ARTIGO JURÍDICO • INTELIGÊNCIA ARTIFICIAL E GOVERNANÇA

Quem pode contrariar a inteligência artificial?

A divergência humana protegida como condição jurídica da supervisão efetiva

Uma reconstrução do chamado direito ao desalinhamento algorítmico

JANDISLEI ANTONIO GENOVA

Advogado e pesquisador independente | São Paulo, agosto de 2026

> ***Nos contextos em que a supervisão humana efetiva é juridicamente exigida, a organização não deve neutralizá-la impondo desvantagem fundada apenas na divergência profissional de boa-fé.***

Resumo

O artigo investiga a seguinte questão: quando normas ou compromissos juridicamente vinculantes exigem supervisão humana de sistemas de inteligência artificial, quais garantias podem proteger a pessoa encarregada de contrariar a saída automatizada? O método combina análise jurídico-dogmática reconstrutiva, comparação funcional e revisão dirigida, não sistemática, com protocolo declarado e corpus encerrado em 19 de agosto de 2026. A literatura já desenvolveu condições de supervisão efetiva e formulou a necessidade de proteger a divergência justificada. A contribuição residual consiste em sistematizar uma arquitetura jurídica: distinguir dever institucional, interesse funcionalmente protegido e eventual direito subjetivo; alocar sujeitos obrigados por esfera de controle; e delimitar causalidade, prova, limites e tutelas. Sustenta-se que, nos regimes em que a supervisão efetiva é exigível, uma medida adversa fundada exclusiva ou substancialmente na divergência de boa-fé pode esvaziar a própria salvaguarda. A qualificação como direito subjetivo depende da fonte normativa, do vínculo e do interesse tutelado. Propõe-se, ao final, uma Matriz Diagnóstica Preliminar, qualitativa e ainda não validada empiricamente.

Palavras-chave: inteligência artificial; supervisão humana; divergência protegida; não retaliação; ônus da prova; governança algorítmica.

Abstract

This article asks which safeguards may protect a person tasked with challenging an automated output when legal rules or binding institutional commitments require human oversight of artificial intelligence systems. The method combines reconstructive legal-dogmatic analysis, functional comparison, and a targeted, non-systematic literature review under a disclosed protocol, with sources reviewed up to 19 August 2026. Prior scholarship has already developed the conditions for effective oversight and the need to protect justified divergence. The article's residual

contribution is to systematise a legal architecture that distinguishes an institutional duty, a functionally protected interest, and a possible claim-right; allocates duty-bearers according to their sphere of control; and specifies causation, evidence, limits, and remedies. It argues that, where effective oversight is legally required, an adverse measure based exclusively or substantially on good-faith divergence may defeat the safeguard itself. Whether the position amounts to an enforceable subjective right depends on the applicable legal source, relationship, and protected interest. The article concludes with a qualitative and not yet empirically validated Preliminary Diagnostic Matrix.

Keywords: artificial intelligence; human oversight; protected divergence; non-retaliation; burden of proof; algorithmic governance.

1. Introdução: a decisão formalmente humana

Uma interface pode reservar a palavra final a uma pessoa e, ao mesmo tempo, tornar impraticável o exercício dessa autoridade. A aceitação da saída algorítmica ocorre em segundos; a rejeição exige abrir telas adicionais, localizar os dados do caso, justificar a discordância, atrasar a fila e explicar-se perante a chefia. O sistema registra um humano no circuito, mas o desenho do trabalho favorece uma resposta específica.

O problema jurídico não está na mera diferença de esforço entre dois botões. Ele surge quando a organização invoca supervisão humana como salvaguarda, transfere responsabilidade ao profissional e, por suas métricas ou relações hierárquicas, cria desvantagem para quem exerce a revisão que declarou existir. A pergunta central deste artigo é delimitada: segundo a fonte normativa e o vínculo aplicáveis, pode a instituição penalizar a divergência fundamentada que materializa a supervisão exigida?

A hipótese defendida é condicional. A partir de um dever jurídico, regulatório ou profissional de supervisão efetiva, propõe-se reconstruir uma posição funcional que preserve a capacidade de divergir de boa-fé. Sua qualificação como direito subjetivo dependerá da fonte aplicável, do vínculo, do titular do interesse tutelado e do sujeito contra quem a pretensão é dirigida. A expressão direito à divergência humana protegida será usada apenas quando essas condições estiverem presentes. Direito ao desalinhamento algorítmico permanece como rótulo autoral secundário, pois desalinhamento já possui significado próprio na literatura de segurança de IA.

O artigo não afirma a existência de um direito fundamental universal com esse nome. Também não presume superioridade humana, nem converte discordância em imunidade profissional. A proposta é mais estreita: reconstruir as consequências jurídicas necessárias para impedir que uma obrigação de supervisão seja neutralizada por incentivos internos incompatíveis com sua finalidade.

O método possui quatro movimentos. Primeiro, revisa-se a literatura que operacionalizou a supervisão humana. Segundo, interpretam-se o Regulamento Europeu de Inteligência Artificial, a LGPD e a

Resolução CNJ nº 615/2025. Terceiro, reconstrói-se a posição de divergência protegida. Por fim, apresenta-se uma matriz qualitativa de diagnóstico. O direito brasileiro funciona como ordenamento principal; o direito da União Europeia é empregado como comparação funcional; e, sem base positiva suficiente, a construção transetorial é apresentada como proposta normativa, não como direito vigente.

A revisão dirigida foi realizada mediante buscas no Google Scholar e no arXiv, seguidas de conferência nas páginas dos periódicos e em repositórios oficiais (EUR-Lex, Planalto, CNJ e EDPS). Usaram-se, em português e inglês, os descritores 'supervisão humana', 'human oversight', 'meaningful human control', 'automation bias', 'protected divergence', 'override' e 'non-retaliation'. Foram incluídos trabalhos que ofereciam conceito, condição de efetividade, evidência empírica ou consequência jurídica diretamente pertinente; comentários genéricos e fontes sem identificação verificável foram excluídos. O corpus foi encerrado em 19 de agosto de 2026. Não houve revisão sistemática, pesquisa de campo, experimento ou validação de instrumento. O protocolo torna explícitos os critérios de seleção, mas não autoriza alegação de exaustividade.

2. Estado da arte e delimitação da originalidade

A literatura de fatores humanos há décadas rejeita a ideia de que basta inserir uma pessoa no fluxo. Bainbridge (1983) mostrou que a automação pode reservar ao operador justamente a tarefa rara e difícil de detectar falhas; Parasuraman e Manzey (2010) sistematizaram riscos de complacência e viés no uso humano da automação. Em sentido complementar, os experimentos de Alon-Barkat e Busuioc (2023) no setor público não revelaram adesão algorítmica universal, o que impede presumir submissão em qualquer contexto.

A distinção entre presença humana e supervisão efetiva está consolidada o suficiente para impedir qualquer reivindicação ampla de novidade. Green (2022) mostrou que políticas de humano no circuito (*human-in-the-loop*) podem legitimar sistemas problemáticos sem corrigir suas falhas. Laux (2024) tratou a supervisão como desenho institucional de desconfiança organizada. O EDPS (2025), em documento técnico que expressamente não pretende oferecer interpretação jurídica, foi além da crítica à presença nominal: vinculou a agência do operador à ausência de medo de consequências implícitas ou explícitas quando ele substitui a saída automatizada.

Sterz et al. (2024) formularam quatro condições de efetividade: poder causal, acesso epistêmico, autocontrole e intenções adequadas. Langer, Lazar e Baum (2025) mostraram que a conformidade não se prova por lista de verificação (*checklist*) quando a efetividade depende do contexto e pode exigir avaliação empírica. Zhu et al. (2026) distinguiram agência operativa da IA e agência avaliativa humana, enfatizando verificação, direcionamento e substituição. Em julho de 2026, van de Sande, Economou-Zavlanos e van Genderen propuseram, para a IA médica, capacidade epistêmica, espaço cognitivo,

autoridade decisória e efetividade da intervenção; afirmaram expressamente que a divergência justificada deve ser esperada, documentada e protegida.

Fonte	Contribuição já existente	Questão residual
EDPS (2025)	A agência do operador depende de não temer consequências implícitas ou explícitas ao substituir a saída.	Documento técnico, sem interpretação jurídica nem regime de titularidade, causalidade, prova ou tutela.
Sterz et al. (2024)	Poder causal, acesso epistêmico, autocontrole e intenções adequadas.	Não estrutura uma posição jurídica de não retaliação, prova e tutela.
Langer; Lazar; Baum (2025)	Dificuldades de testar conformidade; limites de listas de verificação e necessidade de avaliação contextual.	Não oferece regime substantivo para proteger o supervisor divergente.
Zhu et al. (2026)	Agência avaliativa, verificação, contestação, direcionamento e substituição.	Concentra-se no desenho da supervisão, não em consequências adversas impostas ao avaliador.
Van de Sande et al. (2026)	Capacidade epistêmica, espaço cognitivo, autoridade, intervenção efetiva e proteção da divergência justificada.	A proteção é requisito de governança clínica; não se desenvolve um regime jurídico transetorial completo.
Contribuição deste artigo	Sistematiza a arquitetura jurídico-dogmática residual da divergência protegida, sem reivindicar a origem da ideia.	Níveis de proteção, sujeitos por controle, gatilho, causalidade, prova, limites e tutelas.

Fonte: elaboração do autor a partir das obras citadas.

Esse reconhecimento altera a pretensão de originalidade. A contribuição proposta não reside na descoberta de que a divergência precisa de proteção, já presente no EDPS (2025) e em van de Sande et al. (2026). Sua novidade residual é jurídico-dogmática: distinguir dever institucional, interesse funcionalmente protegido e eventual direito subjetivo; identificar o sujeito obrigado por esfera de controle; e formular critérios de causalidade, prova, limites e tutela. A simetria de escrutínio aparece como padrão de governança proposto e só será exigível quando incorporada pela fonte jurídica aplicável.

3. O mapa normativo: deveres existentes e lacuna residual

3.1. Regulamento Europeu de Inteligência Artificial

O artigo 14 do Regulamento (UE) 2024/1689 disciplina a supervisão humana dos sistemas de IA de alto risco. O sistema deve ser concebido para que as pessoas encarregadas compreendam capacidades e limitações relevantes, percebam o risco de confiança excessiva, interpretem a saída e possam decidir não utilizar o sistema, ignorar, substituir ou reverter o resultado, além de interromper a operação quando necessário. O artigo 26, nº 2, dirige-se ao responsável pela implantação (*deployer*): a supervisão

deve ser confiada a pessoas singulares com competência, formação, autoridade e apoio necessários (UNIÃO EUROPEIA, 2024).

A repartição importa. O fornecedor controla desenho, documentação e mecanismos técnicos de intervenção. O responsável pela implantação controla o uso conforme instruções, a designação dos supervisores e o suporte operacional. O empregador, o órgão público ou outro gestor do vínculo controla metas, avaliação, remuneração, disciplina e distribuição de tarefas. Esses papéis podem coincidir, mas a coincidência não deve ser presumida; cada dever acompanha a esfera efetiva de controle. Laux e Ruschemeier (2025) observam que a determinação para reconhecer o viés de automação não elimina suas causas organizacionais. Os artigos 14 e 26 criam condições de intervenção, mas não contêm regime geral de proteção trabalhista ou profissional contra retaliações.

3.2. LGPD e Resolução CNJ nº 615/2025

O artigo 20 da LGPD atribui ao titular o direito de solicitar revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses e impõe ao controlador deveres informacionais sobre critérios e procedimentos, com as ressalvas legais (BRASIL, 2018). O titular afetado, porém, ocupa posição diferente daquela do profissional encarregado de supervisionar ou decidir. O artigo 20 não deve ser apresentado como fonte direta de um direito geral do supervisor.

No Poder Judiciário, a Resolução CNJ nº 615/2025, em texto compilado após a Resolução nº 674/2026, oferece base normativa mais próxima. O artigo 2º, V, inclui participação e supervisão humana; o artigo 3º, VII e VIII, exige supervisão efetiva, periódica e adequada ao risco, além de capacitação contínua. O artigo 10, I, veda soluções que impeçam revisão ou produzam dependência absoluta. O artigo 32 assegura autonomia aos usuários internos, revisão detalhada, acesso a premissas e método, ausência de vinculação à saída e autoridade final. O artigo 34 exige supervisão humana e permite ao magistrado competente modificar o produto gerado por IA (CNJ, 2025).

O conjunto é robusto quanto a acesso, não vinculação e autoridade. Ainda assim, não disciplina expressamente como metas de produtividade, remuneração, avaliação funcional ou auditorias seletivas devem tratar a divergência. Também não define um regime específico de causalidade, prova e tutela para a desvantagem imposta a quem contrariou a saída algorítmica. É nessa faixa residual que a proposta se localiza.

Posição ou esfera	Sujeito central	Função
Revisão da decisão automatizada	Pessoa afetada pelo resultado.	Provocar nova apreciação nos limites da norma aplicável.

Posição ou esfera	Sujeito central	Função
Desenho da supervisão	Fornecedor.	Prover documentação, informação e mecanismos de intervenção no limite do controle sobre o sistema.
Implantação e operação	Responsável pela implantação.	Designar pessoas competentes e assegurar autoridade, formação e apoio operacional exigidos pelo regime.
Relação funcional	Empregador, órgão público ou gestor do vínculo.	Controlar metas, avaliação e medidas adversas sem neutralizar a supervisão exigida.
Divergência humana protegida	Pessoa natural investida na revisão, supervisão ou decisão.	Exercer discordância fundamentada nos limites da posição reconhecida pela fonte aplicável.

Fonte: elaboração do autor. As posições e esferas podem coexistir no mesmo processo; a responsabilidade de cada ator permanece limitada à sua fonte jurídica e ao controle que efetivamente exerce.

3.3. Da obrigação institucional à posição funcional protegida

Uma obrigação de supervisão não gera, por si só, direito subjetivo do supervisor. Em linguagem analítica inspirada em Hohfeld (1913), a existência de um dever só permite afirmar uma pretensão correlata depois de identificar a quem a conduta é devida. No Regulamento Europeu de Inteligência Artificial, por exemplo, os deveres de supervisão controlam riscos e protegem pessoas afetadas; disso não decorre automaticamente legitimidade do operador para exigir tutela em nome próprio.

A reconstrução proposta distingue três níveis. No primeiro, há dever institucional objetivo de organizar a supervisão. No segundo, há interesse funcionalmente protegido do supervisor, porque sua independência pode ser instrumento necessário à finalidade da norma. No terceiro, haverá direito subjetivo exigível pelo supervisor apenas quando lei, estatuto, contrato, regulamento interno vinculante, boa-fé objetiva, proteção da confiança ou regime administrativo lhe atribuir pretensão contra ator determinado.

Quatro perguntas controlam essa passagem: a fonte exige supervisão efetiva? A independência do supervisor integra o seu escopo de proteção? Qual ator controla a condição de trabalho ou a medida adversa discutida? O regime confere legitimidade e tutela ao supervisor? Se as duas primeiras respostas forem positivas, mas as demais permanecerem indeterminadas, o argumento sustenta interpretação finalística ou proposta normativa; não autoriza proclamar um direito subjetivo já consolidado.

Posição funcional protegida é, portanto, o gênero empregado neste artigo. Direito à divergência humana protegida é a espécie exigível quando a correlação entre titular, sujeito obrigado, conduta e tutela puder ser demonstrada. Nos demais casos, a matriz adiante opera como parâmetro de governança e agenda regulatória.

4. Regime jurídico da divergência humana protegida

4.1. Natureza e definição

Neste artigo, divergência humana protegida designa o gênero posição funcional protegida. Sua incidência pressupõe que uma pessoa natural tenha atribuição real de revisar, supervisionar ou decidir sobre saída de IA em contexto no qual a supervisão efetiva seja exigida pelo ordenamento, por norma profissional vinculante ou por compromisso institucional juridicamente vinculante, conforme contrato, regulamento interno, boa-fé objetiva, proteção da confiança ou regime administrativo aplicável. Não se propõe um direito universal para qualquer uso cotidiano de IA.

> ***Quando reunidas as condições de exigibilidade, o direito à divergência humana protegida assegura ao supervisor competente o exercício, de boa-fé e com fundamento profissional, dos poderes de questionar, suspender, desconsiderar, corrigir, substituir ou encaminhar uma saída de IA, sem desvantagem fundada exclusiva ou substancialmente nessa conduta, observadas as regras causais e probatórias do regime aplicável.***

A definição não protege o resultado humano por ser humano. Protege o processo de julgamento independente que a própria supervisão pressupõe. A organização conserva o dever e a prerrogativa de controlar qualidade, coerência, legalidade e resultados. O limite está em usar a simples divergência como atalho para presumir erro, deslealdade ou baixa produtividade.

Elemento	Proposição mínima
Titular	Pessoa natural formal ou funcionalmente incumbida de revisar, supervisionar, decidir ou interromper uma trajetória mediada por IA.
Sujeito organizacional	Empregador, órgão público ou gestor do vínculo, quando a fonte o vincular e no limite do controle sobre metas, avaliação, remuneração, disciplina e tarefas.
Sujeito operacional	Responsável pela implantação, quando a fonte o vincular e no limite do controle sobre designação, formação, autoridade, suporte e condições de uso.
Sujeito técnico complementar	Fornecedor ou desenvolvedor, no limite do controle sobre documentação, informação operacional e mecanismos de intervenção; não se presume dever de não retaliação.
Fato gerador	Uso materialmente relevante de saída de IA em processo que deva conservar supervisão humana efetiva segundo fonte jurídica identificável.
Conduta protegida	Discordância de boa-fé, apoiada em razão profissional identificável e exercida por canal seguro e efetivo, ressalvadas urgência ou impossibilidade justificada.
Conteúdo positivo	Acesso útil, tempo proporcional ao risco, suporte, autoridade, registro e preservação de evidências, distribuídos conforme a esfera de controle de cada ator.
Conteúdo negativo	Vedação, quando exigível, de avaliação, perda remuneratória, bloqueio de promoção, distribuição punitiva de tarefas ou sanção em que a divergência seja causa exclusiva ou fator substancial.

Elemento	Proposição mínima
Causalidade	A medida não teria ocorrido, ou teria sido menos onerosa, sem a divergência; em motivação mista, admite-se prova de razão legítima independente e suficiente, conforme o regime aplicável.
Limites	Não abrange fraude, discriminação, desídia, recusa arbitrária, sabotagem, erro grosseiro ou descumprimento injustificado de protocolo legítimo.
Tutelas possíveis	Conforme base trabalhista, administrativa, civil ou processual: cessação ou revisão da medida, reavaliação independente, correção de métricas, preservação e exibição de registros e reparação quando presentes seus requisitos.

Fonte: elaboração do autor. A incidência concreta depende do regime setorial, funcional, contratual e processual aplicável.

4.2. Não retaliação não significa neutralidade absoluta

Para evitar a abertura da expressão penalização assimétrica, adota-se um critério causal e contrafactual. A proteção incide quando a divergência fundamentada é causa exclusiva ou fator substancial sem o qual a medida adversa não teria sido aplicada, ou teria sido menos onerosa. Motivos mistos não afastam automaticamente a proteção. A organização pode demonstrar razão legítima independente e suficiente, baseada em critérios prévios e aplicados consistentemente a casos comparáveis. O padrão de prova e o efeito dessa demonstração dependem do regime jurídico aplicável.

Também não se exige auditoria numericamente idêntica de concordâncias e discordâncias. Equivalência de escrutínio significa aplicar critérios de diligência comparáveis a casos funcionalmente semelhantes: mesma função, tipo de decisão, faixa de risco, informação disponível, impacto e período de avaliação. Intensidade distinta é legítima quando justificada, antes do resultado, por risco, incerteza ou anomalia. Trata-se de padrão de governança proposto; sua exigibilidade depende de incorporação normativa, contratual ou administrativa.

4.3. Prova, registros e causalidade

O supervisor costuma ter acesso limitado às métricas, aos registros automatizados (*logs*), às comparações de desempenho e às comunicações gerenciais que explicam uma medida adversa. Por isso, a proteção precisa de desenho probatório. Uma alegação responsável deve apresentar lastro mínimo: existência e fundamento da divergência, medida posterior e indícios de conexão, como proximidade temporal, mudança de critério, fala gerencial, auditoria seletiva ou diferença em relação a comparadores. Nenhum desses elementos, isoladamente, encerra a controvérsia.

Quando a organização detém os registros decisivos, a distribuição dinâmica do ônus da prova prevista no artigo 373, § 1°, do Código de Processo Civil pode servir como técnica processual, desde que o juiz reconheça impossibilidade, dificuldade excessiva ou maior facilidade da parte contrária e assegure oportunidade de desincumbimento (BRASIL, 2015). Não se trata de inversão automática. Antes do litígio,

a conservação proporcional de registros reduz a assimetria quando o regime setorial impuser documentação ou quando a organização apresentar a supervisão como salvaguarda institucional demonstrável. Retenção, acesso e proteção de dados continuam submetidos às normas aplicáveis.

4.4. Direito de divergir e dever de intervir

A posição protetiva não exclui deveres profissionais. Em situações definidas por lei, norma profissional ou protocolo legítimo, o supervisor não apenas pode divergir: deve intervir diante de risco conhecido, ilegalidade, dado incompatível ou saída fora do uso pretendido. O conteúdo desse dever depende da profissão e do setor. A não retaliação permite que ele seja cumprido; não substitui o padrão de diligência nem transfere toda responsabilidade à organização.

5. Matriz Diagnóstica Preliminar de Divergência Protegida

A denominação teste sugeriria validade e confiabilidade ainda não demonstradas. Por isso, o instrumento é apresentado como Matriz Diagnóstica Preliminar. Ele organiza evidências de governança; não mede personalidade, não produz nota de conformidade e não autoriza, sozinho, conclusão jurídica.

As seis dimensões possuem base identificável. Acesso epistêmico e poder causal dialogam com Sterz et al. (2024); espaço cognitivo e autoridade, com van de Sande et al. (2026); agência avaliativa e verificabilidade, com Zhu et al. (2026); e a cautela contra listas de verificação, com Langer, Lazar e Baum (2025). EDPS (2025) e van de Sande et al. (2026) já sustentam a necessidade de proteger a intervenção ou divergência. O desenvolvimento específico deste artigo é a arquitetura jurídica residual: níveis de proteção, alocação por controle, causalidade, prova e tutelas.

Dimensão	Pergunta de controle	Evidência observável e sinal de falha
1. Acesso epistêmico	O supervisor recebe finalidade, limites, dados do caso, incerteza e base suficiente para verificar a saída?	Documentação, formação, telas, fontes e amostras de casos. Falha: conclusão sem base operacionalmente verificável.
2. Espaço cognitivo	Há tempo, equipe e recursos proporcionais ao risco e ao custo de verificação?	Carga, filas, níveis de serviço, tempo real e comparação por complexidade. Falha: revisar torna impossível cumprir a meta ordinária.
3. Autoridade e intervenção	A pessoa pode ignorar, corrigir, suspender, substituir ou escalar antes que o efeito seja irreversível?	Permissões, fluxos, testes de intervenção e prazos. Falha: intervenção excepcional, tardia ou apenas simbólica.
4. Divergência protegida	A fonte aplicável protege a discordância fundamentada contra desvantagem causada pela divergência em si?	Lei, contrato, estatuto, política vinculante, avaliações e sanções. Falha: divergir aciona automaticamente perda, suspeita ou exposição.
5. Escrutínio equivalente	Casos funcionalmente semelhantes são examinados por critérios comparáveis, com diferenças justificadas por risco?	Plano amostral, função, tipo de decisão, faixa de risco, período e resultados. Falha: somente o divergente precisa demonstrar diligência.

Dimensão	Pergunta de controle	Evidência observável e sinal de falha
6. Rastreabilidade probatória	Quando houver dever de documentação, é possível reconstruir saída, informação disponível, decisão e consequência?	Registros, versões, retenção, acesso e proteção de dados. Falha: atribui-se responsabilidade sem preservar a cadeia exigível.

Fonte: elaboração do autor a partir da literatura e das normas citadas. Instrumento conceitual não validado; as dimensões de proteção, escrutínio e rastreabilidade funcionam como proposta de governança quando não houver fonte jurídica que as torne exigíveis.

5.1. Protocolo mínimo de aplicação

1. Delimitar a decisão, o uso pretendido da IA, o risco e o sujeito que realmente possui autoridade.
2. Reunir políticas, interfaces, permissões, formação, metas, registros, critérios de auditoria e amostras de casos.
3. Ouvir funções diferentes — supervisor, gestor, equipe técnica, qualidade e pessoas afetadas — quando a investigação permitir.
4. Classificar cada dimensão como evidenciada, parcialmente evidenciada ou não demonstrada, registrando a base da conclusão.
5. Tratar a ausência de evidência como supervisão não demonstrada somente quando houver dever de documentação ou escopo de auditoria previamente definido; não convertê-la automaticamente em ilícito ou falha comprovada.
6. Validar a matriz por estudos de caso, confiabilidade entre avaliadores e comparação com desfechos antes de qualquer pontuação.

As dimensões não devem ser somadas mecanicamente. Uma deficiência pode ser crítica conforme o risco: autoridade inexistente não é compensada por formação excelente. Ao mesmo tempo, a matriz precisa distinguir falha do sistema, inadequação do fluxo e conduta individual. A conclusão deve ser contextual e revisável.

6. Métricas, incentivos e condicionamento institucional da decisão

Medidas institucionais podem alterar o comportamento que pretendem descrever. Espeland e Sauder (2007) chamam esse fenômeno de reatividade: pessoas e organizações ajustam práticas em resposta à avaliação. No trabalho mediado por algoritmos, Kellogg, Valentine e Christin (2020) mostram que sistemas podem direcionar, registrar, classificar, recompensar e restringir condutas. Essas pesquisas não provam que toda meta de produtividade suprima a supervisão; justificam tratá-la como hipótese de risco auditável.

A arquitetura problemática aparece quando o custo da discordância é incorporado à avaliação pessoal, enquanto o custo da concordância errada se dissolve no sistema. A decisão continua formalmente

atribuída a uma pessoa, mas a estrutura de incentivos condiciona a resposta esperada. Condicionamento institucional, aqui, significa reduzir indiretamente o espaço de julgamento por meio da combinação entre interface, tempo, metas, hierarquia e distribuição assimétrica de responsabilidade, mesmo sem ordem explícita para seguir a IA.

Quatro cautelas reduzem o risco sem eliminar gestão de desempenho:

- não transformar taxa de divergência em meta de adesão ou resistência;
- incorporar complexidade e tempo de revisão às métricas de produtividade;
- auditar amostras de concordâncias e discordâncias por critérios de risco e qualidade previamente definidos;
- separar o registro de intervenção, rejeição ou substituição da saída, usado como sinal de segurança e aprendizagem, da presunção disciplinar de não conformidade.

Taxas altas e baixas admitem explicações concorrentes. Poucas divergências podem indicar excelente desempenho do modelo, casos homogêneos, falta de tempo ou receio hierárquico. Muitas podem revelar modelo inadequado, seleção de casos difíceis, independência profissional ou rejeição humana infundada. O indicador só ganha sentido quando relacionado ao contexto, à justificativa e ao desfecho.

7. Duas aplicações delimitadas

7.1. Poder Judiciário

No Judiciário brasileiro, a Resolução CNJ nº 615/2025 permite uma aplicação dogmaticamente controlada. Para magistrados, os artigos 32 e 34 preservam ausência de vinculação e autoridade de modificação. Para servidores e colaboradores, a posição dependerá da tarefa: quem não possui autoridade final precisa ao menos de canal de escalonamento capaz de interromper ou corrigir o fluxo no limite de sua atribuição. Não é correto equiparar todos os usuários internos ao magistrado competente.

A auditoria deveria verificar se a não vinculação declarada é confirmada pelas permissões do sistema, pelo tempo disponível, pelos critérios de produtividade e pelo tratamento dado aos alertas internos. A simples assinatura do magistrado não prova revisão substancial; por outro lado, o uso de IA também não permite inferir ausência de julgamento. A auditoria deve buscar evidência proporcional ao risco de que a autoridade permaneceu efetivamente exercitável.

7.2. Saúde como comparação funcional

A saúde oferece comparação funcional, não demonstração automática do direito brasileiro proposto. Van de Sande et al. (2026) descrevem supervisão médica baseada em capacidade epistêmica, espaço

cognitivo, autoridade decisória e intervenção efetiva, inclusive com proteção da divergência justificada e interpretação cautelosa das taxas de intervenção ou substituição da saída. O paralelo confirma a relevância do problema organizacional, mas qualquer consequência jurídica concreta exige análise das normas sanitárias, profissionais, trabalhistas e de responsabilidade aplicáveis ao caso.

Numa ferramenta de alerta clínico, por exemplo, registrar a discordância pode melhorar segurança e calibração. Esse registro não deve funcionar automaticamente como prova de erro do profissional. Também não deve blindá-lo quando ignora alerta correto sem base adequada. A governança precisa reconstruir o contexto, as informações disponíveis e os resultados, evitando tanto deferência cega ao modelo quanto privilégio abstrato do julgamento humano.

8. Objeções e respostas

8.1. A IA pode estar correta e o humano, errado

A objeção procede, mas não derruba a tese. A proteção recai sobre a independência do processo, não sobre a correção presumida do resultado humano. Divergências permanecem sujeitas a revisão de qualidade, e concordâncias também. O padrão relevante é diligência comparável, não preferência ontológica.

8.2. A proteção pode encobrir negligência

O risco existe se a categoria for mal definida. Por isso, exigem-se boa-fé, fundamento profissional identificável e atuação dentro das competências e dos canais disponíveis, ressalvadas urgências justificadas. Fraude, discriminação, sabotagem, recusa arbitrária e erro grosseiro continuam fora da proteção.

8.3. A supervisão real tem custo

Tem. A resposta adequada é proporcionalidade. Processos de baixo impacto podem usar amostragem e controles leves; decisões de alto impacto exigem tempo e poderes maiores. O custo não desaparece quando a organização o transfere silenciosamente ao supervisor. Se a participação humana é usada como garantia de segurança ou legitimidade, seus recursos precisam integrar o desenho da atividade.

8.4. O artigo 14 do Regulamento Europeu de Inteligência Artificial já contém poderes de intervenção

Contém, e por isso é fundamento central, não adversário da proposta. A contribuição residual examina o que ocorre quando o poder técnico existe, mas seu exercício produz consequência adversa na relação

organizacional. O artigo 26 aproxima-se do problema ao exigir autoridade e suporte; ainda assim, não oferece, por si só, um regime completo de não retaliação, causalidade e tutela profissional.

8.5. Não existe um direito geral consolidado

Correto. A tese é reconstrutiva e condicional. Sua força será maior nos setores em que supervisão efetiva, autonomia profissional ou independência decisória já estejam positivadas. Onde tais premissas não existirem, a matriz serve como proposta regulatória ou de governança, não como declaração automática de direito subjetivo.

9. Limites e agenda de pesquisa

A proposta ainda precisa ser testada em ambientes reais. Estudos de caso devem comparar políticas declaradas, interfaces, carga de trabalho, decisões, medidas de desempenho e resultados. Pesquisas futuras podem investigar se certos desenhos geram dependência funcional ou erosão de expertise, mas este artigo não demonstra tais efeitos. Conceitos mais amplos — risco arquitetônico, dever de cuidado cognitivo e sustentabilidade cognitiva — permanecem em agenda própria e não são pressupostos da tese aqui defendida.

A prioridade empírica é mais modesta: verificar se a divergência fundamentada altera avaliações, remuneração, promoção, exposição disciplinar ou distribuição de tarefas; comparar esse tratamento com casos de concordância; e identificar quem controla as evidências. Somente depois será possível estimar prevalência, causalidade e efetividade das medidas propostas.

10. Conclusão

A supervisão humana pode fracassar mesmo quando a interface contém um botão de rejeição. Informação insuficiente, falta de tempo, hierarquia, métricas e risco profissional podem esvaziar a autoridade formal. A literatura recente já descreve boa parte dessas condições.

Sua contribuição é jurídica e delimitada. Nos contextos em que a supervisão efetiva é exigida, propõe-se reconstruir uma posição funcional que proteja a divergência fundamentada. Ela será direito subjetivo do supervisor somente quando fonte, vínculo, escopo de proteção, sujeito obrigado e tutela puderem ser identificados. A arquitetura formulada aloca deveres conforme o controle exercido, exige fundamento e boa-fé, admite controle de qualidade e oferece critérios para causalidade, prova e remédios conforme o regime aplicável.

A Matriz Diagnóstica Preliminar traduz a tese em perguntas verificáveis, sem simular validação inexistente. Ela desloca a análise da presença nominal do humano para as condições que tornam sua intervenção possível e demonstrável. O resultado mais importante é uma regra de coerência institucional: quem invoca supervisão humana para legitimar uma decisão deve preservar e documentar as condições que tornavam possível a divergência, sem prejuízo das regras de distribuição do ônus da prova aplicáveis ao caso.

> ***Se contrariar a IA é formalmente permitido, mas profissionalmente perigoso, a supervisão continua humana apenas no organograma.***

Referências

ALON-BARKAT, Saar; BUSUIOC, Madalina. Human–AI interactions in public sector decision making: 'automation bias' and 'selective adherence' to algorithmic advice. Journal of Public Administration Research and Theory, v. 33, n. 1, p. 153–169, 2023. DOI: https://doi.org/10.1093/jopart/muac007.

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. Lei n° 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: 19 ago. 2026.

BRASIL. Lei n° 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/l13709.htm. Acesso em: 19 ago. 2026.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). Resolução n° 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. Texto compilado, com alteração pela Resolução n° 674, de 25 de março de 2026. Brasília, DF: CNJ, 2025. Disponível em: https://atos.cnj.jus.br/atos/detalhar/6001. Acesso em: 19 ago. 2026.

ESPELAND, Wendy Nelson; SAUDER, Michael. Rankings and reactivity: how public measures recreate social worlds. American Journal of Sociology, v. 113, n. 1, p. 1–40, 2007. DOI: https://doi.org/10.1086/517897.

EUROPEAN DATA PROTECTION SUPERVISOR (EDPS). TechDispatch #2/2025: human oversight of automated decision-making. Brussels: EDPS, 2025. Disponível em: https://www.edps.europa.eu/system/files/2025-09/25-09-15_techdispatch-human-oversight_en.pdf. Acesso em: 19 ago. 2026.

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.

HOHFELD, Wesley Newcomb. Some fundamental legal conceptions as applied in judicial reasoning. The Yale Law Journal, v. 23, n. 1, p. 16–59, 1913. DOI: https://doi.org/10.2307/785533.

KELLOGG, Katherine C.; VALENTINE, Melissa A.; CHRISTIN, Angèle. Algorithms at work: the new contested terrain of control. Academy of Management Annals, v. 14, n. 1, p. 366–410, 2020. DOI: https://doi.org/10.5465/annals.2018.0174.

LANGER, Markus; LAZAR, Veronika; BAUM, Kevin. On the complexities of testing for compliance with human oversight requirements in AI regulation. arXiv:2504.03300, versão 2, 2025. Disponível em: https://arxiv.org/abs/2504.03300. Acesso em: 19 ago. 2026.

LAUX, Johann. Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act. AI & Society, v. 39, n. 6, p. 2853–2866, 2024. DOI: https://doi.org/10.1007/s00146-023-01777-z.

LAUX, Johann; RUSCHEMEIER, Hannah. Automation bias in the AI Act: on the legal implications of attempting to de-bias human oversight of AI. European Journal of Risk Regulation, v. 16, n. 4, p. 1519–1534, 2025. DOI: https://doi.org/10.1017/err.2025.10033.

PARASURAMAN, Raja; MANZEY, Dietrich H. Complacency and bias in human use of automation: an attentional integration. Human Factors, v. 52, n. 3, p. 381–410, 2010. DOI: https://doi.org/10.1177/0018720810376055.

STERZ, Sarah; BAUM, Kevin; BIEWER, Sebastian; HERMANNS, Holger; LAUBER-RÖNSBERG, Anne; MEINEL, Philip; LANGER, Markus. On the quest for effectiveness in human oversight: interdisciplinary perspectives. In: ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (FAccT), 2024, Rio de Janeiro. Proceedings [...]. New York: ACM, 2024. p. 2495–2507. DOI: https://doi.org/10.1145/3630106.3659051.

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, 2024. Versão consolidada em 27 jul. 2026. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:02024R1689-20260727. Acesso em: 19 ago. 2026.

VAN DE SANDE, Davy; ECONOMOU-ZAVLANOS, Nicoleta; VAN GENDEREN, Michel E. Meaningful oversight of medical AI beyond human in the loop. npj Digital Medicine, v. 9, art. 569, 2026. DOI: https://doi.org/10.1038/s41746-026-02971-1.

ZHU, Liming; LU, Qinghua; DING, Ming; LEE, Sung Une; WANG, Chen. Designing meaningful human oversight in AI. AI and Ethics, v. 6, art. 286, 2026. DOI: https://doi.org/10.1007/s43681-026-01147-7.

JANDISLEI ANTONIO GENOVA

Assinado de forma digital por JANDISLEI ANTONIO GENOVA
Dados: 2026.08.19 12:17:31 -03'00'