

Quem paga por contrariar a inteligência artificial?

A assimetria institucional dos custos da divergência algorítmica

Da divergência protegida à distribuição jurídica dos custos da supervisão

JANDISLEI ANTONIO GENOVA

Advogado e pesquisador independente | São Paulo, agosto de 2026

Quando a organização exige supervisão humana, mas concentra no profissional o custo de divergir e dispersa o custo de aderir ao erro automatizado, ela transforma a salvaguarda em incentivo à conformidade.

Resumo

O artigo propõe o conceito de assimetria institucional dos custos da divergência. A categoria descreve a configuração em que contrariar uma saída de inteligência artificial gera ônus profissional concentrado, imediato e atribuível, enquanto aderir a uma saída errada produz ônus tardio, difuso ou externalizado. Mediante análise jurídico-dogmática, comparação funcional e revisão narrativa dirigida, delimitam-se um teste de relevância jurídica, uma regra de coerência alocativa e uma exigência de escrutínio equivalente. O Regulamento Europeu de Inteligência Artificial, a Diretiva (UE) 2024/2831, a Resolução CNJ nº 615/2025 e categorias do direito brasileiro são examinados sem se postular direito geral já consolidado. A matriz diagnóstica final organiza indícios sobre a distribuição temporal e institucional desses custos e sobre as condições de sua contestação. A tese central é que uma organização não pode invocar supervisão como salvaguarda e tornar profissionalmente mais seguro obedecer à máquina do que exercer o julgamento destinado a controlá-la.

Palavras-chave: inteligência artificial; supervisão humana; divergência algorítmica; alocação de riscos; métricas institucionais; responsabilidade profissional.

Abstract

This article proposes the concept of institutional asymmetry in the costs of divergence. The concept describes a setting in which challenging an artificial intelligence output creates concentrated, immediate, and attributable professional costs, while following an erroneous output produces delayed, diffuse, or externalised costs. Through doctrinal legal analysis, functional comparison, and a targeted narrative review, the article develops a legal relevance test, a rule of allocative coherence, and a requirement of equivalent scrutiny. The EU Artificial Intelligence Act, Directive (EU) 2024/2831, Brazilian National Council of Justice Resolution No. 615/2025, and selected categories of Brazilian law are examined without claiming that a general right is already established. A preliminary diagnostic matrix organises evidence about who bears the costs, when they arise, and how they may be challenged. The central claim is that an organisation cannot invoke oversight as a safeguard while making obedience to the machine professionally safer than the judgment meant to control it.

Keywords: artificial intelligence; human oversight; algorithmic divergence; risk allocation; institutional metrics; professional responsibility.

1. Introdução: o erro que tem dono e o erro que se dissolve

Contrariar uma recomendação algorítmica deixa rastros claros. Consome tempo, exige fundamentação, afeta a fila de trabalho e pode elevar a taxa individual de exceções. Se a decisão humana estiver errada, há um nome, um registro e um instante facilmente identificáveis. Concordar com a recomendação, ao contrário, costuma caber no fluxo esperado. Quando a máquina erra e o profissional apenas adere, a responsabilidade pode se dispersar entre dados, fornecedor, configuração, gestão, política institucional e volume operacional. O custo surge mais tarde — às vezes para a pessoa afetada, às vezes para a organização, às vezes para a coletividade — e raramente é lançado na mesma métrica que premiou a adesão.

Essa diferença não é um detalhe psicológico. É uma arquitetura de incentivos. Mesmo quando o sistema possui botão de rejeição, o profissional pode aprender que utilizá-lo é caro. Revisar reduz produtividade aparente; documentar a exceção cria exposição; escalar a dúvida gera conflito; discordar repetidamente ameaça avaliação, remuneração, reputação ou carreira. A supervisão permanece no organograma, mas a racionalidade local recomenda conformidade.

Contrariar a IA produz um custo individual e imediatamente mensurável; concordar com uma IA errada frequentemente produz um custo tardio, difuso e institucionalmente invisível.

No título, ‘pagar’ é empregado em sentido jurídico-institucional, e não como afirmação de incidência econômica estrita. A pergunta abrange quem suporta tempo, carga, exposição probatória, efeitos métricos, perdas relacionais, consequências disciplinares ou danos de carreira. Esses ônus são heterogêneos e não se convertem automaticamente em dinheiro nem se somam numa unidade comum.

O artigo anterior desta agenda reconstruiu a divergência humana protegida como posição funcional condicional nos contextos em que a supervisão efetiva é exigida (GENOVA, 2026). O passo seguinte é perguntar quem paga para que essa divergência exista. Uma autorização formal de intervir é insuficiente quando seus custos de transação, prova e carreira são individualizados, enquanto os custos da adesão indevida são socializados. A questão deixa de ser apenas se o humano pode contrariar a IA e passa a ser como a instituição distribui os custos das duas rotas decisórias.

A hipótese é delimitada. Nem toda diferença de custo é ilícita; nem toda divergência merece proteção; nem toda adesão automatizada representa falha. A relevância jurídica pressupõe fonte identificável de supervisão, autonomia ou diligência e controle organizacional sobre métricas, fluxos ou evidências. Exige, ainda, divergência de boa-fé apoiada em razão profissional, desvantagem causada pela divergência e ausência de razão legítima, independente e suficiente para o tratamento diferenciado.

A contribuição proposta possui três partes. Primeiro, nomeia e decompõe a assimetria dos custos. Segundo, formula uma regra de coerência alocativa: quem invoca a supervisão como salvaguarda deve internalizar os custos organizacionais previsíveis necessários ao seu exercício e não pode transferi-los integralmente ao supervisor. Terceiro, converte a tese em uma matriz diagnóstica preliminar, sem atribuir-lhe validade empírica ainda inexistente.

2. Método, estado da arte e originalidade residual

O trabalho adota análise jurídico-dogmática e comparação funcional. Parte de normas que atribuem finalidade à supervisão humana, identifica os atores que controlam suas condições de exercício e reconstrói consequências compatíveis com boa-fé, vedação do abuso, distribuição de riscos e efetividade normativa. A comparação europeia não importa automaticamente soluções para o direito brasileiro. Ela demonstra escolhas regulatórias possíveis e testa a coerência do argumento.

A revisão narrativa dirigida foi realizada em fontes oficiais — EUR-Lex, Planalto, Conselho Nacional de Justiça, Câmara dos Deputados, NIST e EDPS — e em páginas de periódicos, editoras e identificadores DOI. O recorte temporal encerrou-se em 22 de agosto de 2026. As buscas combinaram, em português e inglês, os termos ‘supervisão humana’, ‘human oversight’, ‘algorithmic management’, ‘automation bias’, ‘employee voice’, ‘organizational silence’, ‘just culture’, ‘retaliation’, ‘chilling effect’, ‘moral crumple zone’ e ‘metric reactivity’.

Foram incluídas normas e obras diretamente relacionadas a autoridade, suporte, controle, incentivos, evidência ou consequências adversas no exercício da supervisão. Excluíram-se textos de ética geral da IA sem relação demonstrável com a supervisão em práticas decisórias. O corpus final compreende 29 referências. A estratégia é transparente, mas não sistemática nem exaustiva: não houve metanálise, pesquisa de campo ou levantamento jurisprudencial completo. As conclusões são normativas e conceituais.

A literatura já demonstra que a presença de uma pessoa no fluxo não garante controle. Bainbridge (1983) descreveu as ironias da automação. Parasuraman e Manzey (2010) sistematizaram complacência e viés de automação. Alon-Barkat e Busuioc (2023) não encontraram adesão automática geral, mas identificaram padrões de adesão seletiva. Green (2022) mostrou que políticas de supervisão podem legitimar sistemas defeituosos. Sterz et al. (2024) organizaram condições de efetividade; Langer, Lazar e Baum (2025) expuseram a complexidade de testar conformidade; Laux (2024) destacou incentivos inadequados e desenho institucional. O EDPS (2025) afirmou que a agência do operador depende de não temer consequências por substituir a saída.

Também não é nova a percepção de que instituições condicionam a manifestação de dúvidas. Morrison e Milliken (2000) descrevem o silêncio organizacional como retenção coletiva de informações relevantes. Detert e Edmondson (2011) identificam crenças que tornam arriscado falar. A segurança psicológica explica condições interpessoais de aprendizagem (EDMONDSON, 1999), enquanto a just culture investiga confiança, aprendizagem e responsabilização após incidentes (DEKKER, 2016). Esses conceitos explicam por que a pessoa cala ou como a organização reage ao erro; não comparam, por si, a distribuição dos ônus entre divergir da IA e aderir a ela.

Espeland e Sauder (2007) chamam de reatividade a transformação de práticas por rankings e indicadores. Kellogg, Valentine e Christin (2020) mostram que algoritmos podem dirigir, registrar, classificar, recompensar e substituir trabalhadores. Elish (2019), com a metáfora da zona de deformação moral, demonstra como o humano próximo ao sistema pode absorver responsabilidade apesar de controle limitado. Calabresi (1970), em outro domínio, oferece a pergunta que orienta esta investigação: quais custos são reconhecidos, por quem são suportados e quais incentivos sua alocação produz?

A originalidade, portanto, é residual e combinatória. Não consiste em descobrir o medo de contrariar, a força das métricas ou a dispersão da responsabilidade. Consiste em comparar os custos da divergência e da adesão errada nas dimensões de incidência, tempo, visibilidade, atribuição, reversibilidade e prova. A comparação é vinculada a uma fonte jurídica de supervisão e conduz a um teste não circular de relevância, uma regra alocativa e controles verificáveis.

Conceito adjacente	Núcleo explicativo	Diferença específica da tese
Retaliação e efeito inibidor	Explicam sanções ou desestímulos que seguem uma manifestação protegida.	A tese inclui custos anteriores à sanção e compara seu tratamento com o da adesão errada.
Voz e silêncio organizacional	Explicam falar, calar e reter informações diante da hierarquia.	O objeto aqui é uma intervenção funcional sobre saída de IA, ligada a fonte jurídica e a custos comparados.
Segurança psicológica e just culture	Tratam confiança, aprendizagem e responsabilização diante de risco e erro.	A tese examina quem controla e suporta ônus de supervisão, inclusive métricos, probatórios e de carreira.
Zona de deformação moral	Revela culpa concentrada no humano próximo ao sistema apesar de controle limitado.	A assimetria abrange também o custo de prevenir o erro e o contraste entre divergir e aderir.
Gestão algorítmica e reatividade	Expõem direção por sistemas, métricas e efeitos comportamentais dos indicadores.	A tese acrescenta gatilho jurídico, causalidade, defesa legítima e escrutínio comparativo.
Contribuição residual	Integra supervisão, custos e prova numa categoria jurídico-institucional.	Compara as duas rotas no tempo e na organização; ainda requer validação empírica e calibração setorial.

Fonte: elaboração do autor a partir das obras citadas nesta seção. Quadro de distinção conceitual.

3. A assimetria institucional dos custos da divergência

3.1. Definição

Assimetria institucional dos custos da divergência é a configuração em que a organização concentra no profissional, de modo imediato, visível e atribuível, os custos de contrariar uma saída de IA. Ao mesmo tempo, torna tardios, difusos, externalizados ou institucionalmente absorvidos os custos de aderir a uma saída errada.

A palavra institucional é decisiva. O conceito não descreve apenas desconforto subjetivo, aversão individual à tecnologia ou dificuldade técnica inevitável. Ele exige regras, interfaces, metas, rotinas, hierarquias ou práticas probatórias que modifiquem sistematicamente a comparação entre as duas condutas. O objeto não é o sentimento de risco, mas a produção organizacional de risco profissional.

Divergência abrange rejeitar, corrigir, suspender, substituir, escalar ou pedir verificação adicional diante de uma saída automatizada. Adesão significa aceitar ou reproduzir a saída como fundamento material da decisão. Nenhuma das duas rotas é presumida correta. O conceito examina como a instituição trata a incerteza e distribui o custo de exercê-la.

3.2. Cinco espécies de custo

O custo da divergência não se limita à sanção disciplinar. Ele pode aparecer antes, durante e depois da decisão. Para fins diagnósticos, convém distinguir cinco espécies que frequentemente se acumulam:

Espécie	Na divergência	Quando a adesão errada ocorre
Deliberativo	Tempo para verificar, buscar fonte, consultar par e reconstruir o caso.	A verificação é dispensada; a falha aparece apenas se houver contestação ou dano.
Documental e probatório	Dever adicional de justificar a exceção e preservar evidências sob exposição nominal.	A aceitação pode ser registrada como fluxo ordinário, sem motivação comparável.
Métrico	Perda de produtividade aparente, aumento de fila, taxa de exceção ou desvio do padrão.	A rapidez integra o indicador positivo; o erro tardio pode ficar fora da janela de avaliação.
Relacional e reputacional	Conflito com chefia, equipe técnica ou cultura de confiança no sistema; rótulo de resistente.	A conformidade é lida como cooperação até que o problema se torne visível.
Disciplinar, econômico ou de carreira	Redução de remuneração, pior avaliação, tarefas punitivas, bloqueio de promoção ou sanção.	O custo pode ser absorvido por terceiro, orçamento, seguro, revisão futura ou reputação coletiva.

Fonte: elaboração do autor. Tipologia conceitual não validada empiricamente.

Essas espécies formam um vetor multidimensional e não uma soma aritmética. Tempo, reputação, remuneração, carga e exposição probatória são incomensuráveis sem uma convenção empírica que este artigo não fornece. A matriz, por isso, não atribui pesos nem produz um índice agregado.

O custo é material quando, segundo avaliação contextual e documentada, possui aptidão razoável para alterar a conduta do supervisor ou afetar tempo, carga, avaliação, remuneração, tarefas, reputação, carreira ou capacidade de contestação. Ficam fora do núcleo jurídico o mero dissabor e o esforço proporcional inerente à função, salvo quando a repetição ou o desenho organizacional lhes conferir efeito profissional relevante.

3.3. Seis dimensões da assimetria

As espécies acima dizem o que custa. As dimensões seguintes mostram como o custo é distribuído: incidência, temporalidade, visibilidade, atribuição, reversibilidade e controle da evidência. A assimetria é maior quando o mesmo evento concentra no divergente um custo pessoal imediato, legível pela gestão e difícil de reverter. Em contraste, o custo da adesão é diferido, diluído e documentado por quem escolheu o sistema.

A comparação não deve ser feita entre qualquer divergência e qualquer erro algorítmico. O contrafactual adequado reúne casos funcionalmente semelhantes: mesma função, período, classe de risco, informação disponível e complexidade. Sem esse cuidado, diferenças legítimas de tratamento podem ser confundidas com penalização assimétrica.

A igualdade relevante não é a igualdade numérica entre concordar e discordar. É a equivalência de critérios de diligência, risco e qualidade aplicados às duas rotas decisórias.

4. Mapa normativo: bases, comparadores e cautelas

4.1. Regulamento Europeu de Inteligência Artificial

O artigo 14 do Regulamento (UE) 2024/1689 exige que sistemas de alto risco sejam concebidos para supervisão efetiva por pessoas naturais. Entre os poderes previstos estão compreender capacidades e limitações, reconhecer tendência à sobreconfiança, interpretar a saída, decidir não usar o sistema, ignorar, substituir ou reverter o resultado e interromper a operação. O artigo 26 atribui ao responsável pela implantação a designação de pessoas com competência, formação, autoridade e apoio necessários.

Essas disposições oferecem base funcional para a tese, mas não contêm um regime geral de distribuição dos custos da divergência. O calendário foi alterado pelo Regulamento (UE) 2026/1744. A versão consolidada de 27 de julho de 2026 apenas incorpora essa alteração. As seções do capítulo III que abrangem os artigos 14 e 26 passam a aplicar-se, conforme a categoria de alto risco, em 2 de dezembro de 2027 ou 2 de agosto de 2028. O regulamento está em vigor, mas a incidência temporal dessas obrigações exige leitura do artigo 113 e do regime específico. Este artigo não antecipa sua aplicabilidade.

4.2. Diretiva sobre trabalho em plataformas: um comparador setorial expresso

A Diretiva (UE) 2024/2831 fornece o comparador mais direto. Seu artigo 10 exige recursos humanos suficientes para supervisão e avaliação de sistemas automatizados em plataformas digitais. As pessoas encarregadas devem possuir competência, formação e autoridade, inclusive para substituir decisões automatizadas. Devem ser protegidas contra dispensa ou medida equivalente, medidas disciplinares e outros tratamentos adversos quando exercem suas funções.

A norma não cria princípio transetorial automaticamente aplicável fora do trabalho em plataformas, nem produz efeito horizontal indiferenciado no Brasil. Sua importância é argumentativa: o legislador europeu reconheceu que autoridade técnica sem proteção funcional pode ser ineficaz. Em 22 de agosto de 2026, o prazo de transposição pelos Estados-membros — 2 de dezembro de 2026 — ainda não havia terminado. Por isso, a diretiva é utilizada como escolha regulatória setorial e não como regra nacional já uniformemente operacionalizada.

4.3. Direito brasileiro: reconstrução condicional, não direito geral

O direito brasileiro não contém, com alcance geral e essa denominação, uma regra pronta de simetria dos custos da divergência algorítmica. A tese depende da combinação entre a fonte setorial de supervisão e categorias já existentes. No vínculo de emprego, o artigo 2º da CLT atribui ao empregador os riscos da atividade econômica. Isso não elimina responsabilidade profissional do empregado, mas enfraquece a transferência automática ao trabalhador dos custos ordinários de uma salvaguarda que a empresa escolheu ou deve organizar.

Nas relações civis e contratuais, os artigos 186, 187, 422 e 927 do Código Civil oferecem categorias de ilicitude, abuso, boa-fé e reparação, sempre dependentes de seus requisitos. Uma organização pode exercer poderes legítimos de direção, auditoria e avaliação; o problema surge quando o exercício excede a finalidade do próprio arranjo de supervisão ou contradiz a confiança legitimamente criada pela promessa de julgamento independente. O artigo 373, § 1º, do Código de Processo Civil permite distribuição dinâmica do ônus da prova diante de dificuldade excessiva ou maior facilidade da parte contrária. A decisão deve ser fundamentada e oferecer à parte oportunidade de se desincumbir do encargo.

O artigo 20 da Lei Geral de Proteção de Dados assegura ao titular a solicitação de revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses. O dispositivo também prevê informações claras sobre critérios e procedimentos, observados os segredos comercial e industrial. Seu objeto é a posição do titular afetado, e não a distribuição dos custos profissionais suportados por quem supervisiona a decisão. A LGPD oferece, portanto, proteção relevante e adjacente, mas não resolve o problema específico deste artigo.

O Projeto de Lei nº 2.338/2023 integra o contexto legislativo, não o direito vigente. Em 22 de agosto de 2026, após aprovação no Senado, permanecia em tramitação na Câmara dos Deputados e aguardava parecer do relator na comissão especial. A proposição está sujeita à apreciação do Plenário. O projeto pode iluminar escolhas futuras sobre supervisão, governança e direitos, mas não deve ser citado como fonte de obrigação atualmente exigível.

Na Administração Pública, os princípios do artigo 37 da Constituição impõem legalidade, impessoalidade, moralidade, publicidade e eficiência. Eficiência não se reduz a volume. Quando uma norma setorial exige supervisão, a gestão de produtividade deve ser compatível com legalidade, motivação, controle e qualidade do resultado. A tese não autoriza afastar regimes estatutários, competências disciplinares ou critérios válidos de desempenho; exige coerência entre eles e a função de controle humano atribuída.

4.4. Resolução CNJ nº 615/2025

No Poder Judiciário, o fundamento é mais específico. A Resolução CNJ nº 615/2025, em texto compilado após a Resolução nº 674/2026, inclui participação e supervisão humana entre os fundamentos do uso responsável e exige supervisão efetiva, periódica e adequada ao risco. O artigo 32 assegura autonomia dos usuários internos, revisão detalhada, ausência de vinculação à solução e possibilidade de correções; seu parágrafo único impede que o sistema restrinja ou substitua a autoridade final. O artigo 34 exige supervisão humana e modificação, pelo magistrado competente, do produto gerado quando cabível.

A resolução não disciplina expressamente remuneração, promoção, distribuição de tarefas ou causalidade de medidas adversas. Ainda assim, seria contraditório preservar autoridade final na norma e neutralizá-la por indicadores que tratem toda intervenção como ineficiência. A assimetria pode funcionar aqui como critério de auditoria da implementação, não como fonte autônoma de direitos fora do regime aplicável.

Fonte	O que sustenta	O que não autoriza concluir
Regulamentos (UE) 2024/1689 e 2026/1744	Poderes de ignorar, substituir, reverter e interromper; competência, autoridade, apoio e calendário alterado.	Direito geral imediato contra qualquer custo profissional; aplicação uniforme antes do art. 113.
Diretiva (UE) 2024/2831	Proteção setorial expressa do supervisor contra tratamento adverso no trabalho em plataformas.	Transposição automática para todos os setores, países ou vínculos.
LGPD, CLT, Código Civil e CPC	Revisão da decisão automatizada; risco da atividade, boa-fé, abuso, dano e técnica probatória, conforme requisitos.	Direito profissional geral de divergência, responsabilidade objetiva universal ou inversão automática da prova.
PL nº 2.338/2023	Contexto legislativo em elaboração sobre IA, governança e supervisão.	Obrigação vigente ou resultado definitivo do processo legislativo.
Constituição e Resolução CNJ nº 615/2025	Coerência entre eficiência, autonomia, não vinculação, autoridade final e supervisão efetiva.	Imunidade funcional a erro, desídia ou descumprimento de deveres.

Fonte: elaboração do autor a partir das fontes normativas identificadas na primeira coluna. Síntese interpretativa, não transposição automática.

5. Regime jurídico proposto

5.1. Gatilho de relevância jurídica

A assimetria somente ingressa no plano jurídico quando vinculada a uma fonte e a uma relação de controle. O teste separa pressupostos, causalidade, defesa e indícios corroborantes. Essa sequência evita definir a assimetria pelo próprio escrutínio desigual que se pretende demonstrar.

Etapa	Pergunta e conteúdo mínimo	Função no teste
1. Pressuposto normativo	Há lei, regulamento, dever profissional, contrato, estatuto ou política vinculante que exija julgamento humano efetivo?	Sem fonte identificável, resta problema gerencial ou ético, não a mesma reconstrução jurídica.
2. Conduta qualificada	A divergência ocorreu de boa-fé, dentro da competência e com razão profissional identificável?	Exclui recusa arbitrária, fraude, sabotagem e negligência.
3. Controle institucional	O ator questionado controla tempo, métrica, avaliação, remuneração, disciplina, tarefas ou registros relevantes?	Limita o dever à esfera de poder efetivo.
4. Materialidade	Houve desvantagem apta a alterar conduta ou afetar trabalho, remuneração, reputação, carreira ou contestação?	Separa prejuízo relevante de dissabor ou esforço proporcional inerente.
5. Nexo causal	Sem a divergência, a medida provavelmente ocorreria no mesmo momento e da mesma forma?	A resposta negativa sustenta o nexos; proximidade temporal e comparadores são meios de prova.
6. Defesa autônoma	Existe razão legítima, contemporânea, independente e suficiente, documentada por critérios aplicados de modo coerente?	Se demonstrada, afasta ou limita a imputação, ainda que a divergência tenha precedido a medida.
7. Corroboração	Adesões e divergências comparáveis recebem critérios equivalentes de diligência, risco e qualidade?	O escrutínio desigual reforça a inferência; não é pressuposto circular nem exige auditoria numericamente idêntica.

Fonte: elaboração do autor. Teste jurídico proposto, sem pretensão de enunciar direito positivo geral.

5.2. Regra de coerência alocativa

Propõe-se a seguinte regra: a instituição que apresenta a supervisão humana como salvaguarda deve internalizar os custos organizacionais previsíveis necessários ao seu exercício. Não pode concentrar no supervisor os custos de divergir enquanto externaliza ou oculta os custos de aderir ao erro.

Internalizar não significa que a organização pagará todo custo possível ou aceitará qualquer tempo de revisão. Significa tratar tempo, formação, consulta, documentação, redundância e eventual queda de velocidade como componentes da própria arquitetura de segurança. O custo não pode reaparecer, disfarçado, como deficiência individual sempre que o profissional utiliza de boa-fé o mecanismo que a instituição diz exigir.

A mesma direção aparece, em caráter voluntário, no AI Risk Management Framework do NIST (2023): a liderança organizacional assume responsabilidade pelas decisões sobre riscos e as políticas devem diferenciar funções e responsabilidades nas configurações humano-IA. A referência não cria obrigação jurídica no Brasil, mas reforça que supervisão é questão de governança e não simples atributo moral do operador.

A regra propõe correspondência entre controle e responsabilidade, sempre segundo o regime jurídico aplicável. O fornecedor poderá responder, conforme a fonte incidente e seus requisitos, pelo que controla no desenho, na documentação e no mecanismo de intervenção.

O responsável pela implantação poderá responder por recursos, designação, formação e operação. Empregador, órgão ou gestor deverá ser examinado quanto a metas, incentivos, avaliação e medidas adversas sob seu domínio. O supervisor permanece responsável por diligência, competência, boa-fé e observância de seus deveres profissionais. Distribuir responsabilidade sem mapear controle produz tanto impunidade institucional quanto culpabilização indevida do humano.

5.3. Escrutínio equivalente

A principal consequência operacional é o escrutínio equivalente. A organização não precisa auditar igual quantidade de adesões e divergências, nem ignorar que exceções podem revelar riscos específicos. Deve, porém, utilizar critérios comparáveis de qualidade e diligência em amostras funcionalmente semelhantes, justificando diferenças pela classe de risco, pela materialidade, pela frequência e gravidade dos erros ou por finalidade de controle previamente definida.

Auditar apenas quem diverge cria viés de detecção: os erros do divergente são encontrados porque foram procurados; os erros de adesão permanecem invisíveis porque o fluxo padrão não foi examinado. A taxa individual de divergência, isoladamente, é ambígua. Pode indicar prudência, modelo inadequado, casos complexos, resistência improdutiva ou simples diferença de carteira. Transformá-la em proxy automática de desempenho elimina justamente a informação que a supervisão deveria gerar.

5.4. Causalidade, prova e razões legítimas

A causalidade deve evitar dois extremos: exigir confissão explícita da organização ou presumir que toda medida posterior é retaliatória. São indícios relevantes a proximidade temporal, a mudança de critério após a divergência, a concentração de auditorias no divergente e os comentários sobre sua taxa de adesão. Também importam o tratamento diferente de casos comparáveis, a desconsideração de avaliações positivas anteriores e a incompatibilidade entre a justificativa apresentada e os registros contemporâneos.

Em motivação mista, a organização pode demonstrar razão legítima, independente e suficiente. Entre os exemplos estão erro grosseiro, descumprimento de protocolo válido, insubordinação sem relação com o conteúdo técnico da divergência, fraude e conduta discriminatória praticada pelo supervisor.

Também podem ser relevantes a exposição indevida de dados e o desempenho inadequado comprovado por critérios equivalentes. A divergência não deve apagar esses fatos; tampouco esses rótulos podem substituir prova.

Quando métricas, logs, versões do modelo, amostras comparativas e comunicações de gestão estão sob controle exclusivo ou predominante da organização, a preservação e a exibição dos registros tornam-se centrais. A coleta deve observar finalidade, necessidade, proporcionalidade, segurança, prazos de retenção e acesso restrito, evitando converter a supervisão em vigilância indiscriminada. No processo civil brasileiro, eventual distribuição dinâmica do ônus da prova depende do artigo 373, § 1º, e de decisão fundamentada. A dificuldade probatória não autoriza transplantar presunções sem base normativa.

6. Matriz Diagnóstica Preliminar de Assimetria dos Custos

A matriz abaixo organiza evidências; não é escala validada, selo de conformidade ou fórmula de ilicitude. Suas dimensões não devem ser somadas mecanicamente. Um único defeito pode ser crítico — por exemplo, sanção automática por substituir a saída —, enquanto várias diferenças pequenas podem ser justificadas pelo risco. Sem validação, não há pesos, nota de corte ou certificado responsável.

Dimensão	Pergunta de controle	Evidência e sinal de assimetria
1. Incidência	Quem suporta o custo direto de revisar, divergir e corrigir?	Metas, remuneração, carga e tarefas. Sinal: custo da salvaguarda lançado apenas na conta individual.
2. Temporalidade	Quando o custo de cada rota aparece e qual janela a avaliação considera?	Ciclos de avaliação, reclamações, reversões. Sinal: ganho imediato de adesão e erro fora da janela.
3. Visibilidade	Quais condutas entram nas métricas, alertas e auditorias?	Painéis, critérios amostrais, alarmes. Sinal: divergência hiperregistrada e adesão presumida correta.
4. Atribuição	A responsabilidade acompanha poder real de decidir e controlar o sistema?	Permissões, papéis, documentação e cadeia decisória. Sinal: culpa individual com controle institucional limitado.
5. Reversibilidade	O profissional pode contestar a medida e corrigir o registro sem custo desproporcional?	Revisão independente, contraditório, prazos e correção de métricas. Sinal: rótulo ou perda irreversível.
6. Controle da evidência	Quem conserva logs, versões, comparadores e motivos da decisão adversa?	Retenção ilimitada, acesso seguro e trilha de auditoria. Sinal: exige-se prova que só a organização possui.
7. Escrutínio equivalente	Adesões e divergências comparáveis são avaliadas por critérios de risco e qualidade coerentes?	Plano amostral e resultados. Sinal: somente o divergente precisa demonstrar diligência.

Fonte: elaboração do autor. Matriz qualitativa preliminar, sem pesos, nota de corte ou validação empírica.

6.1. Protocolo mínimo de aplicação

1. Delimitar a decisão, a saída de IA, o efeito possível e a fonte concreta de supervisão ou autonomia.
2. Mapear atores e esferas de controle: fornecedor, responsável pela implantação, gestor do vínculo, supervisor e pessoa afetada.
3. Reunir políticas, telas, permissões, metas, tempos reais, amostras, avaliações, comunicações, logs e versões do sistema, com minimização de dados, controle de acesso e prazo de retenção.

4. Construir comparadores por função, risco, complexidade, período e informação disponível, evitando médias brutas.
5. Examinar separadamente custos deliberativos, métricos, relacionais, probatórios e disciplinares.
6. Registrar hipóteses concorrentes e razões legítimas independentes antes de concluir pela assimetria.
7. Aplicar um manual de codificação que defina unidade de análise, evidência admissível, materialidade e hipóteses concorrentes.
8. Pilotar o instrumento em casos independentes e medir concordância entre avaliadores antes de qualquer ponderação, pontuação ou certificação.

Uma aplicação exploratória pode classificar cada dimensão apenas como ausente, inconclusiva ou presente, acompanhada de evidência e hipótese alternativa. Mesmo essa codificação ordinal é provisória. Divergências entre avaliadores devem ser registradas e resolvidas por justificativa, não ocultadas por média numérica.

7. Cenários hipotéticos de aplicação

Os cenários seguintes são construções analíticas. Não descrevem organizações identificadas, não estimam prevalência e não validam a matriz. Servem apenas para testar se a tese distingue situações em contextos jurídicos distintos.

7.1. Poder Judiciário

Considere uma ferramenta que sugira minuta, classificação ou encaminhamento. O magistrado ou servidor que corrige a saída pode gastar mais tempo, registrar mais alterações e reduzir o volume aparente. Se o painel valoriza apenas quantidade e tempo, a intervenção exigida pela Resolução CNJ nº 615/2025 surge como ineficiência individual. Se a concordância errada for descoberta meses depois, o custo aparecerá em recurso, retrabalho, atraso ou perda de confiança, distante da métrica que orientou a conduta.

A resposta não é abolir produtividade nem conceder imunidade. É combinar indicadores de volume com complexidade, risco, retrabalho e qualidade; amostrar adesões e divergências; preservar versões; e impedir que a taxa de alteração seja tratada automaticamente como resistência. A autonomia prevista nos artigos 32 e 34 deve ser verificável na configuração técnica e na gestão do trabalho.

7.2. Relações de trabalho e profissões reguladas

Em empresas, o conflito pode aparecer quando uma ferramenta recomenda crédito, contratação, auditoria, manutenção, diagnóstico ou conduta operacional. A pessoa que diverge assume o ônus de interromper o fluxo e explicar a exceção; a adesão é processada como normalidade. A Diretiva (UE) 2024/2831 mostra, no setor de plataformas, uma solução expressa de proteção do supervisor. Fora desse campo, a incidência depende do regime brasileiro aplicável, do contrato, das normas profissionais e do controle efetivo do empregador ou gestor.

Uma política institucional robusta deve separar o registro de override, usado para segurança e aprendizagem, de uma presunção disciplinar. O mesmo dado pode ser relevante para detectar abuso, mas não deve decidir sozinho. Revisão humana efetiva exige que o tempo de revisão integre capacidade planejada e que erros de adesão também alimentem avaliação e melhoria do sistema.

7.3. Saúde e administração pública como comparações funcionais

Na saúde, a gravidade dos resultados falso-negativos e falso-positivos varia conforme o procedimento; por isso, taxas brutas de concordância são pouco informativas. O profissional deve responder por diligência e protocolo, enquanto a instituição responde por escolha, validação, formação, carga, integração e monitoramento dentro de sua esfera. Estudos sobre supervisão médica significativa reforçam que capacidade epistêmica, espaço cognitivo, autoridade e proteção da divergência justificada são interdependentes (VAN DE SANDE; ECONOMOU-ZAVLANOS; VAN GENDEREN, 2026).

Na administração pública, eficiência, impessoalidade e motivação impedem tanto o desperdício quanto a automatização cega. Uma recomendação pode padronizar decisões e reduzir arbitrariedade, mas também ocultar erro sistemático. A auditoria precisa verificar se a adesão recebe deferência não demonstrada e se o custo de contestar é transferido ao agente que preserva a legalidade. A conclusão depende sempre da norma setorial e da natureza da competência.

8. Objeções e respostas

8.1. Divergir realmente custa mais

Sim. Verificar e afastar uma recomendação pode exigir esforço adicional legítimo.

A tese não elimina esse custo natural; impede que a organização o apresente como falha individual depois de escolher a supervisão como controle de risco. A proporcionalidade permite revisão por amostragem, limiares e escalonamento conforme o impacto.

8.2. A simetria pode incentivar divergência performática

Pode, se for convertida em meta de discordância. Por isso, não se premia divergir nem se exige paridade de taxas. Avaliam-se razão, diligência e qualidade. Fraude, recusa arbitrária e uso estratégico da exceção permanecem sujeitos a controle por critérios transparentes e equivalentes.

8.3. A IA pode ser muito mais precisa que o humano

Mesmo um sistema superior em média pode errar em subgrupos, situações raras ou mudanças de contexto. Quanto melhor o sistema, menor talvez seja a frequência legítima de divergência, mas maior pode ser a dificuldade de detectar falhas excepcionais. Desempenho agregado não torna supérflua a salvaguarda exigida nem justifica penalização automática de toda intervenção.

8.4. O gestor precisa medir desempenho

Precisa. A crítica dirige-se a indicadores incompletos apresentados como neutros. Uma métrica válida deve explicitar finalidade, janela temporal, unidade de análise, risco, complexidade e efeitos comportamentais. Taxa de adesão pode ser sinal investigativo; não é, isoladamente, medida de competência.

8.5. A categoria pode inflar responsabilidade institucional

O risco é controlado pela esfera de controle e pelos requisitos de causalidade. Fornecedor, implantador, empregador, gestor e supervisor não respondem indistintamente. A matriz serve justamente para impedir imputação sem poder real, inclusive a culpabilização da instituição por conduta humana autônoma que ela não poderia razoavelmente prevenir.

8.6. A complexidade dos casos pode explicar a diferença

Pode. Casos difíceis tendem a gerar mais divergência, mais tempo e mais revisão. Essa endogeneidade impede comparar médias brutas entre quem diverge e quem adere. O desenho diagnóstico deve estratificar função, risco, carteira, informação disponível e complexidade. Se a diferença desaparecer após o controle desses fatores, não há base suficiente para imputar assimetria.

8.7. Auditar mais as exceções pode ser racional

Também pode. Exceções são informativas e podem concentrar risco. O escrutínio equivalente não exige amostras numericamente iguais; exige finalidade explícita, critério proporcional e comparação de qualidade. A seleção torna-se problemática quando a divergência é o único gatilho, a adesão recebe presunção de acerto e o resultado da auditoria alimenta sanção sem controle dos casos comparáveis.

8.8. Um modelo superior cria custo de oportunidade

Se o sistema apresenta desempenho robusto e validado, revisar todas as saídas pode desperdiçar recursos e atrasar benefícios. A resposta é calibrar a supervisão por risco, incerteza, mudança de contexto e consequências do erro. A tese não ordena revisão universal; impede que a economia obtida seja contabilizada apenas no fluxo padrão e que o custo residual de segurança seja imputado como deficiência do supervisor.

8.9. A proteção pode produzir risco moral

Sim, se proteger qualquer discordância. O teste exige boa-fé, competência e razão profissional identificável; admite controle de fraude, negligência e recusa arbitrária. A proteção incide sobre o exercício qualificado da função, não sobre o resultado correto garantido nem sobre imunidade disciplinar.

8.10. Registrar tudo ameaça privacidade e autonomia

Esse risco é real. Uma política de evidência pode virar vigilância granular, ampliar assimetrias de poder e expor dados pessoais ou sigilosos. Por isso, rastreabilidade deve observar finalidade, minimização, segurança, segregação de acesso, retenção limitada e revisão independente. Prova suficiente não equivale a monitoramento ilimitado.

8.11. A categoria apenas renomeia conceitos existentes

A objeção seria procedente se a expressão apenas descrevesse medo, retaliação ou culpa concentrada. Sua utilidade residual depende de quatro operações próprias. A categoria compara divergência e adesão errada, decompõe a distribuição dos custos no tempo e na prova, vincula a análise a uma fonte de supervisão e separa pressupostos, nexos, defesa e corroboração. Se não melhorar diagnóstico ou decisão em estudos futuros, deverá ser abandonada ou absorvida pelos conceitos adjacentes.

9. Limites e agenda de pesquisa

A proposta ainda não foi validada em campo. Não se conhece a prevalência das assimetrias, seus limiares de materialidade nem a confiabilidade entre avaliadores. Estudos futuros devem comparar políticas declaradas e práticas reais, observar painéis de desempenho, medir tempo de revisão, acompanhar reversões e entrevistar supervisores, gestores, equipes técnicas e pessoas afetadas.

Três hipóteses são prioritárias. Primeira: quanto maior a concentração nominal de responsabilidade no humano e menor seu controle efetivo, maior o risco de zona de deformação moral. Segunda: métricas de velocidade sem janela de erro tardio reduzem divergências mesmo quando a interface permite override. Terceira: auditoria baseada em risco, aplicada a concordâncias e divergências, aumenta o aprendizado e reduz tanto a conformidade automática quanto a oposição performática.

Também são necessários estudos jurídicos setoriais. Saúde, finanças, advocacia, magistratura, engenharia, aviação e serviço público possuem deveres, independência e regimes de responsabilidade diferentes. A categoria somente será útil se preservar essas diferenças. O desenvolvimento futuro deve testar tutelas preventivas, coletivas, contratuais, administrativas e processuais sem converter uma tese reconstrutiva em direito universal por afirmação.

10. Conclusão

A supervisão humana não fracassa apenas quando falta botão, informação ou competência. Ela também fracassa quando a organização distribui os custos de modo a tornar a concordância profissionalmente segura e a divergência pessoalmente perigosa. Nesse ambiente, o humano permanece formalmente no circuito, mas aprende que sua função real é ratificar.

A assimetria institucional dos custos da divergência nomeia essa arquitetura. Sua relevância jurídica depende de fonte, controle, conduta qualificada, desvantagem, causalidade e escrutínio desigual. A regra de coerência alocativa exige que a instituição que invoca supervisão internalize os custos previsíveis da salvaguarda e alinhe responsabilidade a poder real. O escrutínio equivalente impede que apenas o divergente precise provar diligência.

O Regulamento Europeu de Inteligência Artificial define poderes e condições de supervisão; a Diretiva sobre trabalho em plataformas protege, em setor delimitado, quem a exerce. No Brasil, normas setoriais, risco da atividade, boa-fé, abuso, prova e princípios administrativos oferecem bases condicionais, não um direito geral pronto. Juntas, essas fontes sustentam uma pergunta: se a instituição depende do humano para conter o erro da máquina, por que o custo do controle deve recair apenas sobre quem o exerce?

Uma supervisão que individualiza o custo da divergência e socializa o custo da obediência não controla a automação; ela organiza a conformidade.

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](https://doi.org/10.1016/0005-1098(83)90046-8).
- BRASIL. *Constituição da República Federativa do Brasil de 1988*. Brasília, DF: Presidência da República, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 22 ago. 2026.
- BRASIL. *Decreto-Lei nº 5.452, de 1º de maio de 1943*. Aprova a Consolidação das Leis do Trabalho. Texto compilado. Brasília, DF: Presidência da República, 1943. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del5452compilado.htm. Acesso em: 22 ago. 2026.
- BRASIL. *Lei nº 10.406, de 10 de janeiro de 2002*. Institui o Código Civil. Texto compilado. Brasília, DF: Presidência da República, 2002. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2002/10406compilada.htm. Acesso em: 22 ago. 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/13105.htm. Acesso em: 22 ago. 2026.
- BRASIL. *Lei nº 13.709, de 14 de agosto de 2018*. Lei Geral de Proteção de Dados Pessoais. Texto compilado. Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/13709.htm. Acesso em: 22 ago. 2026.
- CALABRESI, Guido. *The costs of accidents: a legal and economic analysis*. New Haven: Yale University Press, 1970.
- CÂMARA DOS DEPUTADOS (Brasil). *Projeto de Lei nº 2.338, de 2023*. Dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. Brasília, DF: Câmara dos Deputados, 2023. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao/?idProposicao=2487262>. Acesso em: 22 ago. 2026.
- CONSELHO NACIONAL DE JUSTIÇA (CNJ). *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. Texto compilado, com alteração pela Resolução nº 674, de 25 de março de 2026. Brasília, DF: CNJ, 2025. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/6001>. Acesso em: 22 ago. 2026.
- DEKKER, Sidney. *Just culture: restoring trust and accountability in your organization*. 3. ed. Boca Raton: CRC Press, 2016.
- DETERT, James R.; EDMONDSON, Amy C. Implicit voice theories: taken-for-granted rules of self-censorship at work. *Academy of Management Journal*, v. 54, n. 3, p. 461–488, 2011. DOI: <https://doi.org/10.5465/amj.2011.61967925>.
- EDMONDSON, Amy C. Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, v. 44, n. 2, p. 350–383, 1999. DOI: <https://doi.org/10.2307/2666999>.
- 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>.

Referências (continuação)

- EPELAND, 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/data-protection/our-work/publications/techdispatch/2025-09-23-techdispatch-22025-human-oversight-automated-making_en. Acesso em: 22 ago. 2026.
- GENOVA, Jandislei Antonio. *Quem pode contrariar a inteligência artificial? A divergência humana protegida como condição jurídica da supervisão efetiva*. São Paulo: publicação independente, 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>.
- 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*. Preprint, arXiv:2504.03300v2, 2025. Disponível em: <https://arxiv.org/abs/2504.03300>. Acesso em: 22 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>.
- MORRISON, Elizabeth Wolfe; MILLIKEN, Frances J. Organizational silence: a barrier to change and development in a pluralistic world. *Academy of Management Review*, v. 25, n. 4, p. 706–725, 2000. DOI: <https://doi.org/10.5465/amr.2000.3707697>.
- NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Gaithersburg, MD: NIST, 2023. DOI: <https://doi.org/10.6028/NIST.AI.100-1>.
- 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; HERMANN, 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. *Diretiva (UE) 2024/2831 do Parlamento Europeu e do Conselho, de 23 de outubro de 2024, relativa à melhoria das condições de trabalho em plataformas digitais*. Jornal Oficial da União Europeia, 2024. Disponível em: <https://eur-lex.europa.eu/eli/dir/2024/2831/oj>. Acesso em: 22 ago. 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*. Versão consolidada em 27 jul. 2026. Jornal Oficial da União Europeia, 2024. Disponível em: <https://eur-lex.europa.eu/eli/reg/2024/1689/2026-07-27/eng>. Acesso em: 22 ago. 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, 2026. Disponível em: <https://eur-lex.europa.eu/eli/reg/2026/1744/oj/por>. Acesso em: 22 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>.