REVISTA TÓPICOS
ISSN 2965-6672 | Qualis A2

# DA ASSINATURA HUMANA AO CONTROLE HUMANO JURIDICAMENTE SIGNIFICATIVO: LIMITES CONSTITUCIONAIS À INTELIGÊNCIA ARTIFICIAL GENERATIVA NA JURISDIÇÃO BRASILEIRA

**FROM HUMAN SIGNATURE TO LEGALLY MEANINGFUL HUMAN CONTROL: CONSTITUTIONAL LIMITS ON GENERATIVE ARTIFICIAL INTELLIGENCE IN BRAZILIAN ADJUDICATION**

Ciências Sociais Aplicadas • 21/11/2026

REGISTRO DOI: 10.70773/revistatopicos/794992977

Ariel Moraes Chiossi[1]

**RESUMO**

O artigo examina se a revisão humana formal basta para legitimar decisões judiciais apoiadas por inteligência artificial generativa. Emprega pesquisa qualitativa, jurídico-dogmática, bibliográfica e documental, com análise da Constituição, do Código de Processo Civil e da Resolução CNJ nº 615/2025, além de referências comparativas. Sustenta que a assinatura do juiz, por si só, não assegura controle sobre as razões da decisão. Propõe a Matriz de Relevância Algorítmica Judicial, com quatro níveis de uso, e um teste de atribuição humana efetiva, rastreabilidade, contestabilidade e transparência proporcional ao risco. Conclui que o apoio da IA exige salvaguardas graduais e preservação da responsabilidade jurisdicional.

**Palavras-chave:** Inteligência artificial; Poder Judiciário; Devido processo legal; Fundamentação judicial; Supervisão humana.

**ABSTRACT**

The article asks whether formal human review is enough to legitimize judicial decisions aided by generative artificial intelligence. It uses qualitative, doctrinal and documentary research to examine the Brazilian Constitution, the Code of Civil Procedure and CNJ Resolution No. 615/2025, with comparative support. It argues that a human signature alone does not ensure control over the reasons for a decision. It proposes a Judicial Algorithmic Relevance Matrix with four levels of use and a test of effective human attribution, traceability, contestability and risk-based transparency. It concludes that AI support requires safeguards matched to its role and preservation of judicial responsibility.

**Keywords:** Artificial intelligence; Judiciary; Due process; Judicial reasoning; Human oversight.

## 1. INTRODUÇÃO

A inteligência artificial generativa deixou de ser uma tecnologia periférica no debate jurídico. Modelos de linguagem capazes de resumir documentos, localizar padrões, estruturar argumentos e redigir textos em linguagem natural passaram a integrar rotinas de pesquisa, advocacia e gestão judicial. No Poder Judiciário, essa transformação é especialmente sensível porque a tecnologia não se limita a produzir eficiência administrativa: ela pode ingressar na própria infraestrutura informacional pela qual fatos, normas, precedentes e argumentos são selecionados e organizados antes da decisão.

O problema constitucional não é saber se um sistema consegue produzir um texto semelhante a uma sentença. Uma decisão jurisdicional não é definida pela aparência linguística do documento, mas pela existência de uma autoridade competente que aprecia os elementos do processo, enfrenta argumentos relevantes, interpreta as normas aplicáveis, justifica a conclusão e responde institucionalmente por ela. Por isso, a utilização de IA em ambiente decisório deve ser examinada a partir da Constituição, e não apenas de métricas de produtividade ou desempenho técnico.

A Constituição Federal de 1988 assegura que ninguém será privado da liberdade ou de seus bens sem o devido processo legal e garante aos litigantes o contraditório e a ampla defesa, nos termos do art. 5º, LIV e LV (Brasil, 1988). O art. 93, IX, por sua vez, exige fundamentação das decisões judiciais. No plano infraconstitucional, o art. 489, § 1º, do Código de Processo Civil especifica hipóteses em que uma decisão não se considera fundamentada, entre elas a utilização de razões genéricas, a invocação de precedentes sem demonstração de

pertinência e a ausência de enfrentamento de argumentos capazes de infirmar a conclusão adotada (Brasil, 2015, art. 489, § 1º, III, IV e V).

A Resolução CNJ nº 615/2025 consolidou um novo regime de governança de IA no Judiciário brasileiro. O ato incorpora categorias como supervisão humana, auditabilidade, explicabilidade, contestabilidade e classificação de riscos, além de disciplinar especificamente grandes modelos de linguagem e inteligência artificial generativa. A Resolução nº 674/2026 alterou o art. 15, § 1º, I, da norma para modificar a composição do Comitê Nacional de Inteligência Artificial do Judiciário, sem substituir o núcleo material das salvaguardas aplicáveis ao uso da tecnologia.[2]

O avanço regulatório, contudo, não encerra a questão. A própria Resolução nº 615/2025 admite que o uso de IA generativa na elaboração de ato judicial seja mencionado na decisão a critério do magistrado, ao mesmo tempo em que exige registro interno para fins de estatística, monitoramento e auditoria. Surge, então, um problema constitucional de fronteira: em que momento a participação da IA deixa de ser meramente instrumental e passa a ser materialmente relevante para a formação das razões de decidir, tornando insuficiente a transparência apenas interna?

A literatura brasileira recente já desenvolveu argumentos consistentes sobre opacidade, vieses e explicabilidade na decisão judicial assistida por IA. Toledo e Pessoa (2023) examinam a falta de transparência e os vieses no uso de sistemas de IA como apoio à decisão. Valle, Fuentes i Gasó e Ajus (2023) discutem transparência e intervenção do sistema Victor. Mais recentemente, Strachicini e Alencar (2026) sustentam a inteligência artificial explicável como requisito da motivação judicial. Este artigo parte dessa evolução,

mas procura enfrentar lacuna distinta: a insuficiência da categoria genérica de “supervisão humana” quando não se define o conteúdo mínimo da intervenção humana e a intensidade das salvaguardas em cada grau de interferência algorítmica.

A pergunta de pesquisa é a seguinte: a supervisão humana formal é suficiente para legitimar constitucionalmente uma decisão judicial cuja formação tenha sido materialmente auxiliada por inteligência artificial generativa? A hipótese é negativa. Defende-se que a legitimidade depende de controle humano juridicamente significativo, conceito que, neste trabalho, reúne quatro dimensões: atribuição humana efetiva, rastreabilidade da intervenção algorítmica, contestabilidade processual e transparência proporcional ao risco.

Como contribuição propositiva, apresenta-se a Matriz de Relevância Algorítmica Judicial - MRAJ, estruturada em quatro níveis de interferência. A matriz pretende resolver um problema prático: não é razoável impor o mesmo regime a um corretor gramatical e a um sistema que seleciona provas relevantes ou formula juízo conclusivo de aplicação do direito. Ao mesmo tempo, não é constitucionalmente adequado tratar todos os usos de IA como simples ferramentas neutras. O critério decisivo é a relevância material da intervenção tecnológica para as premissas ou para o resultado da decisão.

## 2. METODOLOGIA, DELIMITAÇÃO E ESTADO DA ARTE

A pesquisa é qualitativa, jurídico-dogmática, bibliográfica e documental. Emprega método predominantemente dedutivo: parte de garantias constitucionais estruturantes da jurisdição para avaliar

as condições de admissibilidade de diferentes formas de assistência algorítmica. A análise normativa concentra-se nos arts. 5º, LIII, LIV, LV, LXXVIII e LXXIX, e 93, IX, da Constituição; no art. 489 do Código de Processo Civil; em disposições pertinentes da Lei Geral de Proteção de Dados; e na Resolução CNJ nº 615/2025, considerada em sua redação vigente após a Resolução nº 674/2026.

A perspectiva comparativa é funcional e limitada. A pesquisa não realiza levantamento empírico sobre o comportamento de magistrados, não mede a acurácia de modelos e não pretende demonstrar causalidade sociológica. Suas conclusões são jurídico-interpretativas e propositivas, extraídas da leitura sistemática das normas, da literatura selecionada e dos casos oficiais examinados.

O Regulamento (UE) 2024/1689, conhecido como AI Act, é utilizado para observar a regulação de sistemas de alto risco empregados na administração da justiça e as exigências de supervisão humana. A Recomendação da UNESCO sobre Ética da Inteligência Artificial e a Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial, Direitos Humanos, Democracia e Estado de Direito são consideradas como parâmetros internacionais de governança. Não se realiza transposição automática desses instrumentos ao ordenamento brasileiro.

O recorte material é a utilização de IA generativa ou de sistemas algorítmicos capazes de influenciar a produção das razões de decidir. Ficam fora do núcleo do estudo aplicações puramente administrativas sem interferência na apreciação de fatos, provas, argumentos ou direito aplicável. O objetivo não é discutir personalidade jurídica da IA, responsabilidade civil do fornecedor em abstrato, proteção autoral de conteúdos gerados ou substituição

integral de magistrados por sistemas autônomos, embora alguns desses temas apareçam incidentalmente.

A opção por essa delimitação decorre do estado da arte. Surden (2019) demonstra que "inteligência artificial" agrega técnicas distintas e que a análise jurídica deve evitar antropomorfismos. Selbst et al. (2019) alertam para a inadequação de abstrair algoritmos de seus contextos sociotécnicos. Green (2022), ao analisar políticas governamentais de human oversight, sustenta que a simples inserção de um ser humano no fluxo decisório pode gerar falsa sensação de segurança quando não há condições institucionais para uma revisão efetiva. Citron (2008) demonstra que a automação de decisões públicas também envolve exigências de devido processo tecnológico. Re e Solow-Niederman (2019) defendem que sistemas de justiça artificial devem ser examinados à luz das instituições e dos valores jurídicos que pretendem apoiar. Laux e Ruschemeier (2025) alertam que a supervisão humana pode ser comprometida pelo viés de automação e por mecanismos insuficientes de correção. Esses estudos permitem deslocar a questão da presença nominal do humano para a qualidade jurídica e organizacional da supervisão.

A literatura sobre automação também oferece uma advertência clássica. Bainbridge (1983) identificou a ironia de sistemas automatizados que retiram do humano atividades rotineiras e, simultaneamente, deixam para ele as situações excepcionais em que são necessárias habilidades que a própria automação pode enfraquecer. No contexto contemporâneo, esse risco se aproxima do chamado viés de automação: a tendência de aceitar recomendações automatizadas com verificação insuficiente. O AI Act europeu incorpora expressamente a necessidade de que responsáveis pela

supervisão estejam conscientes desse risco em sistemas de alto risco.

No campo dos modelos de linguagem, Bender et al. (2021) advertiram que fluência linguística não equivale a compreensão nem garante correspondência com fatos. Ji et al. (2023) sistematizam o fenômeno da "alucinação" em geração de linguagem natural, expressão empregada tecnicamente para designar conteúdo não fiel à fonte ou factualmente incorreto, sem atribuir estados mentais humanos ao sistema.[3] Esse risco adquire especial relevância jurídica quando precedentes, dispositivos legais ou fatos processuais inexistentes são apresentados em linguagem plausível.

No Brasil, Toledo e Pessoa (2023) criticam opacidade e vieses na tomada de decisão judicial assistida por IA. Valle, Fuentes i Gasó e Ajus (2023) defendem mecanismos de transparência relacionados ao sistema Victor. Strachicini e Alencar (2026) vinculam explicabilidade à motivação judicial e propõem trilhas de auditoria, técnicas de explicação, supervisão humana significativa e mecanismos de contestação. Para evitar duplicação dessa agenda, o presente trabalho formula uma pergunta posterior: como transformar a supervisão humana significativa em um teste jurídico operacional, capaz de distinguir níveis de relevância algorítmica e deveres processuais correspondentes?

Essa diferença define a originalidade pretendida. O artigo não sustenta apenas que decisões assistidas por IA devem ser transparentes ou explicáveis. Propõe que a intensidade das obrigações constitucionais varie de acordo com a proximidade da tecnologia em relação ao núcleo decisório, e que a facultatividade

de revelar o uso de IA, prevista na disciplina administrativa do CNJ, seja interpretada à luz do devido processo e do contraditório quando a intervenção tecnológica for materialmente relevante.

### 2.1. Critérios de Análise e Procedimento de Aplicação

A construção da MRAJ segue três etapas. Na primeira, identifica-se a finalidade concreta do sistema e o tipo de output produzido. Na segunda, verifica-se se esse output participa da identificação, seleção ou valoração de fatos, provas, argumentos, normas ou precedentes. Na terceira, examina-se se o conteúdo foi materialmente incorporado à fundamentação ou ao resultado e se o conhecimento da intervenção poderia alterar a estratégia de contestação da parte.

A classificação proposta não pretende substituir as categorias oficiais de risco do Conselho Nacional de Justiça. Ela funciona como instrumento analítico para aferir a relevância processual concreta do uso. Para reduzir arbitrariedade, a classificação deve considerar cumulativamente: (i) proximidade da ferramenta em relação às premissas decisórias; (ii) potencial de impacto sobre liberdade, patrimônio, igualdade processual ou outros direitos; (iii) grau de autonomia do sistema; (iv) possibilidade de reconstrução das fontes e transformações realizadas; e (v) capacidade efetiva de revisão pelo magistrado.

O teste é aplicado a três situações-modelo. Primeiro, se a ferramenta apenas corrige ortografia de texto já elaborado pelo magistrado, o uso tende ao Nível 1: exige revisão integral e registro interno, mas não necessariamente comunicação específica na decisão. Segundo, se o sistema resume depoimentos ou seleciona precedentes que

serão utilizados como premissas, o uso tende ao Nível 2: exige conferência direta nos autos e nas fontes, rastreabilidade e, se houver incorporação material, comunicação suficiente para contestação. Terceiro, se o sistema valora a credibilidade de testemunhas, classifica risco individual ou recomenda a conclusão jurídica do caso, o uso tende ao Nível 3: a análise começa pelas vedações e restrições do CNJ; apenas nos usos juridicamente admitidos se examinam as salvaguardas máximas, sem que a revisão meramente formal seja suficiente.

Essas situações são exemplos metodológicos, e não afirmações sobre a existência de sistemas específicos em funcionamento no Judiciário brasileiro. A finalidade é demonstrar como a matriz pode ser utilizada em casos concretos sem converter a classificação em presunção automática de invalidade.

## 3. IA GENERATIVA E A INFRAESTRUTURA COGNITIVA DA DECISÃO

A distinção entre instrumento e infraestrutura cognitiva é central. Um sistema utilizado para converter áudio em texto, corrigir ortografia ou padronizar formatação não ocupa a mesma posição institucional de uma ferramenta que resume depoimentos, seleciona precedentes, sugere enquadramentos jurídicos ou elabora uma minuta com conclusão sobre o mérito. Nos dois casos existe automação; apenas no segundo a tecnologia pode influenciar diretamente a formação das premissas decisórias.

Essa influência não precisa ser determinística. Um sistema pode ser formalmente consultivo e, ainda assim, exercer efeito relevante sobre a decisão. A ordem de apresentação dos argumentos, a seleção de trechos dos autos, o conjunto de precedentes sugeridos e

a formulação inicial de uma hipótese jurídica funcionam como elementos de enquadramento cognitivo. Em ambiente de alta carga processual, a recomendação inicial pode converter-se em ponto de partida dominante para a revisão humana.

O problema é agravado pela assimetria entre forma e confiabilidade. Modelos generativos são capazes de produzir textos coesos, referências aparentes e justificativas linguisticamente sofisticadas. A correção estilística, entretanto, não comprova a existência do precedente citado, a fidelidade ao documento resumido ou a adequação da norma ao caso. O erro pode aparecer não como texto absurdo, mas como uma falsa afirmação revestida de plausibilidade jurídica.

Casos recentes no Superior Tribunal de Justiça ilustram o risco do uso não verificado. Segundo notícia institucional do STJ, de 19 de maio de 2026, referente ao HC 1.094.270, o ministro Rogerio Schietti Cruz registrou falhas graves em petição com referências erradas e trechos de julgados inexistentes, diante de indícios de uso de IA, e determinou comunicação à OAB. O episódio não prova falha estrutural da tecnologia nem envolve decisão judicial automatizada; demonstra, porém, que a aparência de autoridade textual pode introduzir material inexistente no processo se o usuário não realizar conferência independente.

Outro vetor de risco é adversarial. No REsp 2.256.731/PB, julgado em 1º de setembro de 2026, a Sexta Turma do STJ, sob relatoria do ministro Og Fernandes, identificou comando oculto em petição destinado a influenciar sistemas de inteligência artificial e determinou comunicação à OAB e ao Ministério Público Federal. O acórdão qualificou a conduta, em tese, como ato atentatório à

dignidade da justiça e encaminhou a apuração às instituições competentes, sem resolver definitivamente a responsabilidade disciplinar ou penal do subscritor. O episódio evidencia que a governança da IA judicial deve abranger não apenas erro do modelo, mas segurança contra manipulação do próprio fluxo informacional.

Esses exemplos indicam que a confiabilidade da decisão assistida não depende apenas da qualidade estatística do modelo. Depende da cadeia sociotécnica completa: dados inseridos, permissões, interface, procedência das fontes, capacidade de auditoria, treinamento dos usuários, prevenção contra comandos maliciosos e rotinas de verificação. Selbst et al. (2019) denominam de “armadilha da abstração” a tendência de tratar o sistema técnico isoladamente e ignorar o ambiente institucional em que opera.

No Judiciário, essa perspectiva possui consequência constitucional. Se a tecnologia organiza de maneira decisiva aquilo que o julgador vê, omite ou prioriza, a arquitetura do sistema passa a afetar as condições concretas de exercício da jurisdição. O controle de constitucionalidade do uso de IA, portanto, não deve perguntar apenas “quem assinou?”, mas também “como foram formadas as premissas que a pessoa que assinou recebeu, verificou e utilizou?”.

A noção de infraestrutura cognitiva evita dois extremos. De um lado, impede tratar qualquer software como ameaça à independência judicial. De outro, impede reduzir sistemas de alta influência a ferramentas neutras apenas porque a assinatura final permanece humana. Entre a automação administrativa e a decisão totalmente automatizada existe uma ampla zona de assistência decisória que exige critérios jurídicos graduais.

## 4. FUNDAMENTO CONSTITUCIONAL DO CONTROLE HUMANO SIGNIFICATIVO

O primeiro fundamento é o juiz natural e a autoridade competente. O art. 5º, LIII, da Constituição estabelece que ninguém será processado nem sentenciado senão pela autoridade competente. A IA não possui investidura jurisdicional, independência funcional, garantias da magistratura nem regime constitucional de responsabilidade. Isso não impede seu uso como instrumento de apoio, mas impede que a autoridade decisória seja materialmente deslocada para um sistema cuja conclusão o magistrado apenas ratifica sem exame próprio.

O segundo fundamento é o devido processo legal, previsto no art. 5º, LIV. Na dimensão processual, o dispositivo exige que a decisão estatal seja produzida por procedimento compatível com as garantias reconhecidas às partes. A introdução de uma camada algorítmica que produza premissas decisivas e não possa ser identificada, auditada ou contestada pode romper essa estrutura, ainda que o ato final seja formalmente subscrito por autoridade humana.

O terceiro fundamento é o contraditório e a ampla defesa, assegurados pelo art. 5º, LV. No processo contemporâneo, contraditório não se reduz à possibilidade ritual de falar. Para que a participação seja efetiva, a parte deve conhecer os elementos relevantes que influenciam a decisão e ter oportunidade de respondê-los quando juridicamente necessário. Se uma ferramenta externa aos autos introduz precedente, classificação, síntese factual ou inferência que passa a compor a razão decisória, a invisibilidade dessa influência pode impedir contestação útil.

O quarto fundamento é a motivação, no art. 93, IX, da Constituição. A exigência de fundamentação não se satisfaz pela existência física de páginas contendo texto jurídico. A motivação deve permitir controle racional da relação entre fatos, provas, normas e conclusão. O CPC densifica essa garantia no art. 489, § 1º. Uma minuta produzida por IA pode, em tese, preencher formalmente a estrutura de uma decisão e simultaneamente padecer dos mesmos vícios que o art. 489 pretende evitar: razões genéricas, precedentes descontextualizados, omissão de argumentos ou emprego de conceitos indeterminados sem concretização.

O quinto fundamento, quando houver tratamento de dados pessoais, é o direito fundamental à proteção de dados, incluído no art. 5º, LXXIX, da Constituição. O art. 6º da LGPD acrescenta princípios de finalidade, necessidade, qualidade dos dados, transparência, segurança, prevenção, não discriminação e responsabilização (Brasil, 2018). O art. 20 da LGPD trata do direito de solicitar revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem interesses do titular. Esse dispositivo não deve ser transposto automaticamente para qualquer decisão judicial assistida por IA, pois o regime da jurisdição possui estrutura constitucional própria e a norma se refere a decisões unicamente automatizadas. Além disso, a redação vigente do art. 20 não estabelece, de forma expressa, que toda revisão deva ser realizada por pessoa natural; o controle humano jurisdicional aqui defendido decorre principalmente das garantias constitucionais do processo. Ainda assim, seus princípios reforçam a necessidade de governança quando dados processuais são submetidos a sistemas de IA.

A razoável duração do processo, prevista no art. 5º, LXXVIII, também integra a análise. A IA pode contribuir para reduzir tarefas repetitivas, acelerar pesquisa e organizar grandes volumes de informação. Todavia, eficiência não é autorização constitucional para reduzir contraditório, motivação ou responsabilidade. A duração razoável deve ser perseguida dentro do devido processo, não em substituição a ele.

Desses fundamentos resulta uma proposição: a legitimidade da assistência algorítmica depende de preservação substancial, e não apenas formal, da autoria jurisdicional. A assinatura do magistrado é condição necessária, mas deixa de ser suficiente quando não há demonstração institucional de que o conteúdo foi submetido a verificação adequada. O ponto não é exigir prova psicológica do raciocínio interno do juiz. É exigir arquitetura procedimental capaz de tornar plausível, verificável e responsabilizável a intervenção humana.

Essa conclusão aproxima-se da crítica de Green (2022). A exigência genérica de human in the loop pode funcionar como selo de legitimidade sem modificar os riscos do sistema. Para que o humano seja salvaguarda real, precisa possuir autoridade para discordar, informação para compreender a recomendação, tempo e treinamento para verificá-la, e meios técnicos para reconstruir o que o sistema fez. No contexto judicial, a essas condições somam-se deveres constitucionais de fundamentação e abertura ao contraditório.

## 5. A RESOLUÇÃO CNJ Nº 615/2025 E O PARADOXO DA TRANSPARÊNCIA FACULTATIVA

A Resolução CNJ nº 615/2025 fornece o ponto normativo mais específico para o problema. Seu objeto é o desenvolvimento, a utilização e a governança de soluções de IA no Poder Judiciário, com atenção a direitos fundamentais, segurança, transparência e responsabilização. O ato substituiu a Resolução nº 332/2020 após período de vacatio previsto em seu texto e estabeleceu abordagem baseada em risco. Os arts. 3º, 32 e 34 da norma incorporam como princípios e parâmetros a transparência, explicabilidade, contestabilidade, auditabilidade e supervisão humana. O art. 10, I, também veda soluções que impossibilitem a revisão humana ou criem dependência absoluta do usuário em relação ao resultado apresentado. No Anexo de Classificação de Riscos, item AR2, considera de alto risco, entre outras finalidades, a aferição da adequação e a valoração de meios de prova em processos contenciosos quando essas avaliações possam influenciar diretamente a decisão judicial.

A classificação é juridicamente coerente com a Constituição. Valoração de prova e formulação de juízos conclusivos sobre a aplicação de normas ou precedentes aproximam a IA do núcleo da função jurisdicional. Nesses casos, o risco não está apenas em eventual erro probabilístico. Está na possibilidade de afetar liberdade, patrimônio, igualdade processual e formação da razão decisória sem suficiente controle humano.

O regime dos grandes modelos de linguagem e da IA generativa é particularmente relevante. O art. 20, IV, da Resolução estabelece caráter auxiliar e complementar dessas ferramentas e veda a utilização autônoma para tomada de decisões judiciais sem orientação, interpretação, verificação e revisão pelo magistrado, que permanece responsável pelo conteúdo. Para finalidades classificadas

como de alto risco ou risco excessivo, os arts. 19, § 3º, V, e 20, VI, vedam o emprego de modelos privados ou externos ao Judiciário, em coerência com preocupações de segurança, confidencialidade e controle institucional.

A Resolução também exige registros e mecanismos de auditoria. O art. 19, § 6º, prevê registro interno automático do emprego de IA generativa no auxílio à redação de ato judicial, para fins de estatística, monitoramento e eventual auditoria. Ao mesmo tempo, a indicação de que o texto foi auxiliado por IA pode ficar a critério do magistrado. No art. 33, § 3º, a comunicação sobre eventual uso de IA no texto de decisões judiciais é tratada como facultativa ao signatário.

Essa combinação cria o que este artigo denomina "paradoxo da transparência facultativa". O sistema administrativo sabe que a IA foi utilizada e preserva um registro interno, mas a parte pode não saber que premissas relevantes da decisão foram selecionadas, sintetizadas ou estruturadas com auxílio algorítmico. A solução é adequada quando a ferramenta atua em plano meramente formal. Torna-se mais difícil de compatibilizar com o contraditório quando a intervenção é material.

Não se defende que toda correção gramatical, tradução ou reorganização de parágrafo deva aparecer na decisão. Um dever indiscriminado de divulgação produziria ruído e poderia banalizar a transparência. O argumento é mais restrito e propositivo: defende-se uma interpretação constitucional da facultatividade que varie de acordo com a relevância da intervenção. Essa leitura não equivale a afirmar que o CNJ já imponha divulgação individualizada em todo uso cognitivo de IA. Quando o sistema contribui para premissa

fática, probatória ou jurídica que influencia o resultado, a transparência deixa de ser simples opção de estilo e passa a integrar as condições de contestabilidade.

Essa leitura também preserva a competência normativa do CNJ sem convertê-la em substituto da Constituição. A Resolução organiza governança interna e define salvaguardas administrativas. Ela não pode ser lida como autorização para reduzir direitos processuais fundamentais. Assim, a expressão “a critério do magistrado” deve operar plenamente nos usos de baixa relevância e sofrer interpretação restritiva nas hipóteses em que a ocultação da participação algorítmica impeça controle da decisão.

A Resolução nº 674/2026 não modifica essa conclusão, pois alterou especificamente a composição prevista no art. 15. Seu significado para este estudo é apenas confirmar que o texto da Resolução CNJ nº 615/2025 deve ser consultado em versão atualizada.

O Projeto de Lei nº 2.338/2023, por sua vez, permanece em tramitação na Câmara dos Deputados na consulta de 9 de outubro de 2026 e não integra o direito positivo como lei federal.[4] Pode ser considerado contexto legislativo, mas não fundamento normativo vigente.

A abordagem do CNJ converge, em aspectos relevantes, com o AI Act europeu. O Regulamento (UE) 2024/1689 trata determinados sistemas utilizados por autoridades judiciais para pesquisar e interpretar fatos e direito e aplicar a lei a situações concretas como de alto risco. A supervisão humana deve ser efetiva e proporcional aos riscos, ao nível de autonomia e ao contexto de uso, e os responsáveis devem poder compreender limitações, interpretar

resultados, desconsiderá-los, substituí-los ou interromper o sistema. A regulação europeia também alerta para o risco de dependência excessiva do resultado automatizado.

Não se trata de importar a disciplina europeia para o processo brasileiro. A comparação serve para mostrar uma convergência regulatória: quanto mais a IA se aproxima da aplicação do direito e da apreciação de fatos, mais insuficiente se torna a ideia abstrata de presença humana. A supervisão precisa ter conteúdo funcional.

Uma leitura concorrente preservaria a faculdade prevista no art. 33, § 3º, mesmo em usos cognitivamente relevantes, sob o argumento de que a responsabilidade do magistrado, o acesso às razões da decisão e os registros internos já oferecem controle suficiente. A objeção merece consideração: divulgação excessiva pode produzir custos, revelar dados protegidos e confundir explicação técnica com fundamentação judicial. A resposta proposta não é publicidade irrestrita do modelo, mas informação processual limitada ao conteúdo materialmente utilizado e aos elementos necessários para verificar sua origem e pertinência, respeitados o segredo de justiça, a proteção de dados e os direitos de terceiros.

## 6. DA SUPERVISÃO FORMAL AO CONTROLE HUMANO JURIDICAMENTE SIGNIFICATIVO

O conceito de controle humano juridicamente significativo proposto neste artigo procura reunir condições técnicas, institucionais e processuais em um teste aplicável à jurisdição. A expressão “significativo” não significa que todo magistrado deva compreender matematicamente a arquitetura interna do modelo. Significa que a intervenção humana deve ser capaz de produzir diferença jurídica

real: confirmar, rejeitar, corrigir e justificar independentemente o resultado tecnológico.

A primeira dimensão é a atribuição humana efetiva. A decisão precisa ser materialmente atribuível à autoridade competente, e não apenas formalmente assinada por ela.

Isso pressupõe que o magistrado detenha poder real de alterar o resultado, examine criticamente os fundamentos e não trate a recomendação algorítmica como presunção de correção. A autoridade constitucional não pode ser deslocada pela arquitetura da interface.

A segunda dimensão é a rastreabilidade. Deve ser possível reconstruir, em nível compatível com o risco, qual sistema foi utilizado, sua versão, a finalidade do uso, os dados ou documentos relevantes submetidos, os outputs materialmente incorporados e as intervenções humanas subsequentes. Rastreabilidade não exige necessariamente divulgação de código-fonte, pesos de modelo ou segredo industrial. Exige uma trilha capaz de sustentar auditoria e responsabilização.

A terceira dimensão é a contestabilidade. Quando a IA interfere em elemento capaz de influenciar o resultado, as partes precisam dispor de mecanismo para questionar o conteúdo relevante. Isso pode envolver informação sobre o uso, acesso aos elementos não protegidos necessários à compreensão, possibilidade de demonstrar erro de fonte, inadequação do precedente, viés ou inconsistência com os autos e revisão efetiva pela autoridade humana.

A quarta dimensão é a transparência proporcional ao risco. Transparência deve ser funcional, e não ornamental. A divulgação

deve aumentar quando cresce a proximidade da IA com o núcleo decisório. A Recomendação da UNESCO sobre Ética da Inteligência Artificial adota lógica semelhante ao relacionar transparência e explicabilidade ao contexto e ao impacto, com ênfase na possibilidade de compreender e contestar decisões que afetem direitos.

As quatro dimensões são cumulativas, mas sua intensidade varia. Em um corretor ortográfico integrado ao editor, a atribuição humana permanece praticamente intacta, a rastreabilidade pode ser simples e não há necessidade de contraditório específico. Em um sistema que resume a prova testemunhal e sugere quais depoimentos são mais consistentes, a rastreabilidade deve ser detalhada, a revisão precisa ser independente, a utilização deve observar o regime de alto risco e a parte deve poder questionar qualquer premissa materialmente incorporada.

O teste evita transformar “humano no circuito” em fórmula vazia. Uma pessoa que apenas clica “aceitar” não exerce controle relevante. Uma pessoa que tem autoridade para rejeitar, compreende a finalidade do sistema, conhece suas limitações, confere fontes, confronta o output com os autos e registra correções exerce controle qualitativamente diferente. O direito deve distinguir essas situações.

Também é necessário reconhecer limite epistêmico. Nem toda opacidade pode ser eliminada por explicações simplificadas, e nem toda explicação técnica é útil à parte. Por isso, o objetivo jurídico não deve ser exigir transparência absoluta do modelo, mas transparência suficiente para controlar a premissa utilizada no processo. Se o sistema indica um precedente, a questão central é permitir verificar

qual precedente, qual trecho, qual pertinência e se ele existe; não necessariamente revelar cada parâmetro matemático que participou da geração do texto.

Esse enfoque reduz um equívoco comum: confundir explicabilidade do modelo com fundamentação da decisão. A fundamentação é obrigação do órgão jurisdicional. A explicabilidade tecnológica pode ser um instrumento para torná-la controlável, mas não substitui a exposição humana das razões jurídicas. O juiz não pode delegar à "explicação do algoritmo" o dever constitucional de justificar sua conclusão.

Por fim, o controle humano juridicamente significativo exige condições organizacionais. Tempo, treinamento, desenho de interface, alertas de risco, acesso às fontes e governança de incidentes são elementos da proteção. Uma regra que imponha revisão humana, mas estruture o fluxo para aceitação automática em massa, pode falhar na prática. Nesse ponto, a proposta converge com Green (2022): supervisão humana precisa ser apoiada por supervisão institucional.

## 7. MATRIZ DE RELEVÂNCIA ALGORÍTMICA JUDICIAL - MRAJ

A Matriz de Relevância Algorítmica Judicial - MRAJ é uma proposta normativa e analítica deste artigo, não uma classificação oficialmente adotada pelo CNJ.[5] Sua finalidade é converter a ideia de proporcionalidade em critério operacional. A matriz organiza usos de IA em quatro níveis conforme a capacidade de influenciar premissas ou resultado da decisão.

O Nível 0 corresponde à automação administrativa ou procedimental sem interferência substantiva na apreciação do caso. Inclui, por exemplo, organização de filas, conversão de formatos, detecção de duplicidades documentais e tarefas equivalentes, desde que não produzam classificação capaz de afetar direitos ou prioridades de maneira juridicamente relevante. As salvaguardas concentram-se em segurança, proteção de dados, registros institucionais e governança.

O Nível 1 compreende apoio redacional ou formal. São exemplos correção gramatical, padronização, tradução e melhoria de clareza, bem como síntese auxiliar que não seja utilizada como premissa autônoma. A decisão continua formada por razões previamente definidas pelo magistrado. O dever mínimo é revisão integral do texto final e verificação de que a ferramenta não introduziu fatos, citações ou fundamentos inexistentes. O registro interno é adequado; a divulgação individualizada no corpo da decisão, em regra, pode permanecer facultativa.

O Nível 2 abrange apoio cognitivo relevante: pesquisa de precedentes, seleção e síntese de argumentos, organização de fatos, comparação de documentos e sugestões jurídicas capazes de influenciar a fundamentação. Aqui a simples revisão textual não basta. Exigem-se identificação e verificação das fontes, rastreabilidade do uso, validação independente das premissas e documentação suficiente para auditoria. Se o conteúdo algorítmico for materialmente incorporado à razão de decidir, este artigo sustenta que deve haver divulgação específica suficiente para permitir contestação, ainda que a norma administrativa trate a menção à IA, em abstrato, como facultativa.

O Nível 3 envolve interferência decisória sensível: valoração de prova, avaliação de credibilidade, inferência sobre comportamento, formulação de juízo conclusivo de aplicação de normas ou precedentes a fatos concretos, classificação de risco individual ou recomendação capaz de determinar substancialmente o resultado. Essas hipóteses aproximam-se das categorias de alto risco ou de risco excessivo previstas pela Resolução nº 615/2025. A primeira verificação é a admissibilidade do uso: as hipóteses de risco excessivo permanecem vedadas, e a classificação da MRAJ não afasta as restrições a modelos privados ou externos. Nos usos admitidos, o regime deve ser máximo: sistema institucionalmente autorizado, avaliação de impacto quando exigida, supervisão efetiva, auditoria, registro detalhado, proteção contra dependência do output, informação às partes quando a intervenção for material e possibilidade de questionamento.

A matriz não pretende substituir a classificação oficial de riscos. Ela opera em plano diferente. O CNJ classifica finalidades e contextos para fins de governança administrativa. A MRAJ pergunta qual é a relevância processual concreta da intervenção em determinado ato. Um mesmo sistema pode atuar em Nível 1 em uma tarefa e em Nível 2 em outra, dependendo de como seu output é utilizado.

**Quadro 1.** Matriz de Relevância Algorítmica Judicial (MRAJ)

| **Nível** | **Função** | **Salvaguardas mínimas** | **Efeito processual proposto** |
|---|---|---|---|

| | | | |
|---|---|---|---|
| 0 - Administrativo | Automação sem influência substantiva nas premissas decisórias. | Segurança, proteção de dados, governança e registros compatíveis. | Divulgação no ato normalmente desnecessária. |
| 1 - Redacional | Apoio formal, linguístico ou organizacional sem premissa autônoma. | Revisão integral do texto, conferência de citações e registro interno. | Divulgação em regra facultativa. |
| 2 - Cognitivo relevante | Pesquisa, síntese, seleção de fatos ou argumentos e sugestão jurídica incorporável. | Rastreabilidade das fontes, validação independente, registros e revisão crítica. | Divulgação quando houver incorporação material, suficiente para contestação. |
| 3 - Decisório sensível | Valoração da prova, inferência individual, aplicação conclusiva do direito ou recomendação determinante. | Verificação prévia das vedações. Nos usos admitidos: autorização, avaliação de impacto quando cabível, auditoria, supervisão efetiva e segurança reforçada. | Transparência específica e contestabilidade como regra proposta; prevalecem as vedações do CNJ. |

**Fonte:** elaboração do autor.

A aplicação da matriz pode ser feita por três perguntas sucessivas. Primeira: o output da IA contribuiu para identificar, selecionar ou valorar fato, prova, argumento, norma ou precedente? Segunda: esse conteúdo foi materialmente incorporado à justificativa ou ao resultado? Terceira: a parte teria estratégia processual diferente se soubesse como essa contribuição ocorreu? Quanto mais respostas

positivas, maior a relevância algorítmica e mais intenso o regime de transparência e controle.

A vantagem é evitar dois erros simétricos. O primeiro é a hipertransparência indiferenciada, que transforma qualquer uso trivial de software em informação processual obrigatória e dificulta distinguir o que realmente importa. O segundo é a subtransparência, em que aplicações decisivamente influentes permanecem ocultas sob a fórmula genérica de que o magistrado "revisou" o resultado.

Em termos de validade, a mera utilização de IA não deve produzir nulidade automática. O vício jurídico depende da forma como a tecnologia afetou garantias processuais. Se uma ferramenta de Nível 1 foi utilizada, o texto foi integralmente conferido e nenhuma premissa estranha foi introduzida, não há razão para presumir invalidade. Se, ao contrário, sistema de Nível 2 ou 3 introduziu conteúdo decisivo não submetido a controle, ocultou fonte inexistente ou impediu contraditório útil, pode haver violação dos arts. 5º, LIV e LV, e 93, IX, da Constituição, além do padrão de fundamentação do art. 489 do CPC.

A análise da consequência processual deve considerar o nexo entre a falha de governança e o conteúdo da decisão. O objetivo não é criar uma nulidade tecnológica autônoma, mas reconhecer que a IA pode ser o meio pelo qual se produz defeito constitucional já conhecido: decisão sem fundamentação controlável, uso de elemento não submetido ao contraditório ou deslocamento material da autoridade decisória.

## 8. SEGURANÇA ADVERSARIAL, ALUCINAÇÕES E DEVER DE VERIFICAÇÃO

A legitimidade da IA judicial também depende de integridade informacional. O problema não se resume a vieses de treinamento ou opacidade. Modelos de linguagem podem ser induzidos por instruções maliciosas presentes em documentos, fenômeno conhecido como prompt injection. Quando sistemas processam automaticamente conteúdo fornecido pelas partes, texto oculto ou instruções especialmente construídas podem tentar modificar o comportamento do modelo.

O REsp 2.256.731/PB, julgado pela Sexta Turma do STJ em 1º de setembro de 2026, representa decisão judicial relevante para ilustrar esse risco. O acórdão identificou uma instrução oculta na petição, voltada a influenciar a resposta de sistemas de IA, e determinou a comunicação à OAB e ao Ministério Público Federal para apuração. A defesa afirmou desconhecer a origem da inserção, invocou boa-fé e relatou a adoção de auditoria interna. Esse contraponto delimita o alcance do julgamento: o reconhecimento de indícios de manipulação e o encaminhamento para apuração não equivalem à condenação disciplinar ou penal definitiva do subscritor. O caso não demonstra que decisões do STJ sejam produzidas automaticamente; ao contrário, o relator enfatizou que a jurisdição é exercida por ministros e equipes sem confiança cega na tecnologia.

O episódio produz duas consequências para a governança. A primeira é técnica: documentos processuais devem ser tratados como entradas potencialmente adversariais. Sistemas precisam isolar instruções do usuário, detectar conteúdo oculto, limitar permissões e registrar incidentes. A segunda é jurídica: o magistrado

não pode presumir que a informação sintetizada pelo sistema corresponde fielmente ao documento original sem mecanismos adequados de verificação.

O HC 1.094.270, noticiado pelo STJ em maio de 2026, revela risco complementar: referências erradas e trechos de julgados inexistentes em petição com indícios de uso de IA. A importância metodológica do caso é mostrar que "alucinação" não é risco abstrato. Em ambiente jurídico, erro generativo pode assumir formato de precedente, citação ou fundamento que parece verificável. A salvaguarda central é a conferência na fonte oficial.

Para o Judiciário, isso significa que sistemas de Nível 2 ou 3 não deveriam permitir incorporação de citações sem vínculo recuperável com a fonte. Idealmente, a interface deve separar texto gerado de texto extraído, indicar grau de confiança quando tecnicamente adequado, facilitar abertura do documento original e sinalizar quando determinada afirmação não possui fonte verificável. Não se trata de transformar recomendação técnica em requisito constitucional específico, mas de desenhar infraestrutura coerente com o dever jurídico de verificação.

A segurança adversarial reforça a tese do controle humano significativo. Uma revisão humana incapaz de enxergar as entradas, fontes e transformações do sistema não é plenamente capaz de detectar manipulação. A rastreabilidade, portanto, não é luxo administrativo: é condição para reconstruir incidentes que possam afetar a validade do processo.

## 9. DIREITO COMPARADO E GOVERNANÇA INTERNACIONAL: CONVERGÊNCIAS E LIMITES

O Regulamento Europeu de Inteligência Artificial adota abordagem baseada em risco e dedica atenção específica à administração da justiça. O Anexo III, item 8, alínea a, em conjunto com o art. 6º e o considerando 61, identifica como de alto risco determinados sistemas empregados por autoridades judiciais para pesquisar e interpretar fatos e direito e aplicar a lei a situações concretas, ressalvando atividades meramente auxiliares. A lógica é compatível com a distinção proposta pela MRAJ entre automação periférica e interferência no núcleo decisório.

O art. 14 do AI Act exige que sistemas de alto risco sejam projetados para permitir supervisão humana efetiva, proporcional aos riscos, ao nível de autonomia e ao contexto de uso. Entre as capacidades esperadas estão compreender limitações, permanecer atento ao risco de confiar excessivamente no output, interpretar resultados, desconsiderá-los, substituí-los ou interromper o sistema. Esse desenho regulatório demonstra que “humano presente” e “humano capaz de exercer controle” são categorias diferentes.

A Recomendação da UNESCO de 2021 acrescenta dimensão de direitos humanos. O documento associa transparência e explicabilidade à possibilidade de responsabilização e de contestação e afirma que a responsabilidade final não deve ser transferida aos sistemas de IA. A relevância para o processo judicial brasileiro está menos em sua força normativa interna e mais na convergência de princípios com a Constituição e com a Resolução do CNJ.

A Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial, Direitos Humanos, Democracia e Estado de Direito, adotada em 2024, também estrutura a governança a partir de riscos,

impactos, salvaguardas e remédios. Este artigo utiliza o instrumento exclusivamente como referência comparativa, sem pressupor sua aplicação vinculante ao Brasil.

A comparação revela uma tendência: aplicações públicas de alto impacto exigem mais que responsabilização ex post. Exigem avaliação prévia, documentação, supervisão, possibilidade de intervenção e mecanismos de recurso. No Judiciário brasileiro, esses elementos precisam ser traduzidos pelas categorias próprias do processo constitucional: juiz natural, contraditório, fundamentação, publicidade na medida juridicamente cabível e proteção de dados.

Por isso, a importação literal de categorias estrangeiras seria inadequada. O valor do direito comparado é funcional: ajudar a identificar problemas e soluções compatíveis com a arquitetura doméstica. A MRAJ não deriva do AI Act, embora dialogue com sua proporcionalidade baseada em risco; deriva da combinação entre o modelo regulatório do CNJ e as garantias da Constituição brasileira.

## 10. CONSEQUÊNCIAS PARA MAGISTRADOS, TRIBUNAIS, ADVOCACIA E PESQUISA JURÍDICA

Para magistrados, a principal consequência é substituir a lógica de confiança por uma lógica de verificação calibrada ao risco. Quanto mais o output participar da seleção de fatos, provas ou direito, menos suficiente é a revisão estilística da minuta. O magistrado deve poder identificar a origem das fontes, confrontar o resultado com os autos e registrar correções relevantes. O objetivo não é burocratizar a decisão, mas preservar autoria e responsabilidade.

Para tribunais, a governança precisa ser institucional. Devem existir políticas claras sobre quais ferramentas podem ser utilizadas, em

quais dados, para quais finalidades e sob qual nível de risco. Sistemas devem manter logs, versões, controles de acesso, avaliação de incidentes e documentação que permita auditoria. Treinamento deve incluir não apenas “como usar” a ferramenta, mas também vieses, alucinações, prompt injection, proteção de dados e automação da confiança.

Para a advocacia, os casos de 2026 reforçam dever elementar de conferência. A IA pode ser utilizada como suporte de pesquisa ou redação, mas o profissional responde pela peça que assina e deve verificar legislação, precedentes, fatos e citações em fontes oficiais. Conteúdo oculto destinado a manipular sistemas automatizados atinge a integridade do processo e não se confunde com legítima técnica de argumentação.

Para a pesquisa jurídica, surge agenda empírica importante. É necessário medir como magistrados interagem com recomendações de IA, em quais tarefas a supervisão realmente corrige erros, quais interfaces reduzem viés de automação e como diferentes modelos de transparência afetam capacidade de contestação. O direito não deve presumir eficácia da supervisão humana sem evidência sobre suas condições concretas.

Existe também uma agenda normativa. A regulamentação pode evoluir para definir critérios mais objetivos de divulgação quando IA generativa participar materialmente da fundamentação. Um modelo possível é exigir anotação processual padronizada para usos de Nível 2 e declaração específica para usos de Nível 3, sem impor divulgação desnecessária para automação administrativa ou simples correção textual. Essa diferenciação tornaria a transparência mais informativa.

Outra agenda diz respeito à prova de conformidade. Em eventual questionamento, a existência de registros institucionais poderia demonstrar qual versão do sistema foi utilizada, quais fontes foram recuperadas, se houve revisão humana e quais correções foram realizadas. Isso protege não apenas as partes, mas também o próprio magistrado e o tribunal contra alegações impossíveis de verificar por ausência de rastros.

O desafio final é cultural. A melhor governança não trata a IA como oráculo nem como inimiga da jurisdição. Trata-a como tecnologia de alto potencial e risco variável. A função do direito é construir condições para que o ganho de capacidade não se converta em perda de legitimidade.

## 11. CONCLUSÃO

A utilização de inteligência artificial generativa no Poder Judiciário exige mudança de foco. A questão constitucional relevante não é se um humano aparece formalmente no final do fluxo, mas se a decisão permanece materialmente atribuível a uma autoridade humana capaz de compreender, contestar, corrigir e justificar o uso da tecnologia.

A análise normativa sustenta a hipótese de que a supervisão humana formal é insuficiente quando a IA participa de maneira relevante da formação das premissas decisórias. Os arts. 5º, LIII, LIV e LV, e 93, IX, da Constituição, em conjunto com o padrão de fundamentação do art. 489 do CPC, exigem estrutura de controle que preserve competência, devido processo, contraditório e motivação. O art. 5º, LXXIX, e a LGPD acrescentam salvaguardas quando há tratamento de dados pessoais.

A Resolução CNJ nº 615/2025 constitui avanço expressivo ao adotar classificação de riscos, supervisão humana, auditabilidade, contestabilidade e regras próprias para IA generativa. Seu ponto de tensão está na coexistência entre registro interno obrigatório e menção externa facultativa do uso de IA na elaboração de atos judiciais. Este artigo sustenta, como interpretação constitucional proposta, que essa facultatividade é adequada para usos de baixa relevância, mas deve ser interpretada conforme a Constituição quando a IA fornece conteúdo materialmente incorporado à fundamentação.

Como contribuição, formulou-se o conceito de controle humano juridicamente significativo, composto por atribuição humana efetiva, rastreabilidade, contestabilidade e transparência proporcional ao risco. Propôs-se ainda a Matriz de Relevância Algorítmica Judicial, com quatro níveis - administrativo, redacional, cognitivo relevante e decisório sensível - aos quais correspondem salvaguardas progressivamente mais intensas. A matriz permite evitar soluções binárias e é apresentada como proposta interpretativa e regulatória, não como norma atualmente incorporada ao direito positivo.

Nem todo uso de IA precisa ser revelado na decisão; nem todo uso pode permanecer oculto sob a simples afirmação de que houve revisão humana. O elemento decisivo é a capacidade de a tecnologia alterar premissas ou resultado e, consequentemente, afetar o exercício de direitos processuais.

Casos recentes do STJ envolvendo precedentes inexistentes e prompt injection demonstram que a governança precisa incluir verificação de fontes e segurança adversarial. Esses episódios não

autorizam generalizações sobre decisões judiciais automatizadas, mas confirmam que sistemas generativos introduzem riscos específicos de integridade informacional que não podem ser resolvidos pela fluência textual do output.

A IA pode contribuir para uma jurisdição mais eficiente e informacionalmente capaz. A Constituição, contudo, impede que eficiência seja obtida pela invisibilização da origem das premissas decisórias ou pela transferência material da autoridade de julgar. O futuro juridicamente legítimo é o da jurisdição tecnologicamente aumentada e humanamente responsável: a máquina pode auxiliar a construir informações e propostas, mas a autoridade que transforma essas informações em decisão deve continuar sendo humana, identificável, contestável e constitucionalmente responsável.

## REFERÊNCIAS BIBLIOGRÁFICAS


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.

BENDER, Emily M.; GEBRU, Timnit; MCMILLAN-MAJOR, Angelina; SHMITCHELL, Shmargaret. On the dangers of stochastic parrots: can language models be too big? In: ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2021. **Proceedings [...].** New York: ACM, 2021. p. 610–623. DOI: https://doi.org/10.1145/3442188.3445922.

BRASIL. Câmara dos Deputados. **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, 2026. Disponível em:

https://www.camara.leg.br/proposicoesWeb/fichadetramitacao/?idProposicao=2487262. Acesso em: 9 out. 2026.

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. Disponível em: https://www.stj.jus.br/internet_docs/biblioteca/clippinglegislacao/Res_615_2025_CNJ.pdf. Acesso em: 9 out. 2026.

BRASIL. Conselho Nacional de Justiça. **Resolução nº 674, de 25 de março de 2026.** Altera o art. 15 da Resolução CNJ nº 615/2025. Brasília, DF: CNJ, 2026. Disponível em: https://www.stj.jus.br/internet_docs/biblioteca/clippinglegislacao/Res_674_2026_CNJ.pdf. Acesso em: 9 out. 2026.

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: 9 out. 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: 9 out. 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/l13709.htm. Acesso em: 9 out. 2026.

BRASIL. Superior Tribunal de Justiça. **Relator identifica habeas corpus com erros e alucinações criadas por IA e, após ouvir advogado, determina comunicação à OAB.** Brasília, DF, 19 maio 2026. Disponível em: https://www.stj.jus.br/sites/portalp/paginas/comunicacao/noticias/2026/19052026-relator-identifica-habeas-corpus-com-erros-e-alucinacoes-criadas-por-ia.aspx. Acesso em: 9 out. 2026.

BRASIL. Superior Tribunal de Justiça. **REsp 2.256.731/PB.** Relator: Min. Og Fernandes. Sexta Turma. Julgado em: 1° set. 2026. DJEN, 11 set. 2026. Disponível em: https://scon.stj.jus.br/SCON/GetInteiroTeorDoAcordao?num_registro=202600352871&dt_publicacao=11/09/2026. Acesso em: 9 out. 2026.

CITRON, Danielle Keats. Technological due process. **Washington University Law Review,** v. 85, n. 6, p. 1249–1313, 2008.

CONSELHO DA EUROPA. **Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law.** Strasbourg: Council of Europe, 2024. Disponível em: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence. Acesso em: 9 out. 2026.

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

JI, Ziwei et al. Survey of hallucination in natural language generation. **ACM Computing Surveys,** v. 55, n. 12, art. 248, p. 1–38, 2023. DOI: https://doi.org/10.1145/3571730.

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.

RE, Richard M.; SOLOW-NIEDERMAN, Alicia. Developing artificially intelligent justice. **Stanford Technology Law Review,** v. 22, n. 2, p. 242–289, 2019.

SELBST, Andrew D.; BOYD, Danah; FRIEDLER, Sorelle A.; VENKATASUBRAMANIAN, Suresh; VERTESI, Janet. Fairness and abstraction in sociotechnical systems. In: ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2019. **Proceedings [...].** New York: ACM, 2019. p. 59–68. DOI: https://doi.org/10.1145/3287560.3287598.

STRACHICINI, Douglas Lingiardi; ALENCAR, Alison Carvalho de. Transparência decisória na era algorítmica: a inteligência artificial explicável como requisito da motivação judicial. **Revista Rede de Direito Digital, Intelectual & Sociedade,** v. 6, n. 11, p. 279–308, 2026. DOI: https://doi.org/10.5380/rrddis.v6i11.101862.

SURDEN, Harry. Artificial intelligence and law: an overview. **Georgia State University Law Review,** v. 35, n. 4, p. 1305–1337, 2019.

TOLEDO, Cláudia; PESSOA, Daniel Alves. O uso de inteligência artificial na tomada de decisão judicial. **Revista de Investigações Constitucionais,** v. 10, n. 1, e237, 2023. DOI: https://doi.org/10.5380/rinc.v10i1.86319.

UNESCO. **Recommendation on the Ethics of Artificial Intelligence.** Paris: UNESCO, 2021. Disponível em:

https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence. Acesso em: 9 out. 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,** 2024. Disponível em: https://eur-lex.europa.eu/legal-content/pt/ALL/?uri=CELEX:32024R1689. Acesso em: 9 out. 2026.

VALLE, Vivian Cristina Lima López; FUENTES I GASÓ, Josep Ramón; AJUS, Attílio Martins. Decisão judicial assistida por inteligência artificial e o Sistema Victor do Supremo Tribunal Federal. **Revista de Investigações Constitucionais,** v. 10, n. 2, e252, 2023. DOI: https://doi.org/10.5380/rinc.v10i2.92598.

---

[1] Acadêmico de Direito do Centro Universitário de Várzea Grande (UNIVAG). ORCID: https://orcid.org/0009-0003-0059-1840. Endereço institucional: Av. Dom Orlando Chaves, 2655, Cristo Rei, Várzea Grande, MT, Brasil. E-mail: acesse o artigo original para visualizar o e-mail.

[2] A Resolução CNJ nº 674, de 25 de março de 2026, alterou o art. 15, § 1º, I, da Resolução nº 615/2025, relativo à composição do Comitê Nacional de Inteligência Artificial do Judiciário.

[3] Neste artigo, “alucinação” designa tecnicamente conteúdo sem suporte adequado na fonte ou factualmente incorreto; a expressão não atribui consciência ao sistema (Ji et al., 2023).

[4] Situação legislativa consultada em 9 de outubro de 2026 no portal da Câmara dos Deputados: aguardando parecer do relator na

comissão especial. O PL nº 2.338/2023 não é tratado como lei vigente.

[5] A MRAJ é uma formulação analítica e propositiva deste artigo. Não corresponde à classificação oficial de riscos do Anexo da Resolução CNJ nº 615/2025 nem autoriza usos vedados pela norma.