causas prováveis e fatores contribuintes para fins de segurança, mas seu relatório não adjudica culpa civil ou penal nem determina direitos e responsabilidades jurídicas das partes. A leitura realizada neste artigo preserva essa distinção.

Fonte: elaboração do autor com base em National Transportation Safety Board (2019).

O exercício revela a utilidade e o limite da MCMR. Ele impede duas simplificações simétricas: reduzir o acidente à distração da operadora ou dissolver sua conduta em uma abstração sistêmica. A classificação mostra controle terminal sobre a intervenção imediata, controle organizacional sobre arquitetura, redundância, supervisão e registros, e participação estatal na moldura de fiscalização. O resultado é um mapa de questões e evidências; a qualificação jurídica de cada contribuição continua reservada ao regime competente.

### 9.5 Contracaso analítico: quando a opacidade está ausente

Considere-se um cenário hipotético construído apenas para testar a fronteira do conceito. Um profissional independente escolhe a ferramenta, seleciona e registra a versão, configura seus parâmetros, conserva cópia das entradas e saídas, conhece as limitações relevantes e possui formação compatível. Recebe tempo suficiente, pode rejeitar a recomendação sem sanção, dispõe de rota alternativa não automatizada e toma a decisão após consultar diretamente os documentos primários. O fornecedor preserva os registros sob sua esfera e coopera com auditoria. A exposição profissional corresponde à autoridade efetivamente exercida.

Nesse contracaso, a MCMR registraria controle e informação substanciais no agente terminal, possibilidade real de prevenção, rastreabilidade distribuída e ausência de concentração unilateral da prova. Se o profissional validar uma saída manifestamente incompatível com os elementos que examinou, a categoria **opacidade de comando e imputação** não explicará adequadamente sua posição. Permanecem possíveis falhas do produto ou deveres de terceiros, mas a assinatura não será apenas cerimonial. O cenário não é evidência empírica; é um teste lógico de não aplicação que impede a categoria de absorver qualquer decisão assistida por IA.

### 9.6 Critérios de refutação

Uma categoria acadêmica precisa indicar quando **não** se aplica. A hipótese de opacidade de comando e imputação enfraquece ou é afastada quando:

- o agente terminal escolhe livremente o sistema e sua configuração relevante;
- possui informação suficiente sobre dados, versão, limitações e incerteza;
- recebe tempo e treinamento compatíveis com o risco;
- pode rejeitar, interromper ou decidir por rota alternativa;
- suas intervenções são registradas;
- fornecedores e gestores preservam e disponibilizam evidências;

- deveres contratuais e institucionais acompanham capacidades reais;
- a investigação alcança todos os participantes relevantes, sem presunção de causalidade exclusiva da assinatura.

O conceito também será empiricamente enfraquecido se estudos de campo mostrarem que, em cadeias com IA, controle, prova e exposição permanecem normalmente alinhados; que sugestões não alteram decisões sob condições reais de supervisão; ou que mecanismos existentes produzem reconstrução suficiente dos eventos. A abertura à refutação impede que qualquer falha seja retrospectivamente absorvida pela categoria.

## 10 ARQUITETURA DE GOVERNANÇA CONTRA A IMPUTAÇÃO CERIMONIAL

### 10.1 Governar o sistema conjunto

Auditar apenas o modelo é insuficiente. O objeto regulatório deve incluir o sistema humano-IA, seus contratos, incentivos, interfaces e rotinas. Raji et al. (2020) propõem auditoria algorítmica interna ao longo do ciclo de vida; Wieringa (2020) destaca que prestar contas exige definir perante quem, por quê e sobre o quê. Kroll et al. (2017) mostram que mecanismos procedimentais e técnicos podem tornar decisões algorítmicas mais verificáveis sem exigir transparência ingênua de todo o código.

Uma arquitetura mínima deve incluir:

1. inventário de sistemas, componentes, fornecedores e versões;
2. definição de finalidade e usos vedados;
3. avaliação de risco e impacto antes da implantação;
4. testes do desempenho conjunto entre pessoas e sistema, inclusive em casos adversos;
5. documentação de limites, incerteza e mudanças;
6. registro de entradas, saídas, intervenções e rejeições;
7. canais de contestação e escalamento independentes;
8. plano de continuidade sem dependência absoluta da ferramenta;
9. preservação probatória proporcional ao risco;
10. comunicação e investigação de incidentes;
11. auditoria com acesso protegido a informações relevantes;
12. distribuição contratual de deveres que não possa ser anulada pela fragmentação da cadeia.

### 10.2 O direito de recusar

Autoridade formal sem possibilidade prática de recusa não constitui supervisão. Organizações devem examinar metas de produtividade, avaliações de desempenho e sanções informais que favoreçam aceitação automática. Se o trabalhador é responsabilizado por concordar e punido por atrasar, questionar ou interromper, a governança cria incentivo contraditório.

O direito de recusa precisa ser acompanhado de rota alternativa e registro. Uma decisão pode ser escalada para revisão independente, decidida sem o sistema ou suspensa até obtenção de informação. A ausência de alternativa transforma o botão "rejeitar" em gesto sem consequência operacional.

### 10.3 Custódia compartilhada e acesso regulado à prova

Logs não devem ser acumulados sem finalidade, pois podem ampliar vigilância, riscos de segurança e tratamento de dados pessoais. A resposta à concentração privada da capacidade probatória não é retenção ilimitada. É governança proporcional: definir quais registros são necessários, por quanto tempo, sob qual integridade, com quais controles de acesso e para quais procedimentos.

Contratos devem assegurar preservação do estado relevante do sistema, histórico de versões, comunicação de incidentes, exportação em formato utilizável, acesso de auditor autorizado e cooperação em investigações. Segredo industrial pode justificar sigilo perante o público, mas não destruição da capacidade de supervisão por autoridades, peritos ou instâncias independentes.

### 10.4 Responsabilidade em camadas

Responsabilidade compartilhada não significa responsabilidade indistinta. Cada camada deve responder por seus próprios deveres:

- **fornecedor do modelo**: documentação, testes, comunicação de limitações, gestão de versões e incidentes sob seu controle;
- **integrador**: adequação da configuração, interface, segurança e interoperabilidade;
- **organização implantadora**: finalidade, avaliação de impacto, fluxo de trabalho, treinamento, recursos e incentivos;
- **gestor**: alocação de pessoas, tempo, autoridade, monitoramento e resposta a alertas;
- **profissional terminal**: diligência própria, uso dentro de competência, verificação possível, registro de discordância e recusa quando exigível;
- **autoridade pública**: fiscalização, padrões, vias de contestação e acesso à prova.

Essa lista reúne proposições indicativas de governança. A existência, o conteúdo e a exigibilidade de cada dever dependem da lei, do contrato, da atividade regulada e das competências de cada agente; a lista não cria obrigações universais por analogia.

Uma formulação juridicamente mais precisa separa dois planos:

> A responsabilidade substantiva deve ser apurada segundo os pressupostos do regime incidente, considerando conduta, dever, contribuição causal e normativa, previsibilidade e capacidade de prevenção. A custódia da evidência fundamenta deveres probatórios e consequências processuais próprias, sem funcionar como atalho para a condenação.

Controle e benefício auxiliam a localizar posições, incentivos e riscos na cadeia, mas não são, isoladamente, títulos de imputação. Essa separação impede tanto que a assinatura atraia toda a responsabilidade para o fim da cadeia quanto que o argumento sistêmico dispense a demonstração dos requisitos jurídicos.

## 11 OBJEÇÕES E RESPOSTAS

### 11.1 “O profissional sempre deve revisar; logo, o problema é individual”

O dever de revisão é real, mas pressupõe possibilidade de cumprimento. Estudos empíricos demonstram que aconselhamento incorreto pode alterar decisões de especialistas e que explicações simples não neutralizam necessariamente a influência. A resposta adequada é preservar o dever individual e exigir desenho institucional que o torne exequível. Atribuir integralmente a falha ao profissional sem examinar informação, tempo, autoridade e interface converte um dever em ficção.

### 11.2 “Ninguém controla modelos complexos; portanto, não há comando”

Controle não é previsão perfeita. Organizações controlam escolhas de implantação, finalidade, domínio, versão, dados, limiares, acesso, monitoramento e interrupção. A imprevisibilidade de uma saída específica pode reduzir censura em determinado aspecto e aumentar o dever de teste, limitação ou não utilização. Não conhecer tudo não equivale a não decidir nada.

### 11.3 “A categoria incentiva fuga de responsabilidade do usuário”

Esse risco existe. Por isso, a MCMR inclui a conduta terminal e rejeita imunidade presumida. O conceito não transfere automaticamente culpa a fornecedores. Ele exige que a participação de todos seja examinada segundo critérios comuns. Um usuário que tinha conhecimento, autoridade e oportunidade pode continuar responsável; a existência de falhas a montante não elimina sua contribuição.

### 11.4 “Segredo industrial impede abertura”

O artigo não exige publicidade universal de código, dados ou modelos. Exige preservação e acesso regulado a evidências pertinentes, com confidencialidade, perícia e proporcionalidade. Segredo legítimo protege competitividade; não deve tornar direitos e fiscalização impraticáveis.

### 11.5 “Mais logs resolvem o problema”

Registros são necessários, não suficientes. Um volume incompreensível pode produzir opacidade por excesso. Logs podem omitir decisões organizacionais, não capturar conversas ou ser desenhados sem integridade. A governança precisa selecionar eventos relevantes, vincular versões, preservar cadeia de custódia e permitir interpretação.

### 11.6 “Os estudos médicos não se aplicam ao direito, governo ou empresas”

Não se aplicam diretamente. A extrapolação seria indevida. Eles sustentam mecanismo cognitivo limitado: recomendações podem influenciar profissionais e a presença humana não garante independência. A intensidade do efeito deve ser testada em cada domínio. Por isso, o artigo triangula experimentos, estudos de percepção e casos institucionais, sem apresentar a saúde como amostra de toda sociedade.

## 12 AUTONOMIA DECISÓRIA E PROGRAMA DE PESQUISA

Em *Autonomia Decisória na Era da IA*, a autonomia é examinada não apenas como conservação formal da escolha, mas como domínio das condições que tornam a escolha possível (Genova, 2026). A opacidade de comando e imputação desenvolve a face institucional dessa tese. A pessoa pode permanecer habilitada a decidir e assinar enquanto fontes, ordem das informações, opções, tempo e infraestrutura são organizados por terceiros.

O deslocamento possui dois momentos:

1. **antes da decisão**, a influência cognitiva estrutural organiza o campo no qual a vontade se forma;
2. **depois da decisão**, a opacidade de comando e imputação reorganiza o campo no qual a responsabilidade será reconstruída.

A simetria é perturbadora. A montante, o poder é invisível porque aparece como infraestrutura. A jusante, a responsabilidade é visível porque aparece como assinatura. A captura sem destituição preserva a pessoa justamente onde ela é mais útil à cadeia: como fonte de legitimidade antes do dano e como ponto de imputação depois dele.

Essa formulação demanda investigação futura. Ao menos cinco agendas são necessárias:

- estudos etnográficos e organizacionais sobre como profissionais realmente revisam saídas;
- análise de contratos e cadeias de fornecimento para mapear controle e deveres;
- experimentos sobre tempo, interface, autoridade de recusa e rotas alternativas;
- pesquisa processual sobre acesso, preservação e perda de evidências algorítmicas;
- estudos comparados sobre como tribunais e reguladores distinguem intervenção humana real de intervenção cerimonial.

Também será necessário investigar desigualdades. Profissionais com menor poder institucional podem receber maior exposição terminal, enquanto organizações com capacidade contratual preservam controle e distância. A questão não é apenas quem sabe mais sobre a IA, mas quem pode dizer "não", quem suporta o custo da recusa e quem conserva a prova.

## 13 CONCLUSÃO

A inteligência artificial não elimina o ser humano da cadeia. Frequentemente, ela o reposiciona. Fornecedores desenvolvem e atualizam; organizações integram e definem metas; gestores estruturam fluxos; profissionais recebem recomendações; e alguém, no final, assina. A permanência dessa assinatura pode criar conforto normativo: se um humano decidiu, haveria controle; se há controle, a responsabilidade estaria localizada.

Os dados examinados recomendam cautela. Médicos e radiologistas podem ser influenciados por recomendações incorretas, mesmo quando conservam a decisão final. Explicações e treinamento atitudinal simples nem sempre neutralizam o efeito. Casos institucionais mostram que falhas terminais podem coexistir com deficiências de desenho, supervisão, cultura, fiscalização e prova. Normas contemporâneas avançam ao exigir competência, autoridade, registros, capacidade de modificação e supervisão efetiva.

A contribuição do artigo é nomear e operacionalizar a **opacidade de comando e imputação**: o desalinhamento entre quem configura condições e conserva evidências e quem recebe a autoria aparente do ato final. Sua manifestação extrema é a assinatura terminal sem controle material; seu correlato probatório é a concentração privada da capacidade de reconstruir o evento; seu efeito é a pulverização da responsabilidade após uma prévia concentração de poder.

A MCMR traduziu a hipótese em perguntas auditáveis e, nas aplicações demonstrativas, tornou comparáveis Tempe e Horizon sem tratá-los como eventos equivalentes. O contracaso analítico mostrou a fronteira da categoria: quando conhecimento, autoridade, rota alternativa, registros e exposição estão materialmente alinhados no agente terminal, a assinatura recupera valor robusto como indício de autoria decisória.

O conceito não exonera quem assina e não condena automaticamente quem fornece. Ele exige uma investigação mais difícil e mais justa. A pergunta não termina em “quem clicou?”. Deve alcançar quem praticou a conduta, qual dever incidia, quem podia prevenir, interromper, registrar e explicar, e qual relação causal pode ser demonstrada. Em sistemas sociotécnicos, a última mão pode ser apenas a mais visível. A imputação juridicamente legítima começa quando o direito deixa de confundir visibilidade com controle — e também quando se recusa a confundir controle com responsabilidade automática.

## REFERÊNCIAS

ALMADA, Marco; MARANHÃO, Juliano Souza de Albuquerque. Contribuições e limites da Lei Geral de Proteção de Dados para a regulação da inteligência artificial no Brasil. *Direito Público*, Brasília, v. 20, n. 106, 2023. DOI: https://doi.org/10.11117/rdp.v20i106.6957.

ANANNY, Mike; CRAWFORD, Kate. Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, v. 20, n. 3, p. 973-989, 2018. DOI: https://doi.org/10.1177/1461444816676645.

ARTICLE 29 DATA PROTECTION WORKING PARTY. *Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679*. WP251rev.01. Brussels, 2018. Adopted on 3 October 2017; last revised on 6 February 2018; endorsed by the European Data Protection Board on 25 May 2018. Disponível em: https://www.edpb.europa.eu/documents/guideline/automated-decision-making-and-profiling_en. Acesso em: 28 jul. 2026.

BAINBRIDGE, Lisanne. Ironies of automation. *Automatica*, v. 19, n. 6, p. 775-779, 1983. DOI: https://doi.org/10.1016/0005-1098(83)90046-8.

BRASIL. Conselho Nacional de Justiça. **Resolução nº 615, de 11 de março de 2025**. Estabelece diretrizes para o desenvolvimento, utilização e governança de soluções desenvolvidas com recursos de inteligência artificial no Poder Judiciário. Brasília, DF: CNJ, 2025. Texto compilado com as alterações da Resolução CNJ nº 674, de 25 de março de 2026. Disponível em: https://atos.cnj.jus.br/atos/detalhar/6001. Acesso em: 28 jul. 2026.

BRASIL. **Lei nº 8.078, de 11 de setembro de 1990**. Dispõe sobre a proteção do consumidor e dá outras providências. Brasília, DF: Presidência da República, 1990. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm. Acesso em: 27 jul. 2026.

BRASIL. **Lei nº 10.406, de 10 de janeiro de 2002**. Institui o Código Civil. Brasília, DF: Presidência da República, 2002. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2002/l10406compilada.htm. Acesso em: 27 jul. 2026.

BRASIL. **Lei nº 13.105, de 16 de março de 2015**. Código de Processo Civil. Brasília, DF: Presidência da República, 2015. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm. Acesso em: 27 jul. 2026.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709compilado.htm. Acesso em: 27 jul. 2026.

BRASIL. **Mensagem nº 288, de 8 de julho de 2019**. Veto parcial ao Projeto de Lei nº 7.883, de 2017 (nº 53/2018 no Senado Federal), que altera a Lei nº 13.709, de 14 de agosto de 2018. Brasília, DF: Presidência da República, 2019. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/msg/vep/vep-288.htm. Acesso em: 28 jul. 2026.

BRASIL. Superior Tribunal de Justiça (2. Seção). **Recurso Especial nº 1.419.697/RS**. Tema Repetitivo 710. Relator: Ministro Paulo de Tarso Sanseverino, julgado em 12 nov. 2014, DJe 17 nov. 2014. Brasília, DF: STJ, 2014. Disponível em: https://processo.stj.jus.br/repetitivos/temas_repetitivos/pesquisa.jsp?novaConsulta=true&num_processo_classe=1419697&sg_classe=REsp&tipo_pesquisa=T. Acesso em: 28 jul. 2026.

BURRELL, Jenna. How the machine "thinks": understanding opacity in machine learning algorithms. *Big Data & Society*, v. 3, n. 1, p. 1-12, 2016. DOI: https://doi.org/10.1177/2053951715622512.

CAMPOLO, Alexander; CRAWFORD, Kate. Enchanted determinism: power without responsibility in artificial intelligence. *Engaging Science, Technology, and Society*, v. 6, p. 1-19, 2020. DOI: https://doi.org/10.17351/ests2020.277.

COBBE, Jennifer; LEE, Michelle Seng Ah; SINGH, Jatinder. Reviewable automated decision-making: a framework for accountable algorithmic systems. In: ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2021, evento virtual. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York: ACM, 2021. p. 598-609. DOI: https://doi.org/10.1145/3442188.3445921.

COBBE, Jennifer; VEALE, Michael; SINGH, Jatinder. Understanding accountability in algorithmic supply chains. In: ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2023, Chicago. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. New York: ACM, 2023. p. 1186-1197. DOI: https://doi.org/10.1145/3593013.3594073.

CRIMINAL CASES REVIEW COMMISSION. **Post Office "Horizon" cases**. Birmingham: CCRC, [s.d.]. Disponível em: https://ccrc.gov.uk/post-office-horizon-cases/. Acesso em: 28 jul. 2026.

CROOTOF, Rebecca; KAMINSKI, Margot E.; PRICE II, W. Nicholson. Humans in the loop. *Vanderbilt Law Review*, v. 76, n. 2, p. 429-510, 2023. Disponível em: https://scholarship.law.vanderbilt.edu/vlr/vol76/iss2/2/. Acesso em: 27 jul. 2026.

EDWARDS, Christopher et al. The role of patient outcomes in shaping moral responsibility in AI-supported decision making. *Radiography*, v. 31, n. 3, art. 102948, 2025. DOI: https://doi.org/10.1016/j.radi.2025.102948.

ELISH, Madeleine Clare. Moral crumple zones: cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society*, v. 5, p. 40-60, 2019. DOI: https://doi.org/10.17351/ests2019.260.

GAUBE, Susanne et al. Do as AI say: susceptibility in deployment of clinical decision-aids. *npj Digital Medicine*, v. 4, art. 31, 2021. DOI: https://doi.org/10.1038/s41746-021-00385-9.

GENOVA, Jandislei Antonio. *Autonomia decisória na era da IA: formação da vontade, influência cognitiva estrutural e responsabilidade jurídica em ambientes algorítmicos*. Belo Horizonte: Editora Dialética, 2026. ISBN 978-65-288-0467-2.

GREEN, Ben. The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, v. 45, art. 105681, 2022. DOI: https://doi.org/10.1016/j.clsr.2022.105681.

KROLL, Joshua A. et al. Accountable algorithms. *University of Pennsylvania Law Review*, v. 165, n. 3, p. 633-705, 2017.

MATTHIAS, Andreas. The responsibility gap: ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, v. 6, n. 3, p. 175-183, 2004. DOI: https://doi.org/10.1007/s10676-004-3422-1.

MELO, Gustavo da Silva. Inteligência artificial e responsabilidade civil: uma análise do anteprojeto do Marco Legal da Inteligência Artificial e do Projeto de Lei 2338/2023. *Revista IBERC*, Belo Horizonte, v. 7, n. 1, p. 49-65, 2024. DOI: https://doi.org/10.37963/iberc.v7i1.271.

MOREY, Dane A.; RAYO, Michael F.; WOODS, David D. Empirically derived evaluation requirements for responsible deployments of AI in safety-critical settings. *npj Digital Medicine*, v. 8, art. 374, 2025. DOI: https://doi.org/10.1038/s41746-025-01784-y.

MULHOLLAND, Caitlin. Responsabilidade civil e processos decisórios autônomos em sistemas de inteligência artificial (IA): autonomia, imputabilidade e responsabilidade. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. 2. ed. São Paulo: Thomson Reuters Brasil, 2020. p. 327-350.

NATIONAL TRANSPORTATION SAFETY BOARD. *Collision between vehicle controlled by developmental automated driving system and pedestrian: Tempe, Arizona, March 18, 2018*.

Washington, DC: NTSB, 2019. (Highway Accident Report NTSB/HAR-19/03). Disponível em: https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1903.pdf. Acesso em: 28 jul. 2026.

NISSENBAUM, Helen. Accountability in a computerized society. *Science and Engineering Ethics*, v. 2, n. 1, p. 25-42, 1996. DOI: https://doi.org/10.1007/BF02639315.

NOGAROLI, Rafaella. Responsabilidade civil médica e diagnóstico clínico mediado por inteligência artificial: padrões globais de conduta profissional e lições do caso judicial *Sampson v. Heartwise Health Systems* (2023). *Revista IBERC*, Belo Horizonte, v. 9, n. 1, p. 110-139, 2026. DOI: https://doi.org/10.37963/iberc.v9i1.384.

PARASURAMAN, Raja; RILEY, Victor. Humans and automation: use, misuse, disuse, abuse. *Human Factors*, v. 39, n. 2, p. 230-253, 1997. DOI: https://doi.org/10.1518/001872097778543886.

RAJI, Inioluwa Deborah et al. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2020, Barcelona. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York: ACM, 2020. p. 33-44. DOI: https://doi.org/10.1145/3351095.3372873.

REINO UNIDO. Court of Appeal (Criminal Division). **Hamilton and others v Post Office Limited**, [2021] EWCA Crim 577. Londres, 23 abr. 2021. Disponível em: https://www.judiciary.uk/judgments/hamilton-others-v-post-office-limited/. Acesso em: 28 jul. 2026.

REINO UNIDO. High Court of Justice (Queen's Bench Division). **Bates and others v Post Office Ltd (No. 6: Horizon Issues)**, [2019] EWHC 3408 (QB). Londres, 16 dez. 2019. Disponível em: https://www.judiciary.uk/judgments/bates-others-v-post-office/. Acesso em: 28 jul. 2026.

REZAZADE MEHRIZI, Mohammad H. et al. The impact of AI suggestions on radiologists' decisions: a pilot study of explainability and attitudinal priming interventions in mammography examination. *Scientific Reports*, v. 13, art. 9230, 2023. DOI: https://doi.org/10.1038/s41598-023-36435-3.

RUBEL, Alan; CASTRO, Clinton; PHAM, Adam. Agency laundering and information technologies. *Ethical Theory and Moral Practice*, v. 22, n. 4, p. 1017-1041, 2019. DOI: https://doi.org/10.1007/s10677-019-10030-w.

SANTONI DE SIO, Filippo; MECACCI, Giulio. Four responsibility gaps with artificial intelligence: why they matter and how to address them. *Philosophy & Technology*, v. 34, n. 4, p. 1057-1084, 2021. DOI: https://doi.org/10.1007/s13347-021-00450-x.

SANTONI DE SIO, Filippo; VAN DEN HOVEN, Jeroen. Meaningful human control over autonomous systems: a philosophical account. *Frontiers in Robotics and AI*, v. 5, art. 15, 2018. DOI: https://doi.org/10.3389/frobt.2018.00015.

SELBST, Andrew D. et al. Fairness and abstraction in sociotechnical systems. In: CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2019, Atlanta. *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*. New York: ACM, 2019. p. 59-68. DOI: https://doi.org/10.1145/3287560.3287598.

TEPEDINO, Gustavo; SILVA, Rodrigo da Guia. Desafios da inteligência artificial em matéria de responsabilidade civil. *Revista Brasileira de Direito Civil*, Belo Horizonte, v. 21, n. 3, p. 61-86, jul./set. 2019. DOI: https://doi.org/10.33242/rbdc.2019.03.004.

THOMPSON, Dennis F. Moral responsibility of public officials: the problem of many hands. *American Political Science Review*, v. 74, n. 4, p. 905-916, 1980. DOI: https://doi.org/10.2307/1954312.

TSUMURA, Takahiro; YAMADA, Seiji. Effects of knowledge and importance on responsibility in human-AI decision making. *Scientific Reports*, v. 16, art. 2670, 2026. DOI: https://doi.org/10.1038/s41598-025-32513-w.

UNIÃO EUROPEIA. **Tribunal de Justiça. Acórdão de 7 de dezembro de 2023, OQ contra Land Hessen, processo C-634/21**. ECLI:EU:C:2023:957. Luxemburgo: Tribunal de Justiça da União Europeia, 2023. Disponível em: https://infocuria.curia.europa.eu/tabs/redirect/juris/liste.jsf?language=pt&num=C-634%2F21. Acesso em: 28 jul. 2026.

UNIÃO EUROPEIA. **Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024**, que cria regras harmonizadas em matéria de inteligência artificial. *Jornal Oficial da União Europeia*, 12 jul. 2024a. Disponível em: https://eur-lex.europa.eu/eli/reg/2024/1689/oj. Acesso em: 28 jul. 2026.

UNIÃO EUROPEIA. **Diretiva (UE) 2024/2853 do Parlamento Europeu e do Conselho, de 23 de outubro de 2024**, relativa à responsabilidade decorrente dos produtos defeituosos. *Jornal Oficial da União Europeia*, 18 nov. 2024b. Disponível em: https://eur-lex.europa.eu/eli/dir/2024/2853/oj. Acesso em: 28 jul. 2026.

UNIÃO EUROPEIA. **Regulamento (UE) 2026/1744 do Parlamento Europeu e do Conselho, de 8 de julho de 2026**, que altera os Regulamentos (UE) 2024/1689, (UE) 2018/1139 e (UE) 2023/1230 no que diz respeito à simplificação da aplicação das regras harmonizadas em matéria de inteligência artificial (Regulamento Omnibus Digital em matéria de IA). *Jornal Oficial da União Europeia*, série L, 24 jul. 2026. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32026R1744. Acesso em: 28 jul. 2026.

WIERINGA, Maranke. What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In: CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2020, Barcelona. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York: ACM, 2020. p. 1-18. DOI: https://doi.org/10.1145/3351095.3372833.