UFSM

InterAção

ISSN 2357-7975
ACESSO ABERTO

IA., Santa Maria, v. 17, n. 1, e95074, p. 1-25, mar. 2026 • https://dx.doi.org/10.5902/2357797595074
Submissão: 08/01/2026 • Aprovação: 21/02/2026 • Publicação: 20/03/2026

Artigos Publicação Contínua

Direito à explicação no cenário algorítmico: fundamentos e esboço de agenda para uma governança ética da IA

Right to explanation in the algorithmic landscape: foundations and an outline for an ethical AI governance agenda

Otávio Morato[I]
Dierle José Coelho Nunes[II]
Viviane Ramone Tavares[III]

[I]Centro Universitário de Sete Lagoas, Sete Lagoas, MG, Brasil
[II]Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
[III]Mestranda em Compliance pela Ambra University, Orlando, FL, Estados Unidos da América

Resumo

Este artigo investiga o direito à explicação enquanto resposta normativa e estratégica para mitigar a opacidade da inteligência artificial (IA), infraestrutura invisível já consolidada em decisões de alto impacto. Mostra-se como a opacidade algorítmica amplifica riscos (vieses, "alucinações" e falhas) e propõe-se o direito à explicação como mecanismo normativo relevante para assegurar transparência e responsabilidade em decisões automatizadas. Nesse quadro, a inteligência artificial explicável (XAI) é apresentada enquanto abordagem técnica que contribui para viabilizar tal garantia, cuja consolidação como princípio estruturante é demonstrada pela análise de marcos regulatórios. Conclui-se que o direito à explicação é pilar para uma governança ética e democrática da IA, esboçando-se, ao final, uma agenda para a implementação do direito à explicação no Brasil, de modo a compatibilizar a inovação tecnológica com os princípios do Estado Democrático de Direito.

Palavras-chave: Direito à explicação; Opacidade algorítmica; Inteligência artificial explicável (XAI); Governança de IA; Caixa-preta algorítmica

Abstract

This article investigates the right to explanation as a normative and strategic response to mitigate the opacity of artificial intelligence (AI), an invisible infrastructure already embedded in high-impact decision-making. It shows how algorithmic opacity amplifies risks (such as biases, "hallucinations," and failures) and proposes the right to explanation as a relevant normative mechanism to ensure transparency and accountability in automated decisions. In this context, explainable artificial intelligence (XAI) is presented as a technical approach that helps make this guarantee feasible, whose consolidation as a structuring principle is demonstrated through the analysis of regulatory frameworks. The article concludes that the right to explanation is a cornerstone of ethical and democratic AI governance and, finally, outlines an agenda for implementing the right to explanation in Brazil, in a way that reconciles technological innovation with the principles of the Democratic Rule of Law.

Keywords: Right to explanation; Algorithmic opacity; Explainable artificial intelligence (XAI); AI governance; Algorithmic black box

INTRODUÇÃO

A inteligência artificial (IA)[I] deixou de ser uma promessa futurista e consolidou-se como uma tecnologia presente em múltiplas esferas da vida contemporânea. Nos dias atuais, robôs avançados já operam em ambientes reais, enquanto ferramentas outrora consideradas ficção científica, como reconhecimento facial e veículos autônomos, permeiam nosso cotidiano. Para além dos dispositivos aparentes, os algoritmos de IA compõem uma infraestrutura invisível que permeia atividades rotineiras e decisões de alto impacto: avaliam currículos, redigem contratos, controlam câmeras, definem investimentos e até mesmo auxiliam no diagnóstico de doenças. Neste sentido, modelos algorítmicos passaram a modular relações sociais, econômicas e políticas, influenciando desde o entretenimento e a segurança pública até os mercados financeiros e assistentes virtuais (Morato, 2022).

A presença ubíqua da IA, contudo, traz o problema da *opacidade*: parte significativa dos modelos algorítmicos tem funcionamento inescrutável – e o que se conhece são apenas as saídas *(outputs)* produzidas pelo sistema. Nos raros casos em que códigos-fonte são acessíveis, sua interpretação requer conhecimentos avançados

[I] IA pode ser definida como um sistema computacional projetado para operar com diferentes níveis de autonomia e capaz de reagir às variáveis externas, se adaptando ou "aprendendo" com elas após a sua implementação. Isso é potencializado pelo *deep learning*, que usa redes neurais para processar dados complexos, como imagens e linguagem natural. Como veremos adiante, nem todo modelo algorítmico gera opacidade (Morato e Nunes, 2025).

de programação, limitando a supervisão pública dessas tecnologias. A ascensão das IAs generativas, que criam textos, imagens e outros conteúdos aplicando raciocínios complexos para analisar vastos bancos de dados, acentuou esse desafio.

Neste sentido, a *"caixa-preta algorítmica"*[II] levanta questões cruciais: *quais critérios levam um sistema a classificar alguém como suspeito de atividades criminosas? Como verificar se uma decisão automatizada é correta ou justa? Quais mecanismos asseguram que esses sistemas não reproduzam preconceitos históricos ou reforcem desigualdades estruturais?* Mesmo modelos avançados de linguagem natural, como ChatGPT, Gemini ou Claude, que interagem de forma aparentemente transparente, operam sob lógicas quase sempre obscuras para seus usuários – opacidade que se agrava quando o sistema é de *software fechado* (Novaes e Morato, 2025).

A explicabilidade, aqui entendida como a exigência de demonstrar, de forma acessível, como e por que a decisão foi produzida, revela-se uma questão que abrange simultaneamente as dimensões técnica, ética e jurídica. A partir dessa demanda surge a *eXplainable Artificial Intelligence - XAI*[III] (em português: inteligência artificial explicável), que oferece os métodos técnicos capazes de transformar o princípio da explicabilidade em prática concreta, por meio de textos, gráficos, fatores determinantes e exemplos que mostram como o sistema chegou ao resultado (Reigeluth *et al.*, 2025). O *direito à explicação*, por sua vez, institui o dever jurídico de tornar compreensíveis as decisões automatizadas, fazendo com que os mecanismos de XAI resultem em interpretações claras e efetivamente compreensíveis ao usuário.

Diante deste cenário, o presente artigo investiga a estratégia de *explicar* – mais do que simplesmente oferecer transparência – como enfrentamento à opacidade algorítmica. Parte-se da hipótese de que, em contextos decisórios de alto impacto individual ou coletivo, a explicabilidade torna as decisões automatizadas mais legítimas, na medida em que eleva a confiança do destinatário, favorece a identificação e a

[II] Para Frank Pasquale (2015, p. 3), a caixa-preta simboliza tecnologias que, embora registrem dados como os dispositivos de monitoramento de aviões, operam de modo enigmático, tornando visíveis apenas os resultados, e não o percurso decisório que os produz.

[III] A XAI é um conceito técnico universal, não devendo ser confundido com a empresa homônima de Elon Musk, que usa "xAI" como marca comercial.

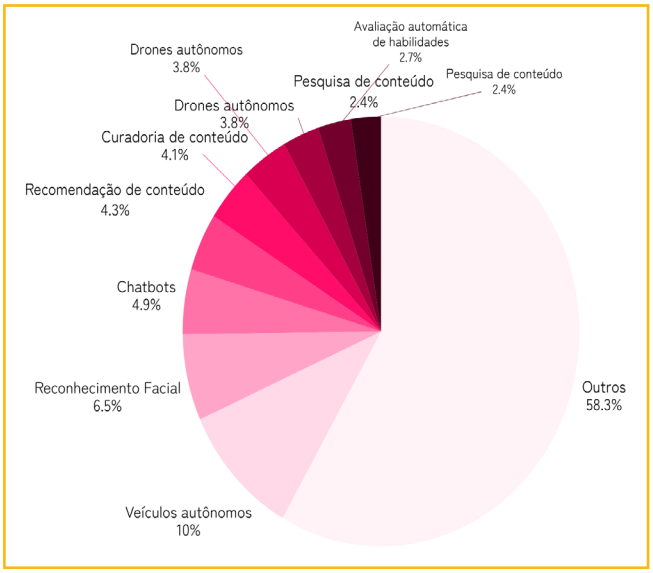
correção de falhas e vieses, viabiliza a contestação substancial e fortalece a atribuição de responsabilidade, reduzindo o risco de violações de direitos. Contudo, não se pretende idealizar uma solução totalizante ou definitiva, sendo o direito à explicação apenas um elemento entre um conjunto mais amplo de medidas regulatórias possíveis neste campo.

Trata-se de uma pesquisa *qualitativa*, *exploratória* e de caráter *hermenêutico-analítico*, orientada pela leitura crítica de marcos legais, diretrizes técnicas e literatura multidisciplinar. A investigação adota um procedimento predominantemente *indutivo*, extraindo categorias a partir da análise de regimes regulatórios nacionais e estrangeiros, em exame de princípios e normas jurídicas relacionadas à proteção de direitos e à responsabilização tecnológica. Para tanto, o percurso metodológico organiza-se em três momentos: primeiro, o mapeamento conceitual dos desafios da opacidade; em seguida, a discussão dos fundamentos jurídicos do direito à explicação; e, por fim, o esboço de diretrizes em uma agenda de implementação prática.

ALUCINAÇÕES, VIESES ALGORÍTMICOS E OS RISCOS DA OPACIDADE

Por mais sofisticados que sejam, os modelos algorítmicos estão longe de ser infalíveis. Respostas incompletas ou erradas (as chamadas “alucinações”), previsões pouco confiáveis e decisões enviesadas aparecem com frequência, por razões que combinam dados, contexto, infraestrutura e design. Este capítulo sintetiza as principais falhas algorítmicas, destacando como sua opacidade transforma erros técnicos em riscos sociais e jurídicos. É importante frisar que *nem toda falha gera prejuízo humano direto*: muitas são detectadas em ambiente de testes ou aparecem em usos triviais, como uma resposta fantasiosa de um assistente virtual. Outras, porém, têm efeitos tangíveis e graves, como acidentes com veículos autônomos ou discriminação em sistemas de reconhecimento facial. O *AI Incident Database* (2025), repositório que indexa danos e quase-danos decorrentes de IA no mundo real, já catalogou mais de mil intecorrências, sendo 233 apenas em 2024 – um recorde anual, com aumento de 56,4% em relação a 2023.

Figura 1 – Falhas da IA por categoria



Fonte: Adaptado de AI Incident Database (2026)

Deve ser ressaltado que um mesmo sistema pode apresentar *falhas sobrepostas* e, ainda, que a IA raramente opera isolada: ela integra um ecossistema técnico, social e econômico que condiciona seu desempenho (e em alguns casos, dificulta a identificação acurada do problema). Quando a estrutura interna do modelo é opaca – uma verdadeira *caixa-preta algorítmica* –, torna-se quase impossível auditar, contestar ou mesmo compreender *como* e *por que* o resultado foi produzido (Morato e Nunes, 2025; Pasquale, 2015).

No livro *Inteligência artificial: o desafio da explicabilidade* (Morato e Nunes, 2025), as falhas algorítmicas são classificadas em quatro categorias. A primeira refere-se aos *dados de entrada*, quando bases insuficientes ou não representativas introduzem distorções e reproduzem desigualdades históricas[IV], como ocorre em sistemas treinados majoritariamente com perfis demográficos homogêneos, que tendem a apresentar vieses sistemáticos na classificação de grupos minoritários.

A segunda envolve *limitações operacionais*, nas quais gargalos de infraestrutura, versões desatualizadas ou integrações precárias degradam o desempenho de

[IV] A qualidade dos dados, conforme delineado no artigo *The six principles of AI-ready data,* é intrinsecamente ligada à explicabilidade, pois dados de baixa qualidade podem introduzir opacidade e dificultar a rastreabilidade das decisões do modelo (Qlik, 2024).

sistemas considerados tecnicamente robustos, a exemplo de modelos que operam corretamente em ambientes controlados, mas falham quando implantados em sistemas legados ou sob restrições severas de processamento e latência.

A terceira diz respeito à *falta de adequação ao contexto*, que surge quando o modelo tem dificuldades em extrapolar o domínio para o qual foi concebido, como em aplicações que transferem sistemas treinados em contextos estrangeiros para realidades institucionais, culturais ou normativas distintas, sem adaptação suficiente.

A quarta e última categoria[v] decorre de *design ou treinamento inadequados*, quando escolhas técnicas (*e.g.:* seleção errônea de arquiteturas, hiperparâmetros, critérios de validação, etc.) produzem modelos ineficientes ou instáveis, comprometendo o desempenho global do sistema, como nos casos em que a otimização excessiva para acurácia sacrifica a generalização ou amplifica comportamentos erráticos em situações não previstas no treinamento.

Figura 2 – Categorias mais comuns de falhas algorítmicas

Fonte: Autores (2026)

[v] A classificação não pretende esgotar todas as causas possíveis de falhas algorítmicas, mas apenas organizar as mais recorrentes ao público não especializado. Certos vieses, por exemplo, podem emergir de propriedades estatísticas do próprio algoritmo de treinamento e da forma como ele explora o espaço probabilístico, escapando a essa taxonomia.

Entre as falhas mais notáveis estão as *alucinações algorítmicas em sistemas generativos*. Esses modelos constroem suas respostas *token* a *token*[VI], prevendo a cada passo quais expressões linguísticas são mais prováveis de aparecer dadas as anteriores. A ausência de um mecanismo interno de checagem factual faz com que preencham lacunas com informações apenas verossímeis, mas incorretas, inventando autores, jurisprudências e até locais fictícios. Em domínios sensíveis, como justiça e saúde, essa opacidade dificulta identificar a origem do erro e atribuir responsabilidade – demandando supervisão humana e ferramentas externas de verificação, para reduzir as chances de alucinação.

Além das falhas, identificam-se também *vieses algorítmicos*. Enquanto as primeiras tendem a refletir problemas operacionais ou de desempenho, os últimos expressam assimetrias presentes nos dados (por exemplo, sub-representação de pessoas negras ou de mulheres em determinado conjunto de dados) e nas decisões humanas incorporadas ao modelo. Em ambientes de alto risco (saúde, justiça, crédito, segurança, *etc.*), confiar na caixa-preta e tratar resultados como verdades inquestionáveis transforma erros computacionais em decisões automáticas, reforçando desigualdades e obscurecendo responsabilidades.

Em plataformas digitais, vieses históricos tornaram-se emblemáticos: do Google Fotos rotulando pessoas negras como "gorilas" à opção de simplesmente bloquear termos em vez de corrigir o modelo; do chatbot Tay[VII], que em horas passou a reproduzir discurso de ódio; ao COMPAS[VIII], cuja pontuação opaca e discriminatória influenciou decisões judiciais sem possibilidade de contestação; ao Apple Card[IX], escrutinado por suposta discriminação de gênero na concessão de crédito; e, mais

[VI] *Token* é a unidade elementar processada pelo modelo de linguagem (como palavras, pedaços de palavras ou sinais de pontuação). A geração de saída *token a token* significa que o sistema escolhe sucessivamente o próximo *token* mais provável, condicionando cada escolha ao contexto dos *tokens* anteriores (*c.f.* Nunes e Morato, 2025).

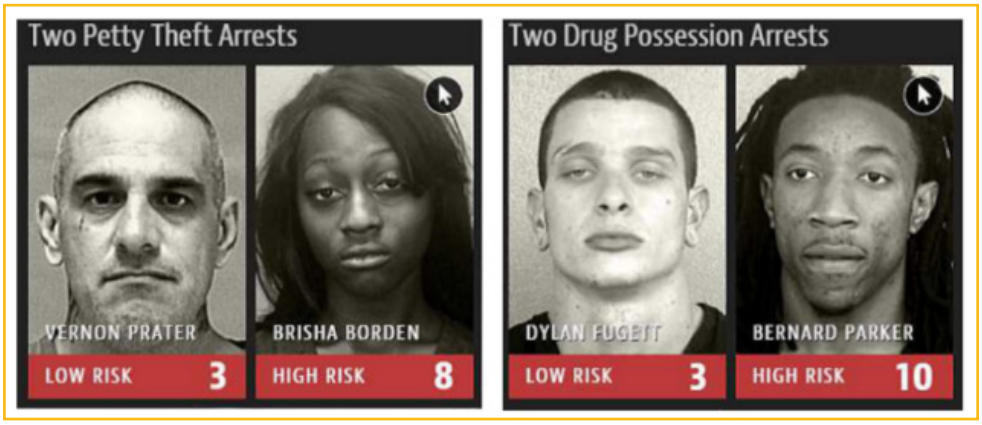
[VII] Lançado pela Microsoft no Twitter em 2016, Tay era um chatbot de IA projetado para aprender com as interações dos usuários. Em menos de 24 horas, foi desativado após ser manipulado por usuários que o ensinaram a reproduzir teorias conspiratórias, racismo e discursos contra minorias, tornando-se um caso emblemático dos riscos do aprendizado de máquina não supervisionado (Cano, 2016).

[VIII] Em 2016, a ProPublica, organização independente de jornalismo investigativo, analisou o COMPAS, um software comercial de avaliação de risco criminal utilizado por tribunais norte-americanos para estimar a probabilidade de reincidência. O estudo revelou que o sistema atribuía pontuações de "alto risco" a réus negros com o dobro de frequência em relação a réus brancos em situações semelhantes, mesmo quando estes últimos tinham antecedentes mais graves. Embora a ferramenta alegasse neutralidade técnica, a investigação mostrou vieses discriminatórios operando de forma opaca e não auditável (Angwin *et al.*, 2016).

[IX] Em 2019, o Apple Card, operado pelo Goldman Sachs, foi alvo de intensa controvérsia após alegações de que seu algoritmo concedia limites de crédito significativamente maiores para homens do que para suas esposas, mesmo com finanças compartilhadas (Knight, 2019).

recentemente, a geradores de imagem que produziram *outputs* racistas, sexualizados ou historicamente distorcidos[X].

Figura 3 – No caso COMPAS, a ferramenta atribuiu baixo risco a réus brancos e alto risco a réus negros, mesmo diante de delitos semelhantes



Fonte: Angwin *et al.* (2016)

Casos concretos sugerem que falhas técnicas e vieses de treinamento costumam andar juntos. No diagnóstico médico, por exemplo, estudos recentes indicam que a IA acerta com frequência nos casos mais comuns, mas erra muito mais nos quadros complexos, raros ou atípicos, justamente porque esses pacientes aparecem menos e pior representados nas bases de dados (Degrave, Janizek e Lee, 2021; Duarte *el al.*, 2025). A falha não é "neutra": ela pesa mais sobre quem foge do perfil padrão aprendido pelo modelo, o que é um tipo de viés. Em veículos autônomos, investigações sobre centenas de acidentes mostram algo semelhante: os sistemas funcionam bem em pistas retas, tempo bom e sinalização clara, mas falham com muito mais frequência sob chuva leve, reflexos solares e em cruzamentos, cenários pouco contemplados ou mal modelados no treinamento (Zhang *et al.*, 2023). De novo, o erro técnico se concentra em determinadas circunstâncias, revelando um viés projetual que expõe algumas situações e algumas pessoas a riscos maiores.

[X] Nunes, Dierle; Morato, Otávio. Inteligência artificial: o desafio da explicabilidade. Salvador: Juspodivm, 2025.

Há também os “acertos enganosos”: o sistema performa bem nos testes, mas pelos motivos errados, apoiado em pistas espúrias. A analogia clássica é a do cavalo Clever Hans (Anders *et al.*, 2022), que “acertava” matemática lendo micro-sinais do treinador. Em visão computacional, um classificador de “lobos vs. huskies” aprendeu a neve do fundo, não as feições do animal (Alves e Morato, 2023). Sem técnicas de interpretabilidade e auditoria, a ilusão permanece, e, em uma caixa-preta, torna-se impossível detectar se o acerto reflete entendimento genuíno ou mera coincidência estatística.

Diante de todos esses problemas, a *opacidade* – barreira para entender como o sistema decide – funciona como um multiplicador de riscos. Ela mascara a origem dos erros, dificulta auditorias e escala problemas pontuais para níveis sistêmicos. Sem visibilidade sobre o processo decisório, falhas e vieses tornam-se indetectáveis, e decisões críticas passam a operar em um regime de confiança cega, inadequado para ambientes de alto impacto. As causas da opacidade são variadas: podem decorrer de sigilo intencional, de limitações na formação técnica dos usuários ou da própria complexidade dos modelos. Em sistemas mais avançados, mesmo o acesso ao código-fonte e a capacidade de interpretá-lo não eliminam a opacidade, pois a altíssima densidade das redes neurais dificulta a compreensão, o questionamento e a contestação do percurso que conduz ao resultado final.

Assim, a governança não pode ser reduzida à otimização de performance ou à melhoria incremental de modelos: ela deve enfrentar a assimetria entre racionalidade algorítmica e racionalidade jurídica. Compreender a opacidade algorítmica é passo essencial para desenvolver regulamentações e práticas que tornem os sistemas mais explicáveis, seguros e justos. Tal compreensão técnica dos riscos é o pressuposto para se construir o arcabouço jurídico de proteção que será analisado a seguir.

EXPLICABILIDADE E FUNDAMENTOS DO DIREITO À EXPLICAÇÃO

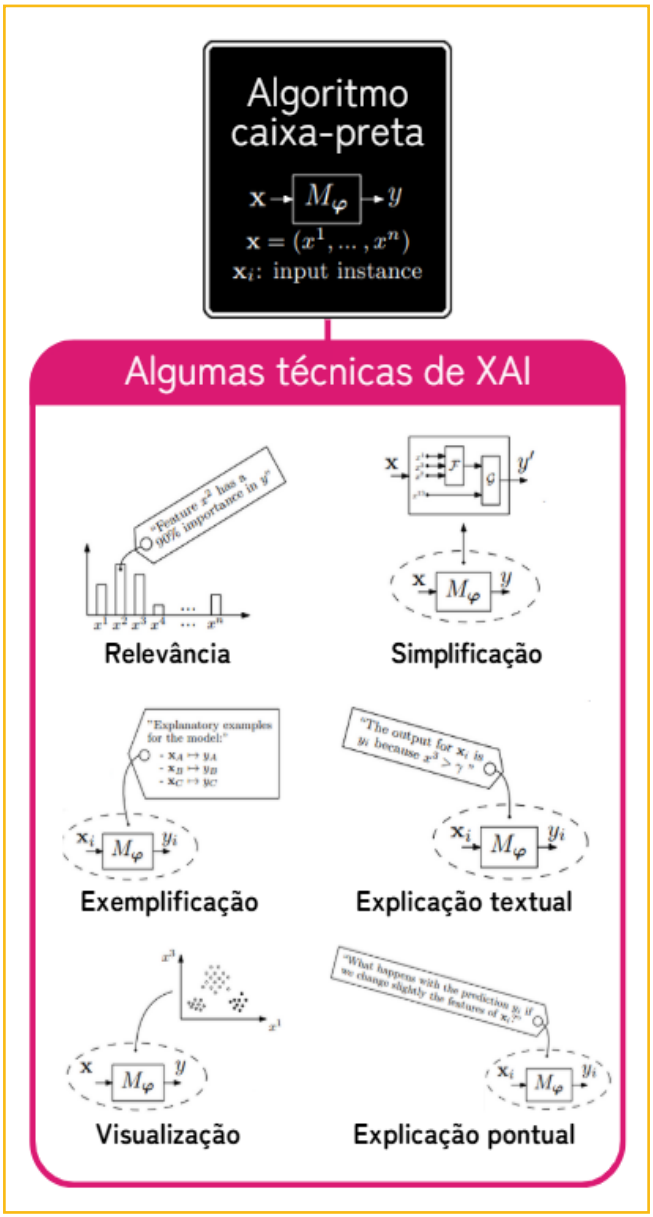
A transparência diz respeito ao acesso a informações sobre o sistema, como código-fonte, dados de treinamento utilizados ou sua arquitetura geral. O mero acesso a esses componentes técnicos, porém, é insuficiente para um controle social efetivo, pois não garante a compreensão da lógica decisória. Isso vale, sobretudo, para modelos opacos cujo funcionamento interno não é diretamente interpretável – como é o caso das redes neurais[XI]. A *explicabilidade* é um passo além: trata-se de traduzir o funcionamento interno de um modelo em termos compreensíveis para um ser humano, respondendo ao "porquê" de uma decisão específica (Reigeluth *et al.*, 2025). Essa diferença é crucial, pois a garantia em jogo não é apenas o acesso a informações, mas o entendimento da lógica por trás de um *output* que afeta a vida do indivíduo, viabilizando a contestação e a responsabilização.

A natureza de uma explicação eficaz depende de seu *público-alvo*. Para o cidadão afetado por uma decisão automatizada, a explicação deve ser clara, contextual e focada nos fatores decisivos, viabilizando seu direito de contestação. Para um auditor ou regulador, a demanda é por uma explicação técnica e auditável, que permita a verificação de conformidade e a detecção de vieses. Já para o desenvolvedor, a explicabilidade funciona como uma ferramenta de depuração e aprimoramento (Morato e Nunes, 2025). Essa multifuncionalidade revela os propósitos centrais da explicabilidade: construir confiança, promover justiça, garantir a responsabilização e assegurar a contestabilidade das decisões algorítmicas.

Se, para o empresário, a explicabilidade assegura a melhoria da conformidade legal, reduzindo os riscos jurídicos associados ao descumprimento de normas regulatórias, há, ainda, um acréscimo na confiança dos funcionários e clientes quanto ao entendimento dos processos automatizados. Principalmente quando se analisa sob o aspecto do grau de lesividade aos direitos fundamentais da tecnologia utilizada.

[XI] Em modelos clássicos, como regressões lineares ou árvores de decisão, o acesso ao código pode efetivamente revelar a lógica do raciocínio algorítmico. Todavia, modelos avançados de redes neurais, aplicados em escala cada vez maior, colocam o problema central da explicabilidade.

Figura 4 – Representação esquemática das principais abordagens de explicação aplicáveis a modelos algorítmicos opacos, incluindo relevância, simplificação, exemplificação, explicação textual, visualização e explicação pontual



Fonte: Adaptado de Arrieta *et al.* (2020)

O debate contemporâneo sobre inteligência artificial (IA) evidencia a necessidade de mecanismos que permitam aos indivíduos compreender as decisões produzidas por sistemas automatizados, especialmente quando estas decisões têm impacto significativo sobre direitos fundamentais. A esse conjunto de garantias convencionou-

se chamar de *direito à explicação*, que fórmula, em termos jurídicos, a demanda por inteligibilidade mínima diante de sistemas tecnicamente sofisticados, porém politicamente assimétricos e socialmente opacos.

Mais do que uma resposta técnica à opacidade, pode-se dizer que o direito à explicação decorre do próprio desenho constitucional brasileiro. Ao lado do *devido processo legal* (art. 5º, LIV), da *autodeterminação informativa* (LGPD) e da *dignidade da pessoa humana*, a explicação opera como garantia procedimental que assegura ao cidadão não apenas acesso a dados, mas capacidade real de compreender os critérios que orientaram uma decisão automatizada e, a partir disso, exercer contraditório e contestação substancial. Trata-se, portanto, de um direito epistêmico, que reivindica inteligibilidade suficiente para restaurar a agência humana diante do poder algorítmico (Brasil, 2018).

Não se pode olvidar da análise do contraditório, principalmente quando se relaciona com o "hodierno perfil poliédrico do contraditório" (Vale e Pereira, 2023), que amplia a possibilidade do questionamento da parte quanto à extensão e natureza do uso da tecnologia no impacto de seus interesses.

O desenvolvimento do direito à explicação e o aperfeiçoamento das técnicas de explicabilidade podem trazer benefícios amplos. Para indivíduos e coletividades, tais garantias estimulam a criação de sistemas mais compreensíveis, auditáveis e confiáveis, contribuindo para a legitimidade das decisões automatizadas (Nunes e Morato, 2023). Para desenvolvedores e empresas, a explicabilidade favorece a otimização dos próprios modelos, possibilitando a correção de erros, a detecção de vieses e o aprimoramento contínuo de funcionalidades.

Os sistemas de IA podem ser classificados conforme seu potencial de impacto sobre direitos fundamentais: os de *alto risco* exigem explicações robustas, compreensíveis e auditáveis para garantir transparência e controle social; os de *baixo risco*, com menor impacto individual ou coletivo, podem oferecer explicações simplificadas, desde que observem princípios de boa-fé, segurança e informação

adequada; já os de *risco excessivo*, cuja opacidade ou finalidade é intrinsecamente perigosa ou discriminatória, como nas tecnologias sem revisão humana ou que fazem inferências biométricas para prever condutas criminosas, devem ser proibidos pelo ordenamento jurídico (CNJ, 2025).

Legislação internacional

A análise dos marcos regulatórios estrangeiros que se segue não se pretende inventário descritivo, mas instrumento crítico para delinear um parâmetro constitucional brasileiro de direito à explicação, capaz de orientar a formulação normativa nacional com densidade garantista.

Na União Europeia, o direito à explicação consolidou-se a partir de dois marcos: o Regulamento Geral sobre a Proteção de Dados - RGPD (União Europeia, 2016), em vigor desde 2018, e o Regulamento Europeu de Inteligência Artificial - *AI Act* (União Europeia, 2024), aprovado em 2024. O RGPD assegura, em seu artigo 22, que indivíduos não sejam submetidos a decisões exclusivamente automatizadas com efeitos jurídicos ou impacto significativo, prevendo ainda a possibilidade de revisão humana e de acesso a informações claras sobre os critérios e a lógica utilizados no processo decisório[XII]. Em 2025, o Tribunal de Justiça da União Europeia reforçou essa proteção ao julgar o caso *Dun & Bradstreet Austria (C-203/22)*, reconhecendo o direito de acesso aos principais fatores que influenciaram decisões algorítmicas e exigindo ponderação entre transparência e proteção de interesses empresariais[XIII] (Gdprhub, 2025). Apesar disso, o RGPD ainda se mostra limitado, pois se aplica apenas a decisões inteiramente automatizadas e não cobre dados não pessoais, além de depender da iniciativa do titular.

O *AI Act* surge para superar essas lacunas, adotando uma abordagem baseada no risco: sistemas classificados como de "alto risco" ficam sujeitos a exigências

[XII] Considerando 71 e artigos 13, 14 e 15 do RGPD detalham esses direitos e orientam que as explicações sejam compreensíveis e contextualizadas.

[XIII] Essa decisão estabeleceu um precedente importante ao invalidar normas nacionais que criavam exceções automáticas em nome do sigilo corporativo, equilibrando proteção empresarial e direito à informação.

rigorosas de explicabilidade e supervisão humana. Seu artigo 86 assegura que qualquer pessoa afetada por tais sistemas receba informações claras sobre o papel desempenhado pela IA e os critérios que embasaram a decisão, enquanto os artigos 13 e 14 estabelecem medidas técnicas que permitem a correção de falhas e a rejeição de resultados automatizados[XIV] (União Europeia, 2024). Inspirado pelas Diretrizes Éticas para uma Inteligência Artificial Confiável, publicadas em 2019, o *AI Act* transformou a explicabilidade em requisito jurídico e operacional para a governança democrática da inteligência artificial na Europa.

No Reino Unido, a explicabilidade é assegurada pelo UK GDPR, versão doméstica do RGPD aprovada após o Brexit[XV]. Esse regulamento manteve a estrutura europeia, garantindo que decisões exclusivamente automatizadas com efeitos jurídicos relevantes sejam acompanhadas de justificativas claras e acessíveis. O *Information Commissioner's Office* (ICO) publica orientações para que organizações forneçam explicações compreensíveis e contextualizadas, enquanto a Estratégia Nacional de Inteligência Artificial enfatiza a importância da auditabilidade em setores sensíveis, como saúde, segurança pública e serviços financeiros (ICO, 2020).

Nos Estados Unidos, a regulação da explicabilidade é fragmentada, pois não há uma lei federal abrangente sobre inteligência artificial. Algumas normas setoriais já impõem obrigações específicas: a *Equal Credit Opportunity Act* exige que instituições financeiras expliquem os motivos da negativa de crédito, detalhando os fatores decisórios, enquanto a *Fair Credit Reporting Act* garante aos consumidores o direito de compreender e contestar decisões baseadas em seus históricos financeiros. Em nível estadual, medidas recentes ampliaram a transparência, como a *Local Law 144*, de Nova York, que impõe auditorias de viés em sistemas de recrutamento automatizados, e a *Senate Bill 21-169*, do Colorado, que proíbe algoritmos discriminatórios em seguros e serviços financeiros. Em 2022, a Casa Branca publicou o *Blueprint for an AI Bill of Rights*, documento não vinculante que reforça a importância da explicabilidade e da

[XIV] Essas obrigações foram diretamente influenciadas pelas Diretrizes Éticas de 2019, que definiram sete requisitos fundamentais para uma IA confiável e introduziram a ferramenta ALTAI para autoavaliação de sistemas.

[XV] Processo político e jurídico pelo qual o Reino Unido deixou a União Europeia, alterando sua vinculação a normas comunitárias.

governança de sistemas automatizados (Estados Unidos da América, 2022).

No Japão, a reforma de 2022 da Lei de Proteção de Informações Pessoais endureceu as regras para decisões automatizadas, obrigando operadores a informar claramente os objetivos e critérios utilizados sempre que tais decisões possam gerar efeitos significativos. Essa reforma também ampliou os poderes da Comissão Nacional de Proteção de Dados, que passou a emitir ordens vinculantes, cujo descumprimento pode gerar sanções penais, como detenção de até seis meses. Ainda que não reconheça a explicabilidade como direito individual, a legislação japonesa a incorporou como princípio ético, promovendo confiança social na adoção de tecnologias digitais.

Na Índia, a *Digital Personal Data Protection Act* (DPDP Act), aprovada em 2023, estabeleceu regras de transparência para decisões automatizadas, exigindo que empresas notifiquem sempre que processamentos envolvendo dados pessoais forem conduzidos por inteligência artificial. A lei também prevê mecanismos de governança algorítmica para prevenir decisões arbitrárias ou discriminatórias, constituindo um passo inicial na consolidação futura de um direito à explicação, ainda dependente de regulamentação detalhada.

No Canadá, o esforço para a regulação da transparência e governança da inteligência artificial sofreu um revés com o arquivamento do Projeto de Lei C-27 no início de 2025, o que resultou na caducidade do *Artificial Intelligence and Data Act - AIDA* antes de sua conversão em lei. O texto propunha mecanismos de supervisão humana e auditorias para sistemas de "alto impacto", visando reduzir a opacidade algorítmica e permitir a contestação de decisões automatizadas. Contudo, a interrupção do processo legislativo deixou o país em um estado de dependência de códigos de conduta voluntários e legislações de privacidade preexistentes. Mesmo como modelo referencial, o AIDA sinalizava uma estratégia preventiva focada na mitigação de riscos sistêmicos, ainda que não reconhecesse, de forma explícita, um direito subjetivo à explicação (Onetrust Dataguidance, 2025).

Avanços no Brasil

No Brasil, o arcabouço normativo existente já aborda a explicabilidade em diferentes frentes. A primeira referência direta surgiu na Resolução CNJ nº 332/2020, que determinou que decisões automatizadas no Judiciário sejam acompanhadas de "explicação satisfatória e passível de auditoria por autoridade humana". Mais recentemente, em 2025, a Resolução do CNJ nº 615, que incorporou a explicabilidade como um princípio basilar para o desenvolvimento, governança, auditoria, monitoramento e uso responsável de soluções de IA no Poder Judiciário, reforçou essas diretrizes ao prever a descontinuidade de sistemas com baixa explicabilidade (CNJ, 2025). A normativa traz uma exigência de explicabilidade que transcende a mera transparência superficial, de modo que suas decisões e operações sejam compreensíveis e auditáveis pelos operadores. Adicionalmente, a Lei Geral de Proteção de Dados (LGPD) introduziu instrumentos de supervisão, como o Relatório de Impacto à Proteção de Dados (RIPD), que, em cenários de alto risco, exige a documentação da lógica decisória para fornecer um nível mínimo de inteligibilidade e prestação de contas (Brasil, 2018; CNJ, 2020).

Paralelamente, há importantes propostas legislativas em debate que buscam consolidar o direito à explicação. O Projeto de Lei nº 2.338, de 2023 (Marco da IA), incorpora a explicabilidade como princípio e define os elementos mínimos das explicações. De forma semelhante, o Projeto de Lei do Senado nº 4, de 2025 (de atualização do Código Civil), propõe a criação de um Livro de Direito Civil Digital, posicionando a explicabilidade como elemento central para a validade de decisões automatizadas (Brasil, 2023; Brasil, 2025).

O documento ISO (2025) sobre padrões internacionais para IA responsável sublinha a importância de frameworks de governança que integrem considerações éticas, culturais e sociais, destacando que a transparência e a interpretabilidade são essenciais para o alinhamento ético e a conformidade regulatória. Assim, a XAI não é

meramente um requisito técnico, mas uma demanda sociojurídica que visa assegurar a responsabilidade e a equidade no ecossistema da inteligência artificial.

Tanto as normas existentes quanto os projetos em debate convergem, portanto, para uma abordagem baseada em risco, que tem no Marco da IA sua expressão mais clara. Nesse modelo, a intensidade do dever de explicação é proporcional ao impacto da decisão automatizada, exigindo-se justificativas robustas e auditáveis para sistemas de alto risco, enquanto contextos de menor impacto demandam requisitos mais brandos de transparência. Demonstrado assim que a regulamentação visa majoritariamente o enfrentamento da opacidade algorítmica, transformando necessidade social de justiça em requisitos técnicos de transparência e interpretabilidade.

Com os fundamentos jurídicos e o panorama regulatório estabelecidos, torna-se imperativo traduzir essa base teórica em uma agenda de implementação exequível para o contexto brasileiro. A partir desse diagnóstico, é possível esboçar uma agenda inicial para o contexto brasileiro, orientada à conversão de princípios em procedimentos concretos e instrumentos de contestação, como se propõe na próxima seção.

POR UMA AGENDA NACIONAL DO DIREITO À EXPLICAÇÃO

Como mostramos, a IA não se limita a uma conquista técnica ou a uma tecnologia de propósito geral: trata-se de uma infraestrutura normativa e política, capaz de modular comportamentos e redistribuir poder, tensionando categorias jurídicas tradicionais. Longe de serem neutros, os algoritmos incorporam vieses, apresentam erros e frequentemente produzem resultados que afetam diretamente indivíduos e coletividades, problemas potencializados pela opacidade que dificulta a fiscalização e a responsabilização. A superação desse quadro demanda uma "virada explicativa", isto é, a transformação das caixas-pretas em sistemas compreensíveis por diferentes públicos.

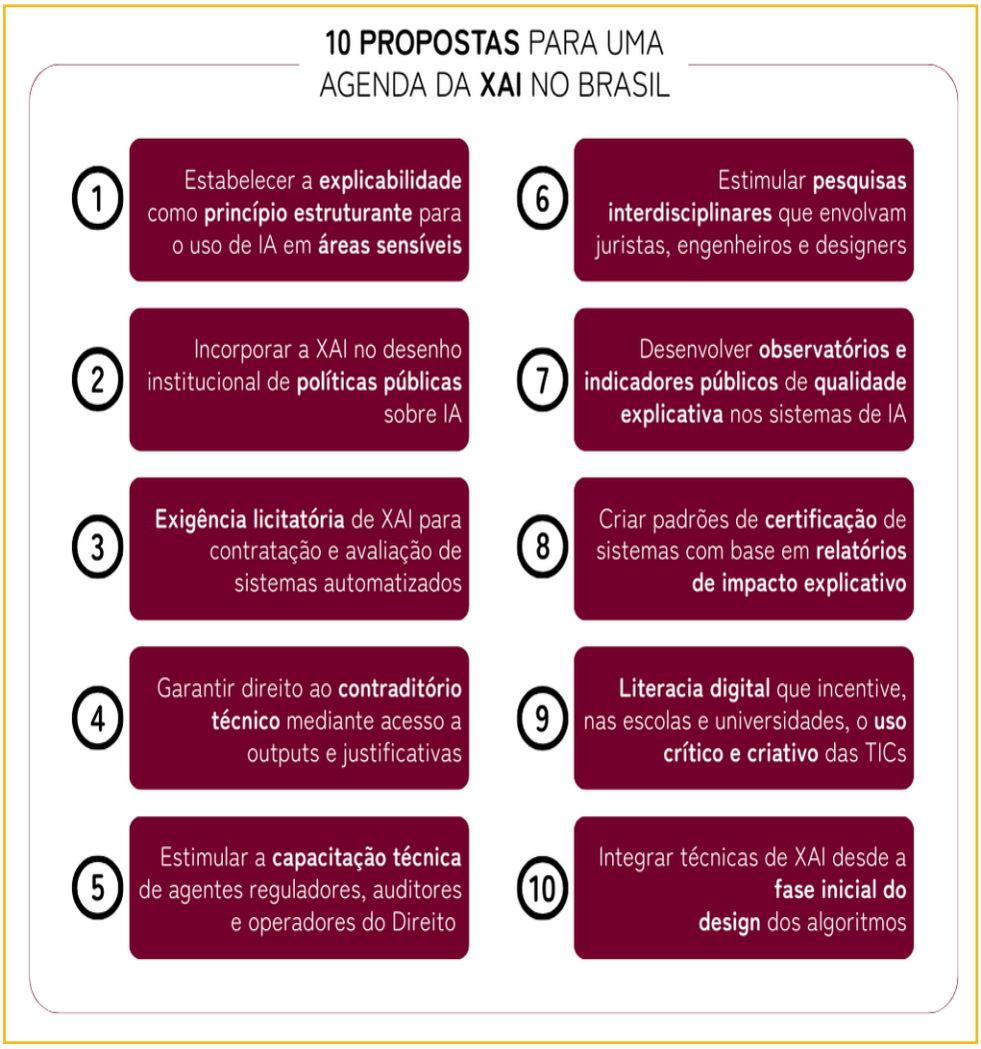
Para tanto, é preciso traduzir os fundamentos do direito à explicação em uma agenda concreta. Um ponto de partida estratégico é o poder de compra e regulação

do Estado. A exigência de explicabilidade como requisito obrigatório em processos licitatórios para a aquisição de sistemas de IA, especialmente em áreas sensíveis como saúde e segurança, mudaria o mercado, forçando fornecedores a desenvolverem produtos transparentes. Essa medida pode ser complementada pela criação de um selo público de *explicabilidade* ou um sistema de certificação algorítmica, que ateste, por meio de auditorias independentes, que um sistema cumpre requisitos mínimos de interpretabilidade e justiça. Tal certificação se basearia em relatórios de impacto explicativo, documentos que as empresas seriam obrigadas a elaborar, detalhando não apenas o funcionamento do modelo, mas também os riscos epistêmicos, sociais e jurídicos de sua opacidade.

Paralelamente, o sistema de justiça deve se aparelhar para essa nova realidade. É fundamental a capacitação contínua de agentes públicos e operadores do Direito para que possam interpretar e auditar decisões algorítmicas, evitando a aceitação acrítica de seus resultados. A explicabilidade das tecnologias consubstancia-se em requisito de aceitação social. Mais do que treinamento, é preciso criar mecanismos processuais, como a institucionalização do "contraditório técnico", que garanta às partes o direito de questionar não apenas a decisão, mas a própria lógica e a arquitetura do modelo que a produziu, com acesso a peritos e à documentação relevante. A consolidação dessa agenda dependerá também da litigância estratégica por parte da sociedade civil e do Ministério Público, levando ao Judiciário casos emblemáticos que ajudem a firmar teses garantistas sobre o tema.

Finalmente, essa agenda só se sustenta com o fomento a uma cultura de explicabilidade que transcenda o direito. É preciso incentivar a *pesquisa interdisciplinar*, conectando juristas, cientistas da computação, designers e cientistas sociais na construção de soluções contextualizadas e eficazes, promovendo a explicabilidade como princípio. A criação de *observatórios públicos*, com participação multissetorial, pode monitorar a implementação de IA no país e desenvolver indicadores de qualidade explicativa: métricas auditáveis e comparáveis para avaliar o grau de transparência dos sistemas em operação.

Figura 5 - Dez propostas para uma agenda de Inteligência Artificial Explicável no Brasil



Fonte: Autores (2026)

Por isso, é indispensável combinar capacitação contínua, monitoramento pós-implantação, auditorias independentes, documentação transparente e mecanismos de contestação com revisão humana real. A governança responsável exige abandonar o tecnocentrismo e reconhecer que justiça, contraditório e devido processo não cabem em métricas isoladas: exigem controle público, prestação de contas e integração com instituições sociais. Essa exigência decorre diretamente das limitações técnicas apresentadas, pois sistemas que operam com incerteza estatística, opacidade estrutural e dependência de dados históricos não conseguem, por si só, incorporar princípios jurídicos de contestação, imparcialidade e revisão.

Essas propostas configuram uma agenda mínima para compatibilizar inovação tecnológica com os princípios do Estado Democrático de Direito, abrindo caminho para uma governança algorítmica mais democrática e inclusiva. Seus desdobramentos dependem do aprofundamento técnico, jurídico e institucional, bem como da articulação entre Direito, técnica e cultura. Embora a XAI não seja suficiente para enfrentar isoladamente os desafios trazidos pela IA, ela constitui um ponto de partida indispensável para mitigar os riscos da opacidade, da automatização acrítica e da erosão das garantias institucionais, abrindo caminho para uma governança algorítmica mais democrática e inclusiva.

CONCLUSÃO

A consolidação da inteligência artificial como infraestrutura invisível de poder, operando decisões com consequências jurídicas e sociais profundas, colocou em evidência um paradoxo democrático: sistemas que impactam vidas concretas permanecem, em grande medida, ininteligíveis para aqueles que são por eles governados. É dessa fricção entre automatização e inteligibilidade que emerge a urgência por um princípio jurídico capaz de reabrir espaço para o contraditório e a responsabilização. A partir dessa perspectiva, explorou-se, neste trabalho, o direito à explicação enquanto resposta normativa à opacidade tecnológica, compreendendo a explicabilidade como exigência jurídica e democrática de inteligibilidade mínima das decisões automatizadas.

Nesse quadro, o direito à explicação projeta-se como instrumento de afirmação da cidadania algorítmica, ao instaurar um dever de inteligibilidade nas decisões produzidas por sistemas automatizados. Ao exigir que os critérios decisórios sejam compreensíveis e contestáveis, a explicabilidade reconduz o poder técnico ao espaço público do contraditório e do devido processo, preservando a centralidade da agência humana em ambientes governados por algoritmos. A explicação funciona como mecanismo de redistribuição de poder cognitivo, transformando resultados opacos

em decisões passíveis de interpretação, questionamento e revisão. Nesse cenário, a explicabilidade assume função constitucional: fortalece a legitimidade das decisões automatizadas, reativa o controle democrático sobre infraestruturas invisíveis e impede que o cálculo estatístico se converta silenciosamente em autoridade normativa. Em certos contextos de risco elevado, a opacidade torna-se inadmissível, impondo a adoção de modelos dotados de explicabilidade, ainda que isso resulte na majoração de custos ou implique em sacrificar ganhos marginais de eficiência.

Reivindicar explicações é, portanto, reivindicar lugar no processo decisório, reafirmando que nenhum modelo técnico substitui o princípio fundamental segundo o qual toda decisão que afeta pessoas deve poder ser compreendida, questionada e respondida. A partir desse enquadramento, o direito à explicação deixa de figurar como promessa regulatória e se converte em tarefa constitucional permanente das instituições democráticas brasileiras. Essa tarefa é hermenêutica e prática: requer justificativas inteligíveis hoje e a construção contínua de capacidades públicas para explicá-las amanhã.

A efetivação desta agenda, contudo, é um campo de batalha travado em múltiplas frentes. É uma disputa técnica, entre a performance dos modelos e sua interpretabilidade; uma disputa comercial, entre a transparência e o segredo industrial; e, no cenário brasileiro, uma disputa jurídico-política, entre a urgência da garantia e um arcabouço legal fragilizado pelo veto à revisão humana obrigatória[XVI]. O sucesso, portanto, dependerá da capacidade dos agentes sociais e políticos de construir um arranjo institucional robusto, que saiba arbitrar esses conflitos com base em princípios democráticos.

XVI Veto Presidencial ao § 3º do art. 20 da Lei nº 13.709/2018 (LGPD), sancionada pelo então presidente Jair Bolsonaro. O dispositivo vetado previa que a revisão da decisão automatizada seria realizada por pessoa natural, tornando a revisão humana não mais explícita ou obrigatória por lei.

REFERÊNCIAS

AI INCIDENT DATABASE. **Responsible AI Collaborative**, 2025. Disponível em: https://incidentdatabase.ai. Acesso em: 24 nov. 2025.

ALVES, Marco Antônio Sousa; MORATO, Otávio. "Da 'caixa-preta' à 'caixa de vidro': o uso da explainable artificial intelligence (XAI) para reduzir a opacidade e enfrentar o enviesamento em modelos algorítmicos". **Direito Público**, vol. 18, n. 100, 2021. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/5973. Acesso em: 20 dez. 2025.

ANDERS, Christopher; WEBER, Leander; NEUMANN, David; SAMEK, Wojciech; MÜLLER, Klaus-Robert; LAPUSCHKIN, Sebastian. Finding and removing Clever Hans: Using explanation methods to debug and improve deep models. **Information Fusion**, vol. 77, 2022, pp. 261-95, Disponível em: https://doi.org/10.1016/j.inffus.2021.07.015. Acesso em: 24 nov. 2025.

ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks. **ProPublica**, 23/5/2016, Disponível em: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Acesso em: 24 nov. 2025.

ARRIETA, Alejandro; DÍAZ-RODRIGUEZ, Natalia; DEL SER, Javier; BENNETOT, Adrien; TABIK, Siham; BARBADO, Alberto; GARCIA, Salvador; GIL-LOPEZ, Sergio; MOLINA, Daniel; BENJAMINS, Richard; CHATILA, Raja; HERRERA, Francisco. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. **Information Fusion**, vol. 58, 2020, pp. 82-115. Elsevier. Disponível em: www.sciencedirect.com/science/article/pii/S1566253519308103. Acesso em: 20 nov. 2025.

BRASIL. **Lei Geral de Proteção de Dados Pessoais (LGPD)**. Lei nº 13.709, de 14 de agosto de 2018. Planalto. Disponível em: www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 20 nov. 2025.

BRASIL. **Projeto de Lei n° 2338, de 2023 (Marco da Inteligência Artificial)**. Senado Federal. 3/5/2023. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/161795. Acesso em: 28 nov. 2025.

BRASIL. Senado Federal. **Projeto de Lei nº 4, de 2025 (Reforma do Código Civil)**. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/166998. Acesso em: 20 nov. 2025.

CANO, Rosa. O robô racista, sexista e xenófobo da Microsoft acaba silenciado. **El País**, 25/3/2016. Disponível em: www.elpais.com/tecnologia/2016-03-25/robot-racista-sexista-xenofobo-microsoft-tay. Acesso em: 20 out. 2025.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). **Resolução n. 332**, de 21 de agosto de 2020. Brasília: CNJ, 2020. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em: 20 out. 2025.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). **Resolução nº 615**, de 11 de março de 2025. Brasília: CNJ, 2025. Disponível em: https://atos.cnj.jus.br/files/original155530202503146 7d4517244566.pdf. Acesso em: 18 set. 2025.

DEGRAVE, Alex; JANIZEK, Joseph; LEE, Su-In. AI for Radiographic COVID-19 Detection Selects Shortcuts over Signal. **Nature Machine Intelligence**, vol. 3, 2021, pp. 610–619, Disponível em: www.nature.com/articles/s42256-021-00338-7. Acesso em: 28 nov. 2025.

DUARTE, Livia; SOUZA, Silene; AGUIRRE, Kevin; SOUSA, Keyla. Inteligência artificial no diagnóstico médico: entre precisão algorítmica e raciocínio clínico humano. **Revista Caderno Pedagógico**, vol. 22, no. 11, 2025, pp. 1–19. Disponível em: https://ojs.studiespublicacoes.com.br/ojs/index.php/cadped/article/view/19612. Acesso em: 28 nov. 2025.

ESTADOS UNIDOS DA AMÉRICA. Executive Office of the President. Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People. **The White House**, 2022. Disponível em: https://www.whitehouse.gov/ostp/ai-bill-of-rights/. Acesso em: 23 nov. 2025.

INFORMATION COMMISSIONER'S OFFICE (ICO). **Explaining decisions made with AI**. Londres, 2020. Disponível em: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/. Acesso em: 23 nov. 2025.

GDPRHUB. CJEU - C-203/22 - Dun & Bradstreet Austria. 4/4/2025. Disponível em: https://gdprhub.eu/index.php?title=CJEU_-_C-203/22_-*Dun*%26_Bradstreet_Austria. Acesso em: 23 nov. 2025.

ISO. **Policy Brief**: Harnessing International Standards for responsible AI development and governance. Disponível em: https://www.iso.org/publication/PUB100498.html. Acesso em: 18 out. 2025.

KNIGHT, Will. The Apple Card Didn't 'See' Gender – and That's the Problem. **Wired**, 19/11/2019, Disponível em: www.wired.com/story/the-apple-card-didnt-see-gender-and-thats-the-problem. Acesso em: 18 out. 2025.

MORATO, Otávio. **Governamentalidade algorítmica***:* democracia em risco? 1. ed. São Paulo: Dialética, 2022.

MORATO, Otávio; Nunes, Dierle. **Inteligência artificial***:* o desafio da explicabilidade. Salvador: Juspodivm, 2025.

NOVAES, Thiago; MORATO, Otávio. Máquinas abertas e rizomáticas: reflexões sobre o software livre e de código aberto (FLOSS). **Revista da Faculdade de Direito da UFMG**, vol. 87, 2025. https://revista.direito.ufmg.br/index.php/revista/pt_BR/article/view/3046. Acesso em: 09 dez. 2025.

NUNES, Dierle; MORATO, Otávio. O uso da inteligência artificial explicável enquanto ferramenta para compreender decisões automatizadas: possível caminho para aumentar a legitimidade e confiabilidade dos modelos algorítmicos? **Revista Eletrônica do Curso de Direito da UFSM**, vol. 18, no. 1, 2023. Disponível em: https://periodicos.ufsm.br/revistadireito/article/view/69329. Acesso em: 18 out. 2025.

ONETRUST DATAGUIDANCE. Canada: Bill C-27 dies after Parliament is prorogued. **DataGuidance**, 13/02/ 2025. Disponível em: https://www.dataguidance.com/news/canada-bill-c-27-dies-after-parliament-prorogued. Acesso em: 04 out. 2025.

PASQUALE, Frank. **The Black Box Society***:* The secret algorithms that control money and information. Harvard University Press, 2015.

QLIK. The six principles of AI ready data. **Qlik**, 2024., Disponível em: https://pages.qlik.com/rs/049-DKK-796/images/The%20Six%20Principles%20of%20AI%20Ready%20Data.pdf. Acesso em 28 nov. 2025.

REIGELUTH, Tyler; VIANA, Diego; MORATO, Otávio. Openness and closure: explainable artificial intelligence and Simondon's 'technical mentality'. **Journal of Human-Technology Relations**, 3, 2025. Disponível em: https://journals.open.tudelft.nl/jhtr/article/view/8091. Acesso em: 02 jan. 2026.

UNIÃO EUROPEIA. **Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016**. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=urisery%3AOJ.L_.2016.119.01.0001.01.POR&toc=OJ%3AL%3A2016%3A119%3AFULL. Acesso em: 18 out. 2025.

UNIÃO EUROPEIA. Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, que estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento Inteligência Artificial). **Jornal Oficial da União Europeia**, vol. 67, 12 de julho de 2024, http://data.europa.eu/eli/reg/2024/1689/oj. Disponível em: https://artificialintelligenceact.eu/ai-act-explorer/. Acesso em: 18 out. 2025.

VALE, Luis Manoel Borges do; PEREIRA, João Sérgio dos Santos Soares*.* **Teoria Geral do processo tecnológico**. São Paulo: Thompson Reuters Brasil, 2023.

ZHANG, Yuxiao; CARBALLO, Alexander; YANG, Hanting; TAKEDA, Kazuya. Perception and sensing for autonomous vehicles under adverse weather conditions: a survey. **ISPRS Journal of Photogrammetry and Remote Sensing**, vol. 196, 2023, pp. 146-177. Disponível em: https://doi.org/10.1016/j.isprsjprs.2022.12.021. Acesso em: 28 nov. 2025.

Autoria

1 Otávio Morato

Doutor em Direito pela Universidade Federal de Minas Gerais; Professor
https://orcid.org/0000-0002-0541-7353 • otaviomorato@gmail.com

2 Dierle José Coelho Nunes

Doutor em Direito Processual pela Pontifícia Universidade Católica de Minas Gerais; Professor
https://orcid.org/0000-0003-4724-5956 • dierlenunes@gmail.com

3 Viviane Ramone Tavares

Especialização em Direito Processual Civil pela Universidade Federal de Uberlândia; Mestranda
https://orcid.org/0000-0001-9252-3930 • vivianeramone@gmail.com

Como citar este artigo

MORATO, O.; NUNES, D. J. C.; TAVARES, V. R. Direito à explicação no cenário algorítmico: fundamentos e esboço de agenda para uma governança ética da IA. **InterAção**, Santa Maria, v. 17, n. 1, e95074, p. 1-25, mar. 2026. DOI 10.5902/1980509895074. Disponível em: https://dx.doi.org/10.5902/2357797595074. Acesso em: dia mês abreviado. ano.