DYEGO GIRÃO ANJOS
KARLA LÔBO FAGUNDES
ALAN DAVID SOUZA SALUSTIANO DA SILVA
MADISON LOPES INOCENCIO



DIREITO NA ERA DA INTELIGÊNCIA ARTIFICIAL

Deepfakes, Proteção de Dados, Responsabilidade Algorítmica e Prova Digital

EDITORA MANUAL

DIREITO NA ERA DA INTELIGÊNCIA ARTIFICIAL

Deepfakes, Proteção de Dados, Responsabilidade Algorítmica e Prova Digital

Dyego Girão Anjos • Karla Lôbo Fagundes

Alan David Souza Salustiano da Silva • Madison Lopes Inocencio

Organização: Kátia Karolini Amaro El Alam

EDITORA MANUAL • 2026

DIREITO NA ERA DA INTELIGÊNCIA ARTIFICIAL

Deepfakes, Proteção de Dados, Responsabilidade Algorítmica e Prova Digital

• • •

Dyego Girão Anjos • Karla Lôbo Fagundes

Alan David Souza Salustiano da Silva • Madison Lopes Inocencio

Organização: Kátia Karolini Amaro El Alam

EDITORA MANUAL • 1ª edição • 2026

EXPEDIENTE EDITORIAL

TÍTULO

Direito na Era da Inteligência Artificial: Deepfakes, Proteção de Dados, Responsabilidade Algorítmica e Prova Digital

EDITORA

EDITORA MANUAL CNPJ: 12.072.139/0001-20

(27) 99750-4756 www.editoramanual.com.br

editoramanualoficial@gmail.com

Rua Costa Brandão,785, Loja A, São Benedito, Cariacica, Espírito Santo – ES, CEP: 29145-340

REGISTROS

ISBN: 978-65-83121-30-1

DOI: 10.29327/5950127

AUTORIA

Dyego Girão Anjos

Karla Lôbo Fagundes

Alan David Souza Salustiano da Silva

Madison Lopes Inocencio

ORGANIZAÇÃO

Kátia Karolini Amaro El Alam

EDIÇÃO

1ª edição, 2026

EDITOR-CHEFE

Afonso Henrique Souza de Assis

PRODUTOR EDITORIAL

Kátia Karolini Amaro El Alam

REVISÃO

Kátia Karolini Amaro El Alam

CAPA

Claude

BIBLIOTECÁRIA RESPONSÁVEL

Janaina Ramos

REVISÃO DE TEXTO E CAPA

Kátia Karolini Amaro El Alam

DIAGRAMAÇÃO

Kátia Karolini Amaro El Alam

CORPO EDITORIAL

Dra. Mônica Aparecida Bortolotti – Universidade Estadual do Paraná

Dr. Walmir Fernandes Pereira – MUST University

Dr. Luiz Gonzaga Lapa Junior – Universidade de Brasília

Dr. Gabriel Arruda Burani - Univerdidad de Ciencias Empresariales e Sociales

Dr. Januario Neto Pereira Sarmento – Instituto Federal de Educação, Ciência e Tecnologia do Tocantins

Dr. Tiago Piñeiro Martins – Secretaria Municipal de Educação de Salvador

Dra. Lucimara Cabreira Benitez – Universidade Estadual de Mato Grosso do Sul

Dr. Abner Alexandre Nogueira - Universidade Federal de Alfenas

Me. Afonso Henrique Souza de Assis – Secretaria Estadual de Educação do Espírito Santo

Me. Elizabete Nascimento Aguiar – Universidade de Passo Fundo

Me. Swelen Freitas Gabarron – Universidade Tuiuti do Paraná

Me. José Neto de Oliveira Felippe – Faculdade de Caldas Novas

Me. Sueli Cristina Merotto Pereira – Secretaria Municipal de Educação - Prefeitura de Marataízes

Me. Guilherme de Souza Marques – Universidade Federal do Rio de Janeiro

Catalogação na publicação

Elaborada por Bibliotecária Janaina Ramos – CRB-8/9166

D598

Direito na era da inteligência artificial: deepfakes, proteção de dados, responsabilidade algorítmica e prova digital / Organização de Kátia Karolini Amaro El Alam. – Cariacica/ES: Manual, 2026.

Autores: Dyego Girão Anjos, Karla Lôbo Fagundes, Alan David Souza Salustiano da Silva, Madison Lopes Inocêncio.

Livro em PDF

ISBN 978-65-83121-30-1

1. Direito e tecnologia. 2. Inteligência artificial. 3. Proteção de dados. 4. Prova digital. I. Alam, Kátia Karolini Amaro El (Organização). II. Título.

CDD 340.0285

Índice para catálogo sistemático

I. Direito e tecnologia

Obra publicada em Setembro de 2026.

© 2026. Todos os direitos reservados.

Nenhuma parte desta obra poderá ser reproduzida, armazenada ou transmitida por qualquer meio sem observância da legislação aplicável e dos direitos autorais.

APRESENTAÇÃO

Direito na Era da Inteligência Artificial: Deepfakes, Proteção de Dados, Responsabilidade Algorítmica e Prova Digital examina conflitos jurídicos que já deixaram de ser hipótese de futuro. Sistemas capazes de classificar pessoas, produzir textos, imagens e vozes sintéticas, apoiar decisões e operar sobre grandes volumes de dados passaram a interferir em direitos da personalidade, relações de consumo, atuação estatal, processos judiciais e estruturas de responsabilidade.

A obra parte de uma premissa central: a novidade tecnológica não suspende o Direito, mas altera as formas pelas quais direitos podem ser atingidos, provas podem ser produzidas, danos podem se propagar e decisões podem ser tomadas. Por isso, os capítulos articulam Constituição, proteção de dados, responsabilidade civil e penal, direito do consumidor, prova digital, transparência algorítmica, uso de inteligência artificial pelo poder público e os marcos regulatórios em construção.

O percurso também evita tratar a inteligência artificial como categoria única. Deepfakes, sistemas de recomendação, ferramentas generativas, reconhecimento automatizado e mecanismos de decisão produzem riscos distintos e exigem respostas proporcionais ao contexto, ao grau de autonomia, aos dados envolvidos e às consequências para indivíduos e coletividades. A análise privilegia responsabilidade, explicabilidade, rastreabilidade, supervisão humana e proteção de direitos fundamentais.

Destinada a estudantes, pesquisadores, profissionais do Direito, gestores públicos e leitores interessados nas transformações digitais, a obra propõe uma leitura crítica e aplicada da relação entre inteligência artificial e ordem jurídica. O desafio não é escolher entre inovação e proteção, mas construir condições para que o desenvolvimento tecnológico permaneça compatível com dignidade, liberdade, igualdade, segurança jurídica e possibilidade efetiva de contestação.

SUMÁRIO

INTRODUÇÃO 10

CAPÍTULO 1 — INTELIGÊNCIA ARTIFICIAL E OS NOVOS DESAFIOS DO DIREITO CONTEMPORÂNEO 13

CAPÍTULO 2 — DEEPFAKES, CLONAGEM DE VOZ E PROTEÇÃO DOS DIREITOS DA PERSONALIDADE 23

CAPÍTULO 3 — PROTEÇÃO DE DADOS PESSOAIS E TRATAMENTO AUTOMATIZADO DE INFORMAÇÕES 31

CAPÍTULO 4 — RESPONSABILIDADE CIVIL POR DANOS DECORRENTES DE SISTEMAS DE INTELIGÊNCIA ARTIFICIAL 40

CAPÍTULO 5 — INTELIGÊNCIA ARTIFICIAL, CRIMES DIGITAIS E RESPONSABILIDADE PENAL 48

CAPÍTULO 6 — PROVA DIGITAL E CONTEÚDOS PRODUZIDOS OU MANIPULADOS POR INTELIGÊNCIA ARTIFICIAL 56

CAPÍTULO 7 — DECISÕES AUTOMATIZADAS, TRANSPARÊNCIA E RESPONSABILIDADE ALGORÍTMICA 64

CAPÍTULO 8 — IA NAS RELAÇÕES DE CONSUMO E PROTEÇÃO DO CONSUMIDOR 72

CAPÍTULO 9 — USO DE INTELIGÊNCIA ARTIFICIAL PELO PODER PÚBLICO E SEUS LIMITES 81

CAPÍTULO 10 — ÉTICA, REGULAÇÃO E PERSPECTIVAS JURÍDICAS PARA A INTELIGÊNCIA ARTIFICIAL 87

CONSIDERAÇÕES FINAIS 98

REFERÊNCIAS 103

SOBRE OS AUTORES 113

INTRODUÇÃO

Quando a inteligência artificial deixa de ser apenas ferramenta

A inteligência artificial (IA) chegou ao Direito antes que o Direito pudesse decidir, com tranquilidade, o que fazer com ela. Sistemas capazes de classificar pessoas, recomendar condutas, estimar riscos, reconhecer padrões, produzir textos, criar imagens e imitar vozes passaram a participar de relações sociais e econômicas que já eram juridicamente reguladas. O problema, portanto, não começa com a pergunta sobre a existência de uma "lei da inteligência artificial". Começa quando tecnologias novas passam a interferir em bens jurídicos antigos: identidade, honra, imagem, privacidade, igualdade, liberdade, patrimônio, confiança, prova, devido processo e responsabilidade. O desafio contemporâneo não é construir o Direito a partir do zero, mas descobrir até onde suas categorias tradicionais continuam suficientes quando a ação humana passa a ser mediada por sistemas que operam em escala, aprendem a partir de dados e produzem resultados nem sempre inteiramente previsíveis.

Essa perspectiva evita dois equívocos recorrentes. O primeiro é tratar toda aplicação de inteligência artificial como se apresentasse o mesmo risco. Um filtro de spam, um sistema de recomendação de conteúdo, um modelo que decide quais currículos serão priorizados, um reconhecimento facial empregado pelo Estado e uma ferramenta generativa capaz de produzir uma fotografia falsa não ocupam a mesma posição jurídica. O segundo equívoco é imaginar que a novidade técnica elimina as normas existentes. No Brasil, a Constituição Federal já protege dignidade, igualdade, intimidade, vida privada, imagem, devido processo e, desde a Emenda Constitucional nº 115/2022, reconhece expressamente a proteção de dados pessoais como direito fundamental. A Lei Geral de Proteção de Dados Pessoais (LGPD), o Marco Civil da Internet, o Código Civil, o Código de Defesa do Consumidor (CDC), as normas penais, processuais e administrativas continuam incidindo sobre relações mediadas por inteligência artificial (BRASIL, 1988; BRASIL, 1990; BRASIL, 2002; BRASIL, 2014; BRASIL, 2018).

O que muda é a forma como esses direitos podem ser atingidos. Em sistemas algorítmicos, uma decisão pode depender de milhares de variáveis, de bases de dados construídas durante anos e de escolhas realizadas por diferentes agentes ao longo do ciclo de vida tecnológico. Um resultado discriminatório pode não nascer de uma regra explicitamente preconceituosa, mas de dados históricos, variáveis substitutas, objetivos mal definidos ou diferenças de representação entre grupos. Uma informação

falsa pode não ser escrita por uma pessoa identificável, mas sintetizada por um modelo generativo e difundida em segundos. Uma decisão automatizada pode produzir impacto sobre crédito, emprego, consumo ou acesso a serviços sem que a pessoa afetada consiga perceber quais fatores foram determinantes. Por isso, autonomia, discriminação, transparência e responsabilidade tornaram-se temas inseparáveis da discussão jurídica sobre IA (DONEDA et al., 2018; MENDES; MATTIUZZO, 2019; MITTELSTADT et al., 2016).

O Brasil atravessa, em 2026, uma fase particularmente interessante dessa construção. Não existe, até a data de fechamento desta versão, uma lei geral de inteligência artificial em vigor. O Projeto de Lei (PL) nº 2.338/2023, aprovado pelo Senado e remetido à Câmara dos Deputados, permanece aguardando parecer na Comissão Especial. O texto em exame adota a centralidade da pessoa humana, a proteção de direitos fundamentais e uma lógica de classificação de riscos, mas sua condição jurídica precisa permanecer clara: trata-se de projeto legislativo, e não de direito positivo vigente (CÂMARA DOS DEPUTADOS, 2026a; CÂMARA DOS DEPUTADOS, 2026b). Ao mesmo tempo, a regulação não está parada. A Agência Nacional de Proteção de Dados (ANPD) vem trabalhando sobre tratamento automatizado e revisão de decisões, e o Conselho Nacional de Justiça (CNJ) já dispõe de disciplina específica para desenvolvimento, utilização e governança de IA no Poder Judiciário, atualizada em 2026 (ANPD, 2025a; CNJ, 2025; CNJ, 2026).

No plano internacional, o movimento regulatório confirma que a preocupação deixou de ser apenas ética ou prospectiva. O Regulamento Europeu de Inteligência Artificial (AI Act) tornou-se, em termos gerais, aplicável em 2 de agosto de 2026 e impõe deveres diferenciados conforme finalidade e risco, inclusive obrigações de transparência para determinadas interações com IA e conteúdos sintéticos. As regras de alto risco possuem cronograma específico, alterado em 2026. Em paralelo, a Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial, Direitos Humanos, Democracia e Estado de Direito foi aberta à assinatura em 2024 como primeiro tratado internacional juridicamente vinculante dedicado ao tema. A Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) e o National Institute of Standards and Technology (NIST) também convergem, embora por instrumentos de natureza distinta, em ideias como proteção de direitos, segurança, transparência, supervisão, rastreabilidade e responsabilização ao longo do ciclo de vida dos sistemas (CONSELHO DA EUROPA, 2024; NIST, 2023; OCDE,

2024; UNESCO, 2021; UNIÃO EUROPEIA, 2024a; COMISSÃO EUROPEIA, 2026).

Essa convergência não significa que exista uma resposta única para todos os conflitos. Pelo contrário. A inteligência artificial expõe tensões que o Direito precisa tratar sem abandonar suas próprias premissas. Transparência pode ser necessária, mas não basta quando uma explicação tecnicamente complexa continua inacessível à pessoa afetada. Consentimento pode ser relevante, mas não resolve relações marcadas por assimetrias profundas e tratamentos de dados difíceis de antecipar. Supervisão humana pode ser exigida, mas se torna vazia quando o responsável apenas confirma automaticamente aquilo que o sistema sugeriu. Responsabilidade pode ser distribuída entre desenvolvedor, fornecedor, instituição e usuário, mas não pode ser diluída até o ponto em que ninguém responda por uma decisão que produziu dano (MENDES; FONSECA, 2020; NIST, 2023; SELBST et al., 2019).

É nesse espaço entre capacidade técnica e consequência jurídica que esta obra se posiciona. Os capítulos não foram organizados como uma lista de tecnologias. A sequência acompanha a transformação do problema: começa pelo impacto geral da inteligência artificial sobre as categorias jurídicas; passa pela manipulação sintética da identidade, pelo tratamento automatizado de dados e pela responsabilidade civil e penal; entra no terreno sensível da prova digital e da decisão algorítmica; examina consumo e poder público; e termina discutindo ética, regulação e os limites que uma sociedade democrática deve impor ao uso da automação. A pergunta que atravessa o livro não é se a tecnologia deve ou não existir. É mais exigente: em quais condições seu uso pode ser juridicamente legítimo quando afeta pessoas concretas.

Essa pergunta exige abandonar tanto o fascínio quanto o medo como critérios normativos. A inovação pode ampliar acesso, reduzir custos, organizar informação e apoiar decisões complexas. Também pode ampliar vigilância, acelerar discriminações, dificultar contestação, multiplicar falsificações convincentes e deslocar poder decisório para sistemas cuja lógica permanece distante de quem sofre seus efeitos. O Direito contemporâneo precisa, portanto, preservar algo que nenhuma sofisticação técnica torna dispensável: a possibilidade de identificar quem exerce poder, compreender quais direitos estão em jogo, contestar decisões e atribuir responsabilidade por consequências. Esse é o critério que permite enfrentar a tecnologia sem perder de vista a estrutura jurídica das relações que ela transforma.

CAPÍTULO 1

INTELIGÊNCIA ARTIFICIAL E OS NOVOS DESAFIOS DO DIREITO CONTEMPORÂNEO

O problema jurídico começa antes da resposta

A expressão "inteligência artificial" reúne sistemas muito diferentes entre si. Alguns classificam informações; outros calculam probabilidades, recomendam ações, reconhecem padrões biométricos ou produzem conteúdo novo a partir de instruções. Essa diversidade importa juridicamente porque a tecnologia, isoladamente considerada, não define o risco. O significado de um sistema depende da finalidade, do contexto, dos dados utilizados, das pessoas afetadas e da consequência atribuída à sua saída. Um mesmo método computacional pode ser banal quando organiza fotografias pessoais e profundamente sensível quando identifica pessoas em um espaço público ou influencia a concessão de um benefício. Regular a IA, por isso, exige olhar menos para o rótulo tecnológico e mais para a função que o sistema ocupa dentro de uma relação social.

Esse raciocínio aproxima a discussão jurídica de uma compreensão sociotécnica da inteligência artificial. Modelos não existem fora das instituições que os constroem e utilizam. Antes do resultado aparecer na tela, alguém definiu um objetivo, selecionou dados, escolheu métricas, estabeleceu limiares, decidiu quais erros seriam tolerados e inseriu o sistema em determinado fluxo. Depois da saída, outra pessoa ou organização decide como interpretá-la e que efeito produzir. Selbst et al. (2019) chamam atenção para o risco de abstrair sistemas algorítmicos de seu contexto social e imaginar que problemas de justiça possam ser resolvidos apenas por ajustes matemáticos. No Direito, essa advertência é central: uma decisão pode ser tecnicamente consistente e ainda ser juridicamente inadequada porque o procedimento, o critério ou o efeito viola direitos.

A discussão sobre neutralidade ilustra bem esse ponto. Algoritmos não carregam intenção moral própria, mas isso não significa que seus resultados sejam neutros. Dados refletem práticas anteriores, escolhas institucionais, desigualdades de acesso e modos históricos de registrar pessoas. Variáveis aparentemente objetivas podem funcionar como substitutas de características protegidas. Uma base de dados pode conter poucos exemplos de determinado grupo e levar o modelo a errar mais justamente onde há maior vulnerabilidade. Mendes e Mattiuzzo (2019) mostram que discriminação algorítmica não depende de uma ordem explícita para discriminar; ela

pode surgir da interação entre classificação, perfilamento e estruturas sociais já desiguais. O desafio jurídico, portanto, não termina quando não se encontra uma regra discriminatória escrita no código.

É por isso que os direitos fundamentais precisam aparecer antes das soluções técnicas. A Constituição Federal oferece um ponto de partida que independe da tecnologia utilizada. Dignidade da pessoa humana, igualdade, liberdade, intimidade, vida privada, imagem, contraditório, ampla defesa e devido processo continuam sendo parâmetros para avaliar atos públicos e privados. A proteção de dados pessoais, expressamente incorporada ao catálogo de direitos fundamentais, reforça que o controle sobre informações relacionadas à pessoa não pode ser reduzido a interesse econômico sobre bases de dados (BRASIL, 1988). Quando um sistema influencia oportunidades, reputação ou acesso a direitos, a discussão jurídica precisa começar pela posição da pessoa afetada, e não pela sofisticação do modelo.

A Lei Geral de Proteção de Dados Pessoais torna essa relação ainda mais concreta. O art. 20 da LGPD assegura ao titular o direito de solicitar revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluindo decisões destinadas à formação de perfis pessoais, profissionais, de consumo e de crédito. A lei também prevê o fornecimento de informações claras e adequadas sobre critérios e procedimentos utilizados, observados os segredos comercial e industrial (BRASIL, 2018). Esse desenho é importante por dois motivos. Primeiro, porque reconhece que a automação pode produzir efeitos jurídicos relevantes. Segundo, porque mostra que a proteção não se resume ao momento da coleta dos dados: alcança também aquilo que é inferido e decidido a partir deles.

Ainda assim, o art. 20 da LGPD não resolve todas as controvérsias. A própria noção de decisão "unicamente" automatizada pode ser difícil de aplicar quando existe uma intervenção humana formal, porém pouco substantiva. Um analista que recebe uma recomendação algorítmica e a confirma em segundos realizou, tecnicamente, uma ação humana; a questão jurídica é saber se essa participação foi suficiente para alterar a natureza do processo decisório e oferecer proteção efetiva. A ANPD tem discutido exatamente os contornos do tratamento automatizado, da revisão e dos direitos dos titulares, o que demonstra que a aplicação prática da LGPD ainda está em construção (ANPD, 2025a; ANPD, 2026a). A presença de uma pessoa no fluxo não pode ser transformada em ficção capaz de neutralizar obrigações de transparência e contestação.

Há também um problema de autonomia. Sistemas algorítmicos não apenas respondem a escolhas; muitas vezes moldam o ambiente no qual escolhas são feitas. Recomendações, ordenações, previsões e personalizações selecionam o que aparece primeiro, o que recebe destaque e quais opções parecem mais relevantes. Doneda et al. (2018) observam que a combinação entre intensa utilização de dados e processos automatizados cria impactos sobre personalidade e autonomia que não se deixam explicar apenas pela ideia clássica de consentimento. Em ambientes altamente mediados por dados, o Direito precisa considerar não apenas se alguém clicou em "aceito", mas se o desenho da relação preservou condições materiais para liberdade, compreensão e contestação.

Essa preocupação é reforçada pela literatura que propõe uma proteção de dados menos dependente do consentimento como solução universal. Mendes e Fonseca (2020) mostram que assimetrias informacionais, complexidade tecnológica e dependência de serviços reduzem a capacidade do titular de administrar sozinho todos os riscos do tratamento. A resposta jurídica passa, então, por mecanismos de accountability, desenho de sistemas, avaliação prévia de riscos e deveres atribuídos a quem decide coletar, desenvolver e utilizar tecnologia. Essa mudança é decisiva para a inteligência artificial: a proteção não pode ser transferida integralmente para o indivíduo por meio de termos de uso extensos que ele não tem condições reais de avaliar.

Os modelos generativos acrescentam uma nova camada. Sistemas tradicionais de classificação ou previsão produzem uma categoria, uma pontuação ou uma estimativa. Modelos generativos podem produzir linguagem, imagens, vídeos e áudio com aparência de conteúdo humano. Isso desloca o problema jurídico da decisão para a própria confiança na informação. Uma fotografia já não serve, por si só, como forte indício de que determinado evento ocorreu; uma gravação de voz pode reproduzir alguém que nunca pronunciou aquelas palavras; um texto convincente pode conter fatos ou citações inexistentes. Em agosto de 2026, as regras europeias de transparência para determinados conteúdos sintéticos e deepfakes já passaram a integrar o regime aplicável do Regulamento Europeu de Inteligência Artificial (AI Act), indicando que integridade informacional se tornou objeto regulatório explícito (COMISSÃO EUROPEIA, 2026).

O ponto comum entre esses exemplos é a redistribuição de poder. Quem controla um modelo pode definir critérios, selecionar dados, modificar interfaces e ampliar a escala de decisões que antes dependiam de análise individual. Mittelstadt et al. (2016)

organizam os problemas éticos dos algoritmos em torno de evidências insuficientes ou enviesadas, opacidade, resultados injustos, impactos sobre autonomia e dificuldades de rastrear responsabilidade. No plano jurídico, essas categorias convergem para uma pergunta simples de formular, mas difícil de responder: quando a tecnologia participa de uma decisão que afeta direitos, quais garantias precisam acompanhar essa participação?

Quando decisão, autoria e responsabilidade se dispersam

O Direito foi construído, em grande parte, para reconstruir relações entre sujeitos, condutas e consequências. A responsabilidade civil pergunta quem praticou o ato, qual dever foi violado, que dano ocorreu e como se estabelece o nexo causal. O Direito Penal exige critérios ainda mais rigorosos de imputação pessoal. O processo depende de autoria, autenticidade, contraditório e possibilidade de verificar fontes. A inteligência artificial não elimina essas categorias, mas pode tornar mais difícil localizar cada elemento dentro de uma cadeia tecnológica distribuída.

Considere um sistema que produz uma recomendação prejudicial. O modelo pode ter sido desenvolvido por uma empresa, treinado com dados adquiridos de terceiros, adaptado por outro fornecedor, integrado a uma plataforma por uma instituição e utilizado por um profissional que desconhece detalhes de sua arquitetura. Se o resultado causa dano, a expressão "a IA errou" descreve o evento de modo tecnicamente pobre e juridicamente inútil. Sistemas não ingressam no ordenamento como sujeitos autônomos apenas porque produzem saídas complexas. É necessário reconstruir decisões humanas e organizacionais: quem definiu a finalidade, quem tinha controle sobre dados e parâmetros, quem conhecia os limites, quem poderia impedir o uso, quem tinha dever de monitorar e quem transformou a recomendação em efeito concreto.

Essa reconstrução é importante porque a linguagem da autonomia pode confundir níveis diferentes. Um sistema pode operar com elevado grau de autonomia funcional, realizando tarefas sem intervenção humana a cada etapa, sem que isso lhe atribua autonomia moral ou personalidade jurídica. Doneda et al. (2018) já apontavam que a sofisticação das aplicações de IA provocava debates sobre subjetividade e personalidade, mas a resposta jurídica não precisa criar uma "pessoa eletrônica" para enfrentar os problemas atuais. Em grande parte dos conflitos, é mais útil identificar deveres dos agentes humanos e institucionais que projetam,

disponibilizam e utilizam o sistema do que deslocar responsabilidade para a própria máquina.

O desafio aparece também na causalidade. Em sistemas de aprendizado, o resultado pode depender de interações entre dados, arquitetura e contexto de uso que nenhum participante isolado consegue explicar integralmente. Isso não significa que a causalidade jurídica se torne impossível. Significa que documentação, rastreabilidade e governança passam a ter função probatória. Registrar versões do modelo, origem dos dados relevantes, alterações, parâmetros de implantação e decisões humanas ao redor da saída deixa de ser mero detalhe de engenharia. Torna-se condição para reconstruir o que ocorreu quando um direito foi afetado.

A ideia de accountability procura justamente impedir que complexidade técnica se converta em irresponsabilidade prática. O NIST organiza seu framework de gestão de riscos em funções de governar, mapear, medir e gerenciar, concebendo o risco como algo que deve ser acompanhado durante o ciclo de vida, e não apenas verificado no momento de lançamento (NIST, 2023). Embora seja um instrumento voluntário e não uma fonte de obrigação jurídica no Brasil, ele ilustra uma mudança importante de perspectiva: sistemas confiáveis dependem de processos, responsabilidades e evidências capazes de mostrar como decisões foram tomadas e como riscos são tratados.

Transparência, por sua vez, precisa ser entendida como meio e não como fim. Publicar milhares de linhas de código não garante que a pessoa afetada compreenda por que recebeu determinado resultado. Explicabilidade juridicamente relevante precisa estar conectada à possibilidade de contestar. Pádua e Lorenzetto (2024), ao discutirem decisões estatais auxiliadas por IA, aproximam explicabilidade de accountability e da exigência de justificação do poder público. Essa associação é particularmente importante: quanto maior o impacto de uma decisão sobre direitos, menos aceitável se torna uma arquitetura em que o destinatário conhece o resultado, mas não consegue identificar critérios suficientes para questioná-lo.

No setor público, o problema ganha intensidade porque a decisão não é apenas uma escolha privada. Administração e Judiciário estão vinculados a legalidade, motivação, igualdade, devido processo e controle. A Resolução CNJ nº 615/2025 classifica como situações de alto risco determinados usos de IA capazes de apoiar avaliação de provas, interpretação de fatos e formulação de juízos conclusivos sobre aplicação de normas a casos concretos (CNJ, 2025). Em 2026, a Resolução CNJ nº

674 promoveu alteração pontual na composição do Comitê Nacional de Inteligência Artificial do Judiciário, sem modificar esse núcleo da classificação de risco (CNJ, 2026). A escolha regulatória é reveladora: não se proíbe toda automação, mas se reconhece que certas funções se aproximam do núcleo de atividades em que fundamentação, imparcialidade e responsabilidade não podem ser reduzidas a eficiência computacional.

A mesma lógica vale para relações privadas. Em seleção de candidatos, concessão de crédito, seguros, publicidade, consumo e gestão de trabalhadores, modelos podem ordenar pessoas sem produzir uma decisão final formalmente automática. Ainda assim, a recomendação pode determinar quais casos receberão atenção humana e quais serão descartados. O problema jurídico não é apenas saber se houve "automação total", mas compreender o peso real da saída algorítmica. Um sistema que influencia sistematicamente decisões humanas pode exercer poder normativo de fato mesmo quando a assinatura final pertence a uma pessoa.

Esse deslocamento também afeta autoria. Modelos generativos conseguem produzir conteúdos que parecem resultado direto de criação ou manifestação humana. O Direito autoral precisa distinguir ferramenta, contribuição humana e produção automatizada; os direitos da personalidade precisam lidar com voz, rosto e gestos sintetizados; o processo precisa avaliar autenticidade; e o Direito Penal precisa separar criação maliciosa, uso fraudulento e circulação de conteúdo. Não há uma única regra capaz de resolver todos esses problemas porque a palavra "gerado por IA" descreve apenas o método de produção, não o contexto, a intenção nem o dano.

A prova digital talvez seja o ponto em que essa transformação se torna mais intuitiva. O processo jurídico sempre dependeu da possibilidade de reconstruir acontecimentos passados. Documentos, fotografias, áudios e vídeos adquiriram valor precisamente porque podiam conectar um evento a um registro. Conteúdos sintéticos não destroem a prova digital, mas aumentam o custo da confiança. Autenticidade passa a depender mais de metadados, cadeia de custódia, perícia, origem, consistência contextual e mecanismos de verificação. A existência de deepfakes não significa que toda imagem seja falsa; significa que a inferência "parece real, logo aconteceu" se tornou juridicamente mais perigosa.

Por isso, o problema da inteligência artificial não pode ser reduzido à opacidade de modelos. Há sistemas transparentes capazes de causar danos e sistemas complexos que podem ser utilizados de forma segura quando existem controles adequados. A

questão jurídica é mais ampla: quem pode produzir determinado efeito, sob quais condições, com quais deveres de informação, documentação e supervisão, e que remédio existe quando a decisão viola direitos. A centralidade da pessoa não se concretiza pela simples presença de um humano no circuito; concretiza-se quando há agência, possibilidade de compreensão proporcional ao risco e mecanismos reais de contestação.

Entre regras existentes e novos regimes de risco

A expansão da inteligência artificial criou uma tentação regulatória compreensível: imaginar que cada nova capacidade exige uma lei completamente nova. Essa estratégia pode produzir normas rapidamente obsoletas. Sistemas mudam de arquitetura, modelos são atualizados e aplicações surgem em ritmos incompatíveis com processos legislativos. Ao mesmo tempo, confiar apenas em princípios genéricos pode deixar pessoas expostas diante de usos de alto impacto. O desafio é construir uma regulação que combine normas existentes, deveres setoriais e instrumentos específicos de gestão de risco.

Hoffmann-Riem (2019) observa que a inteligência artificial oferece ao Direito não apenas um problema, mas uma oportunidade de repensar formas regulatórias capazes de lidar com incerteza e transformação tecnológica. Em vez de buscar controlar matematicamente cada mecanismo interno, a regulação pode concentrar-se em finalidades, riscos, procedimentos, documentação e responsabilidades. Essa lógica aproxima-se do que diferentes marcos internacionais passaram a adotar: quanto maior a capacidade de afetar direitos, maior deve ser a exigência de governança.

No Brasil, esse movimento aparece em camadas. A Constituição Federal estabelece os limites superiores; a LGPD regula tratamento de dados e oferece direitos relevantes diante de decisões automatizadas; normas civis, consumeristas, penais e processuais continuam disciplinando danos, relações de consumo, ilícitos e prova; autoridades setoriais podem impor deveres próprios. O CNJ, por exemplo, já criou um regime de governança e classificação de risco para IA no Judiciário. A ANPD mantém inteligência artificial e tecnologias emergentes entre seus temas prioritários e vem construindo parâmetros para decisões automatizadas (ANPD, 2025a; ANPD, 2026a; CNJ, 2025). Isso demonstra que a ausência de uma lei geral não equivale a ausência de Direito.

O Projeto de Lei nº 2.338/2023 procura acrescentar uma camada horizontal a esse cenário. O texto aprovado no Senado e atualmente em análise na Câmara adota,

entre seus fundamentos, centralidade da pessoa humana, direitos humanos, igualdade, privacidade, proteção de dados, defesa do consumidor, integridade da informação e segurança cibernética. Também estrutura obrigações a partir de categorias de risco (CÂMARA DOS DEPUTADOS, 2026b). Essas escolhas dialogam com tendências internacionais, mas precisam ser apresentadas com precisão: enquanto o processo legislativo não terminar, elas funcionam como referência de debate e possível direção normativa, não como deveres legais gerais já vigentes.

A União Europeia oferece o exemplo comparado mais desenvolvido de regulação horizontal por risco. O AI Act entrou em vigor em 2024 e tornou-se, em termos gerais, aplicável em 2 de agosto de 2026. Algumas obrigações já haviam começado antes, enquanto regras específicas para sistemas de alto risco seguem cronograma posterior, alterado pelo pacote de simplificação de 2026. A arquitetura combina práticas proibidas, deveres para modelos de propósito geral, transparência para determinadas interações e conteúdos sintéticos, além de requisitos mais intensos em contextos sensíveis (UNIÃO EUROPEIA, 2024a; COMISSÃO EUROPEIA, 2026). Mais importante do que copiar categorias europeias é perceber o princípio regulatório: o nível de controle deve acompanhar a gravidade e a probabilidade do impacto.

No plano dos direitos humanos, a Convenção-Quadro do Conselho da Europa desloca a discussão para um nível ainda mais fundamental. O instrumento foi concebido para assegurar que atividades durante o ciclo de vida de sistemas de IA sejam compatíveis com direitos humanos, democracia e Estado de Direito, ao mesmo tempo em que admite inovação tecnológica (CONSELHO DA EUROPA, 2024). A formulação é relevante porque recusa a ideia de que desenvolvimento tecnológico e proteção de direitos sejam objetivos necessariamente opostos. A questão é institucional: inovação legítima precisa ocorrer dentro de estruturas que permitam prevenir, identificar e reparar violações.

A mesma convergência aparece em instrumentos não vinculantes. Os Princípios de IA da OCDE, atualizados em 2024, enfatizam direitos humanos, transparência, explicabilidade, robustez, segurança e accountability. A Recomendação da UNESCO de 2021 situa a governança de IA dentro de uma perspectiva de dignidade, direitos e impactos sociais. O NIST, por sua vez, transforma parte desses valores em uma linguagem operacional de gestão contínua de riscos. Floridi et al. (2018) já haviam organizado uma proposta ética em torno de beneficência, não maleficência, autonomia, justiça e explicabilidade. As fontes têm naturezas jurídicas diferentes, mas

convergem em algo essencial: confiança não é uma propriedade declarada pelo fornecedor; precisa ser sustentada por práticas verificáveis.

É por isso que a regulação contemporânea se afasta da ideia de um único momento de conformidade. Em sistemas de IA, riscos podem surgir no treinamento, na validação, na implantação, na atualização, na mudança do contexto e no uso não previsto. Um modelo adequado em determinada população pode produzir resultados diferentes em outra. Uma ferramenta originalmente usada para organizar informação pode ser posteriormente empregada para tomar decisões. Um sistema pode continuar tecnicamente igual enquanto a sociedade ao redor dele muda. A governança, portanto, precisa acompanhar o ciclo de vida e possuir mecanismos para suspender, corrigir ou abandonar usos que deixaram de ser defensáveis.

Essa abordagem também impede que transparência seja usada como substituto para responsabilidade. Uma empresa pode informar que utiliza IA e ainda assim manter um processo injusto. Um órgão público pode explicar tecnicamente seu sistema e continuar sem oferecer meio efetivo de recurso. Uma plataforma pode rotular conteúdo sintético e não responder adequadamente a uma fraude. O Direito precisa articular dever de informar com dever de prevenir, justificar, monitorar e reparar. Accountability não se resume a dizer quem utilizou o algoritmo depois que o dano ocorreu; começa na definição prévia de quem deveria ter controlado cada risco.

A complexidade aumenta porque os sistemas circulam entre fronteiras. Um modelo pode ser desenvolvido em um país, operado em infraestrutura localizada em outro, integrado a um serviço brasileiro e produzir efeitos sobre pessoas submetidas a regimes jurídicos distintos. Dados podem ser coletados em uma relação e reutilizados em outra. Conteúdos sintéticos atravessam plataformas em segundos. Nenhuma jurisdição consegue tratar esses fluxos apenas por regras territoriais simples. Daí a importância crescente de interoperabilidade regulatória, padrões técnicos e cooperação internacional, sem que isso elimine a autonomia constitucional de cada Estado para definir seus próprios limites.

O ponto de equilíbrio não está em regular pouco ou muito, mas em regular o que importa. O Direito precisa perguntar qual função o sistema exerce, quais pessoas podem ser atingidas, que bem jurídico está exposto, quão reversível é o dano, quem possui capacidade real de controle, quais informações precisam ser preservadas e como o resultado poderá ser contestado. Essas perguntas são mais estáveis do que a

arquitetura de um modelo específico e ajudam a construir uma regulação tecnologicamente informada sem se tornar tecnologicamente dependente.

Ao final, a inteligência artificial não cria um mundo jurídico paralelo. Ela atravessa o mundo existente e torna alguns de seus pressupostos menos seguros. A assinatura pode ser automatizada; a imagem pode ser sintetizada; a decisão pode resultar de uma pontuação; o autor aparente pode não ter produzido o texto; o dano pode emergir de milhares de decisões semelhantes; a discriminação pode não estar visível em uma regra. O papel do Direito é preservar aquilo que torna o exercício de poder legítimo: direitos, justificação, contestabilidade, proporcionalidade e responsabilidade.

O primeiro conflito concreto nasce exatamente dessa perda de segurança. Durante muito tempo, rosto e voz funcionaram como sinais poderosos de identidade, e fotografias, vídeos e gravações carregaram uma presunção social de ligação com acontecimentos reais. A inteligência artificial generativa enfraqueceu essa presunção ao tornar possível fabricar representações convincentes de pessoas que jamais disseram ou fizeram aquilo que o arquivo parece registrar. A partir daí, a questão deixa de ser apenas quais riscos a IA cria em abstrato e passa a atingir personalidade, consentimento, honra, imagem e confiança jurídica quando a própria identidade pode ser sintetizada.

CAPÍTULO 2

DEEPFAKES, CLONAGEM DE VOZ E PROTEÇÃO DOS DIREITOS DA PERSONALIDADE

Quando imagem e voz deixam de provar presença

Durante muito tempo, ver e ouvir funcionaram como atalhos de confiança. Uma fotografia podia ser manipulada, uma gravação podia ser editada e uma testemunha podia mentir, mas a imagem de um rosto em movimento e a voz reconhecível de uma pessoa ainda carregavam forte aparência de presença. A inteligência artificial generativa alterou esse equilíbrio ao reduzir drasticamente o custo de produzir conteúdos nos quais alguém parece dizer, fazer ou vivenciar algo que nunca ocorreu. O problema jurídico não está apenas na falsidade. Está no uso de elementos centrais da identidade humana para fabricar uma manifestação atribuída a uma pessoa real.

O termo deepfake costuma ser associado a vídeos hiper-realistas, mas a categoria prática é mais ampla. Sistemas atuais podem substituir rostos, sincronizar movimentos labiais, reproduzir timbre e entonação, gerar fotografias inteiras e criar diálogos a partir de poucos segundos de voz. O resultado pode ter qualidade desigual e nem sempre pretende enganar. Há usos legítimos em cinema, dublagem, acessibilidade, preservação cultural, educação, arte e sátira. Por isso, nenhuma análise juridicamente séria pode partir da fórmula segundo a qual todo conteúdo sintético é ilícito. A técnica descreve como algo foi produzido; a licitude depende de quem foi representado, do consentimento, da finalidade, do contexto, da possibilidade de confusão e do dano que a utilização pode produzir (CHESNEY; CITRON, 2019).

Essa distinção é essencial porque o deepfake opera sobre direitos que já existiam antes da tecnologia. A Constituição Federal protege intimidade, vida privada, honra e imagem e assegura indenização pelo dano decorrente de sua violação. O Código Civil, por sua vez, inclui entre os direitos da personalidade a proteção do nome, da palavra, da imagem e da vida privada, permitindo medidas para impedir ou fazer cessar usos juridicamente indevidos (BRASIL, 1988; BRASIL, 2002). Uma representação sintética não fica fora desse sistema apenas porque foi calculada por um modelo. Se o conteúdo apropria-se da identidade de alguém, o ponto de partida continua sendo a tutela da pessoa.

A imagem merece atenção especial porque o dano pode existir mesmo quando ninguém acredita integralmente no conteúdo. Uma montagem sexual ou humilhante, por exemplo, pode atingir dignidade, honra e reputação pela própria associação

pública entre a vítima e a cena fabricada. Em outro contexto, a imagem de alguém pode ser utilizada para vender um produto que jamais endossou. O Superior Tribunal de Justiça (STJ) consolidou, pela Súmula 403, que a indenização pela publicação não autorizada da imagem para fins econômicos ou comerciais independe de prova do prejuízo (STJ, 2009). O enunciado nasceu muito antes da popularização dos modelos generativos, mas sua lógica continua relevante: há situações em que a exploração indevida da imagem constitui, por si, violação juridicamente significativa.

A voz possui proteção equivalente em sua dimensão personalíssima. No Recurso Especial (REsp) nº 1.630.851/SP, a Terceira Turma do STJ reconheceu expressamente que a voz integra os direitos da personalidade e que seu uso pode ser objeto de disposição voluntária desde que a limitação não seja permanente nem geral e permaneça dentro da autorização concedida (STJ, 2017). O caso tratava de uma gravação real, não de clonagem por inteligência artificial. Justamente por isso é útil: demonstra que o núcleo jurídico do problema não nasceu com o sintetizador de voz. A novidade está na capacidade de separar a voz do ato original de fala e fazer com que ela produza, indefinidamente, novas mensagens que o titular nunca gravou.

Essa separação torna insuficiente a pergunta "a voz era pública?". Uma entrevista disponível na internet pode permitir que terceiros a ouçam; isso não significa, automaticamente, autorização para extrair características vocais e criar novas falas em nome do entrevistado. O mesmo vale para fotografias publicadas em redes sociais. A publicidade do arquivo original e a legitimidade de reutilizá-lo para reconstruir a identidade de alguém são questões diferentes. Em direito da personalidade, contexto e finalidade importam porque consentir com uma exposição determinada não equivale a abandonar o controle sobre todos os usos tecnologicamente possíveis.

A própria proteção de dados acrescenta uma dimensão que não deve ser confundida com o direito à imagem. A LGPD classifica como dado pessoal sensível o dado biométrico quando vinculado a uma pessoa natural. Face e voz podem servir a processos biométricos e, quando tratadas nessa condição, submetem-se às exigências reforçadas da lei (BRASIL, 2018). A Agência Nacional de Proteção de Dados incluiu, no mapa de temas prioritários, a utilização de dados biométricos — especificamente face e voz — para criação de mídias por inteligência artificial, relacionando o tema a riscos de privacidade, dignidade, fraude e reputação (ANPD, 2025b). Isso revela que o deepfake possui duas faces jurídicas: o conteúdo final que representa alguém e o tratamento de dados que tornou essa representação possível.

A capacidade de fabricar presença produz ainda um dano menos intuitivo. Quando conteúdos falsos se tornam tecnicamente plausíveis, conteúdos verdadeiros também podem ser negados com maior facilidade. Chesney e Citron (2019) descrevem esse efeito como um "dividendo do mentiroso": a existência social dos deepfakes oferece a quem foi legitimamente gravado um argumento pronto para rejeitar provas autênticas. A tecnologia, assim, não ameaça apenas quem é falsamente representado. Ela pode enfraquecer a confiança em registros verdadeiros.

Esse efeito encontra respaldo empírico no campo da comunicação política. Vaccari e Chadwick (2020), em experimento com vídeo político sintético, observaram que deepfakes não precisam necessariamente convencer o público de uma falsidade para produzir impacto; aumentar incerteza pode reduzir a confiança nas notícias que circulam em redes sociais. Para o Direito, a conclusão é importante. A tutela não pode olhar apenas para o indivíduo enganado. Conteúdos sintéticos podem afetar reputação, confiança pública, formação da vontade e a própria qualidade do ambiente informacional em que decisões jurídicas e políticas são tomadas.

Consentimento tem finalidade, contexto e limites

A facilidade técnica de copiar uma identidade pode criar a impressão de que o consentimento resolveria qualquer conflito: havendo autorização, o uso seria legítimo; faltando autorização, seria ilícito. A realidade é mais complexa. Direitos da personalidade admitem usos consentidos, mas a autorização não costuma funcionar como licença irrestrita. O próprio precedente do STJ sobre voz enfatiza que a disposição voluntária deve respeitar os limites do negócio realizado (STJ, 2017). Em ambientes digitais, essa exigência torna-se decisiva porque um único arquivo pode servir como matéria-prima para incontáveis usos futuros.

Imagine uma atriz que autoriza a gravação de sua voz para um filme. A autorização para aquele papel não responde, sozinha, se o estúdio poderá utilizar o material para criar novas falas em uma continuação, treinar um modelo que gere publicidade ou licenciar uma réplica vocal a terceiros. O problema não está apenas na existência de uma assinatura contratual, mas na extensão da finalidade autorizada. Quanto mais distante o novo uso estiver da expectativa que justificou a coleta original, maior a necessidade de examinar especificidade, informação e equilíbrio contratual.

O mesmo raciocínio vale para a imagem. Fotografar alguém para uma campanha não implica, necessariamente, poder construir uma versão sintética daquela pessoa e fazê-la aparecer em cenários nunca contratados. A transformação tecnológica altera

qualitativamente a exploração da identidade: deixa-se de reutilizar um registro para produzir novas performances. Essa diferença recomenda que contratos envolvendo voz, imagem e modelos digitais delimitem finalidades, duração, meios de exploração, possibilidade de treinamento, sublicenciamento, usos póstumos, mecanismos de retirada e formas de atualização do consentimento. O Direito não precisa criar um novo direito de personalidade para perceber que a autorização antiga pode não cobrir uma nova forma de apropriação.

Também não se pode presumir que toda utilização não autorizada de imagem ou voz seja automaticamente censurável. A liberdade de expressão, informação, criação artística e crítica possui proteção constitucional igualmente robusta. Na Ação Direta de Inconstitucionalidade (ADI) 4.815, o Supremo Tribunal Federal (STF) afastou a exigência de autorização prévia para biografias, interpretando os arts. 20 e 21 do Código Civil de forma compatível com as liberdades comunicativas, sem eliminar a possibilidade de responsabilização posterior por abusos (STF, 2015). O precedente não trata de deepfakes, mas impede uma solução excessivamente simples: direitos da personalidade não conferem poder absoluto de controlar toda representação pública da pessoa.

A tensão aparece com clareza em sátiras, paródias e obras artísticas. Uma representação sintética pode ser justamente o recurso expressivo utilizado para criticar uma figura pública ou construir ficção. Nesses casos, a análise deve considerar se o contexto deixa suficientemente clara a natureza criativa da obra, se há interesse público, se existe intenção de enganar, se a pessoa é exposta a dano desproporcional e se o uso ultrapassa o que pode ser protegido pela liberdade artística. Rotular todo deepfake como fraude sacrificaria expressão legítima; tratar toda criação como liberdade artística deixaria direitos da personalidade sem proteção efetiva.

A qualidade do aviso ao público faz diferença. Um vídeo evidentemente satírico possui capacidade de engano distinta de uma gravação produzida para simular vazamento clandestino. Um clone de voz usado em peça teatral não ocupa a mesma posição de uma chamada telefônica que imita um familiar para solicitar transferência bancária. Em ambos os casos existe conteúdo sintético; apenas no segundo a autenticidade aparente é instrumento central do golpe. A finalidade, portanto, não é detalhe subjetivo. Ela ajuda a definir quais bens jurídicos estão em risco e quais remédios são adequados.

A vulnerabilidade do representado também importa. Crianças, pessoas idosas e indivíduos sem projeção pública podem sofrer danos particularmente difíceis de reparar quando sua identidade é transformada em conteúdo viral. A circulação em massa altera a escala do dano: uma imagem pode ser copiada, recortada, reenviada, indexada e reaparecer depois de removida da publicação original. Monteiro e Carrá (2026) observam que danos digitais desafiam a responsabilidade civil justamente porque violações da honra, privacidade, imagem e reputação assumem características próprias no ambiente virtual. No caso dos deepfakes, velocidade e replicabilidade fazem com que tutela inibitória e remoção rápida possam ser tão importantes quanto indenização posterior.

Esse ponto também explica por que direito à personalidade e proteção de dados não devem ser tratados como sinônimos. Uma pessoa pode ter sua imagem violada por conteúdo criado sem tratamento relevante de dados pessoais pelo réu; da mesma forma, pode haver tratamento irregular de dados biométricos sem que qualquer deepfake seja publicado. Os regimes se sobrepõem em certas situações, mas protegem dimensões diferentes. Manter essa distinção é importante para evitar que a LGPD seja convertida em resposta universal para todo problema de identidade digital.

Há ainda situações envolvendo pessoas falecidas. O Código Civil reconhece legitimidade de familiares para requerer proteção em hipóteses relacionadas à imagem de morto ou ausente, e o desenvolvimento de avatares póstumos torna essa previsão cada vez menos teórica (BRASIL, 2002). Uma reconstrução histórica autorizada por herdeiros, um documentário e uma publicidade que faz uma pessoa falecida "recomendar" um produto levantam questões distintas. A morte não transforma a identidade em recurso livre de qualquer limite, mas tampouco autoriza concluir que toda reconstrução digital dependa necessariamente de consentimento familiar, sobretudo quando liberdade artística, memória e interesse histórico estiverem presentes.

O critério mais seguro é resistir a respostas absolutas. Em vez de perguntar apenas se houve autorização, é preciso reconstruir qual autorização existia, para qual finalidade, em que contexto, com que expectativa de audiência e com qual potencial de dano. Essa análise respeita uma característica essencial dos direitos da personalidade: eles protegem a pessoa sem convertê-la em proprietária de todo discurso social no qual sua identidade apareça.

Do dano individual à confiança pública

Os conflitos envolvendo deepfakes deixam de ser exclusivamente privados quando a manipulação da identidade interfere em processos coletivos. O exemplo mais desenvolvido no Brasil é o Direito Eleitoral. A Resolução nº 23.610/2019 do Tribunal Superior Eleitoral (TSE), modificada em 2024 e novamente atualizada pela Resolução nº 23.755/2026 para as Eleições Gerais de 2026, estabelece deveres específicos de informação para conteúdo sintético e proíbe, para prejudicar ou favorecer candidatura, deepfakes em áudio, vídeo ou combinação de ambos que criem, substituam ou alterem imagem ou voz de pessoa viva, falecida ou fictícia (TSE, 2026). O regime é mais restritivo que a tutela civil geral porque o bem protegido inclui a integridade do processo eleitoral.

A atualização de 2026 foi além. Entre outras medidas, proibiu a publicação, republicação e o impulsionamento de novos conteúdos sintéticos produzidos ou alterados por IA que utilizem imagem, voz ou manifestação de candidatura ou pessoa pública no período entre 72 horas antes e 24 horas depois do término do pleito, ainda que o conteúdo esteja rotulado. A norma também admite inversão motivada do ônus da prova em determinadas representações quando a comprovação técnica da manipulação digital for excessivamente onerosa ao autor (TSE, 2026). Essa escolha é juridicamente reveladora: o problema não é apenas definir o ilícito, mas enfrentar a assimetria probatória criada por uma tecnologia cuja falsificação pode exigir perícia sofisticada.

A experiência eleitoral também mostra por que transparência não é uma solução única. Em alguns usos, informar que uma peça foi criada com IA pode ser suficiente para evitar que o público confunda ficção com registro factual. Em outros, a própria finalidade do conteúdo é ilícita e nenhum rótulo a torna legítima. A distinção entre dever de revelar e proibição de utilizar aparece igualmente no AI Act europeu. O art. 50 do AI Act impõe obrigações de marcação técnica a fornecedores de sistemas que geram conteúdo sintético e exige, de quem utiliza sistemas para produzir ou manipular deepfakes, divulgação de que o conteúdo foi artificialmente gerado ou manipulado, com tratamento específico para obras evidentemente artísticas, criativas, satíricas ou ficcionais (UNIÃO EUROPEIA, 2024a).

Essa abordagem reconhece uma limitação prática: detectar deepfakes apenas olhando para o arquivo será cada vez menos confiável. A resposta tende a combinar rotulagem, metadados, mecanismos de proveniência, preservação do original e

contexto de publicação. Nenhuma dessas medidas é infalível. Marcas podem ser removidas, metadados podem desaparecer e detectores podem perder desempenho quando modelos geradores mudam. Por isso, autenticidade não deve depender de um único "detector de IA". No plano jurídico, o mais importante é construir um conjunto de sinais capaz de sustentar ou enfraquecer a confiança no conteúdo.

Esse problema já toca diretamente o campo da prova digital e produz consequências imediatas na tutela dos direitos da personalidade. Uma pessoa atingida por vídeo falso pode precisar demonstrar rapidamente que nunca participou da gravação. O ônus social da dúvida recai, muitas vezes, primeiro sobre a vítima. Mesmo depois de uma perícia confirmar a manipulação, cópias podem continuar circulando. O dano, assim, não depende apenas da capacidade de enganar um juiz; depende da velocidade com que a falsidade entra no espaço público e da dificuldade de retirar dela a aparência inicial de autenticidade.

As plataformas digitais ocupam posição relevante nesse percurso porque ampliam ou reduzem a circulação do conteúdo. O regime brasileiro de responsabilidade dos provedores sofreu alteração significativa com o julgamento conjunto dos Temas 533 e 987 da repercussão geral, concluído em junho de 2025. Em 17 de junho de 2026, ao concluir o julgamento de embargos no Tema 987, o STF ajustou a tese e manteve o reconhecimento da inconstitucionalidade parcial e progressiva do art. 19 do Marco Civil da Internet, com regimes diferenciados conforme a natureza do conteúdo e do serviço e ressalva às regras próprias da legislação eleitoral (STF, 2026). O resultado impede que a discussão sobre deepfakes seja resolvida por uma fórmula única de "só remover com ordem judicial" ou, no extremo oposto, por dever irrestrito de vigilância prévia.

Essa cautela é necessária porque plataformas não criam todos os conteúdos que hospedam, mas podem participar de sua disseminação, recomendação e impulsionamento. A resposta jurídica precisa distinguir autoria do deepfake, disponibilização de ferramentas, contratação de publicidade, denúncia de perfil inautêntico, ciência do ilícito e comportamento posterior da plataforma. A responsabilidade de quem sintetiza a voz de uma vítima não se confunde com a responsabilidade de quem hospeda o arquivo; ambas, por sua vez, diferem da posição de quem financia a distribuição. Separar esses papéis é essencial para não diluir a responsabilidade nem impor a um único agente obrigações incompatíveis com sua capacidade real de controle.

Há ainda um interesse coletivo em preservar ambientes nos quais registros autênticos continuem capazes de produzir confiança. Se qualquer vídeo inconveniente puder ser descartado como possível deepfake, a tecnologia passa a servir também como mecanismo de negação. Se, por outro lado, qualquer gravação realista for recebida como verdadeira, falsificações tornam-se instrumentos de chantagem, fraude, perseguição e manipulação política. O Direito precisa operar entre esses extremos, protegendo a reputação de quem é falsamente representado sem instituir uma presunção geral de falsidade sobre toda mídia digital.

A melhor resposta, portanto, não está em perseguir uma impossível pureza informacional. Está em criar condições para que identidade sintética seja reconhecível quando necessário, que usos sem consentimento possam ser interrompidos quando violarem direitos, que vítimas disponham de vias rápidas de proteção e que liberdade de expressão continue preservada em usos legítimos. O equilíbrio não será obtido por um único dispositivo legal. Ele depende da articulação entre a Constituição Federal, os direitos da personalidade, a proteção de dados, a responsabilidade civil, as regras de plataforma, as normas eleitorais e as garantias processuais.

É justamente na passagem do resultado para o percurso que a proteção de dados se torna inevitável. Para fabricar uma voz, reconstruir um rosto ou personalizar um avatar, sistemas precisam receber, extrair, armazenar ou aprender a partir de informações ligadas a pessoas. O deepfake mostra o dano no resultado final; a proteção de dados obriga o Direito a olhar para o que veio antes. Quem podia coletar aquele material? Para qual finalidade? O dado poderia ser reutilizado para treinar um modelo? Que deveres surgem quando informações pessoais deixam de apenas descrever alguém e passam a permitir que sistemas produzam novas inferências e representações em seu nome? A identidade sintetizada, assim, conduz ao problema mais amplo do tratamento automatizado de dados pessoais.

CAPÍTULO 3

PROTEÇÃO DE DADOS PESSOAIS E TRATAMENTO AUTOMATIZADO DE INFORMAÇÕES

Dados não apenas descrevem: também permitem inferir

A discussão sobre proteção de dados muda de natureza quando aplicada à inteligência artificial. Em sistemas tradicionais, era intuitivo imaginar uma base de dados como conjunto de informações que alguém forneceu ou que uma instituição registrou: nome, endereço, histórico de compras, localização, renda, diagnóstico, fotografia. Modelos de aprendizado de máquina, entretanto, não se limitam a recuperar aquilo que já está armazenado. Eles procuram relações entre variáveis e produzem inferências: probabilidade de inadimplência, afinidade política, risco de fraude, propensão a determinada doença, perfil de consumo ou chance de abandonar um serviço. O dado juridicamente relevante, portanto, não é apenas o que entrou no sistema. Também importa aquilo que o sistema consegue concluir a respeito da pessoa.

Essa diferença toca o núcleo da Lei Geral de Proteção de Dados Pessoais. A LGPD adota conceito amplo de dado pessoal: informação relacionada a pessoa natural identificada ou identificável. A lei não exige que a informação tenha sido declarada pelo titular nem que revele diretamente seu nome. Se uma inferência puder ser relacionada a uma pessoa identificada ou identificável e for utilizada para classificá-la, avaliá-la ou orientar decisões a seu respeito, a proteção jurídica não desaparece apenas porque o conteúdo foi "descoberto" pelo modelo e não preenchido em um formulário (BRASIL, 2018). A inteligência artificial torna particularmente visível essa passagem do dado fornecido para o dado derivado.

Wachter e Mittelstadt (2019) chamam atenção para esse ponto ao mostrar que sistemas de big data e IA produzem inferências muitas vezes não intuitivas, difíceis de verificar e capazes de afetar privacidade, reputação e autonomia. Uma pessoa pode revelar informações aparentemente banais e, pela combinação com padrões extraídos de grandes populações, ser associada a características que nunca decidiu comunicar. No Brasil, Magrani e Miranda (2025) aproximam o problema do direito fundamental à proteção de dados e defendem que inferências utilizadas em decisões relevantes precisam ser justificáveis, transparentes e compatíveis com a singularidade do indivíduo. O desafio não é impedir toda inferência, algo incompatível com grande

parte da utilidade da análise de dados, mas impedir que correlações opacas adquiram autoridade automática sobre a vida das pessoas.

A distinção entre dado pessoal e dado pessoal sensível amplia essa preocupação. A LGPD atribui proteção reforçada a informações sobre origem racial ou étnica, convicção religiosa, opinião política, filiação sindical ou a organização de caráter religioso, filosófico ou político, saúde, vida sexual e dados genéticos ou biométricos vinculados a pessoa natural (BRASIL, 2018). Em sistemas de IA, porém, uma categoria sensível pode ser inferida a partir de variáveis que, isoladamente, não pertencem a esse rol. Padrões de consumo, localização, linguagem, rede de contatos ou histórico de navegação podem funcionar como sinais indiretos de condições íntimas. Retirar a coluna "religião" ou "saúde" de uma base não garante, por si só, que o modelo deixe de reconstruir informação equivalente por outras vias.

Essa possibilidade mostra por que a proteção de dados não pode ser reduzida à identificação de campos sensíveis em uma planilha. O risco depende do conjunto de informações, da capacidade analítica aplicada e da finalidade. Um CEP pode ser apenas informação geográfica em determinado contexto; em outro, pode funcionar como variável substituta para renda, raça ou vulnerabilidade social. Um horário de acesso pode parecer irrelevante, mas, combinado com localização e comportamento, contribuir para inferências sobre rotina de trabalho ou condição de saúde. O Direito precisa avaliar o tratamento como processo, e não apenas a natureza abstrata de cada variável.

Também é equivocada a ideia de que informação disponível publicamente deixou de ser protegida. A LGPD disciplina o tratamento de dados cujo acesso é público e exige consideração da finalidade, da boa-fé e do interesse público que justificaram sua disponibilização, além de preservar os princípios e direitos dos titulares (BRASIL, 2018). Esse detalhe ganha enorme importância em modelos treinados a partir de conteúdos coletados em larga escala na internet. Uma fotografia aberta, uma postagem antiga ou um comentário em fórum público podem ser acessíveis tecnicamente; disso não decorre uma autorização geral para qualquer reutilização futura.

O caso envolvendo a política de privacidade da Meta, em 2024, materializou esse conflito no Brasil. A ANPD suspendeu cautelarmente o uso de dados pessoais para treinamento de sistemas de IA generativa ao identificar, em análise preliminar, problemas relacionados à hipótese legal invocada, transparência, exercício de direitos e

proteção de crianças e adolescentes. A Autoridade observou, entre outros pontos, que conteúdos compartilhados em redes sociais ao longo de anos não eram necessariamente publicados sob a legítima expectativa de que seriam empregados para treinar modelos que sequer existiam quando o material foi disponibilizado (ANPD, 2024b). Posteriormente, a suspensão foi levantada após aprovação de um plano de conformidade com restrições, deveres de informação e mecanismo facilitado de oposição, sem que isso convertesse o caso em uma declaração geral de que todo treinamento baseado em conteúdo público é lícito (ANPD, 2024c).

O episódio é importante porque desloca a pergunta de "o dado estava disponível?" para "o novo tratamento é compatível com a relação em que o dado surgiu?". Essa mudança é coerente com os princípios da finalidade, adequação, necessidade, transparência e não discriminação previstos na LGPD. A inovação tecnológica não apaga o contexto original da coleta. Quanto mais distante a nova finalidade estiver das expectativas razoáveis do titular, maior será a necessidade de justificar juridicamente a reutilização, informar de forma compreensível e criar salvaguardas proporcionais ao risco.

Proteção de dados, portanto, não significa congelar informações no instante em que foram coletadas. Dados podem ser reutilizados quando existe fundamento jurídico e compatibilidade com as regras aplicáveis. O que a LGPD rejeita é a transformação da disponibilidade técnica em licença irrestrita. Na inteligência artificial, essa distinção é decisiva porque o valor econômico dos dados frequentemente está justamente em usos que não eram evidentes no momento da coleta. O sistema encontra relações novas; o Direito precisa perguntar se a organização tinha legitimidade para procurá-las e utilizá-las.

Treinar também é tratar dados

Um erro recorrente na discussão pública sobre inteligência artificial é tratar o treinamento do modelo como etapa neutra, anterior ao uso propriamente dito. Juridicamente, isso não se sustenta quando há operações realizadas com dados pessoais. A LGPD define tratamento de forma ampla, abrangendo coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, processamento, arquivamento, armazenamento, avaliação, modificação, comunicação, transferência, difusão e extração (BRASIL, 2018). Preparar uma base, selecionar registros, transformar textos em representações matemáticas e ajustar parâmetros a partir desses dados fazem parte de um processo de tratamento e precisam ser examinados como tal.

O primeiro passo é identificar uma hipótese legal adequada. Consentimento é uma delas, mas não é a única e não deve ser apresentado como solução automática. Dependendo da situação, a LGPD admite outras bases legais para dados pessoais não sensíveis, como cumprimento de obrigação legal, execução de contrato, exercício regular de direitos, proteção do crédito e legítimo interesse. Dados sensíveis obedecem a um regime próprio e mais restritivo. O legítimo interesse, por exemplo, não funciona como fundamento para tratamento de dados pessoais sensíveis e exige análise concreta de finalidade, necessidade, expectativas legítimas e salvaguardas (BRASIL, 2018; ANPD, 2024a).

O Guia Orientativo sobre Legítimo Interesse da ANPD propõe um teste de balanceamento estruturado em finalidade, necessidade e balanceamento com salvaguardas. Essa abordagem é particularmente útil para sistemas de IA porque impede que "inovação" seja utilizada como justificativa abstrata. O controlador precisa demonstrar qual interesse pretende atender, por que aquele tratamento é necessário e se os direitos e liberdades do titular prevalecem no caso concreto (ANPD, 2024a). Uma empresa não satisfaz esse dever apenas afirmando que mais dados tornarão o modelo melhor. A necessidade jurídica não se confunde com conveniência estatística.

Essa distinção coloca em tensão a cultura de aprendizado de máquina e o princípio da necessidade. Em engenharia, acumular mais variáveis e mais exemplos pode parecer desejável porque amplia possibilidades de modelagem. Em proteção de dados, deve-se limitar o tratamento ao mínimo necessário para alcançar finalidades legítimas, com dados pertinentes, proporcionais e não excessivos (BRASIL, 2018). Não existe contradição inevitável entre os dois campos, mas existe uma obrigação de projeto: demonstrar por que determinada informação contribui de maneira justificável para a finalidade e quais dados podem ser excluídos sem comprometer o objetivo.

O problema se torna mais difícil em modelos de propósito geral. Diferentemente de um classificador criado para uma tarefa claramente delimitada, modelos generativos podem ser treinados sobre grandes conjuntos de texto, imagem, áudio e código para desempenhar usos que serão definidos somente depois. Essa abertura desafia o princípio da finalidade, pois é mais difícil especificar antecipadamente todas as aplicações futuras. O fato de a arquitetura permitir múltiplos usos não elimina a necessidade de governança. Pelo contrário: aumenta a importância de documentar

origem dos dados, categorias envolvidas, critérios de inclusão e exclusão, mecanismos de oposição, retenção e condições de reutilização.

O risco não termina quando o treinamento acaba. Pesquisas em segurança demonstram que modelos podem revelar informações sobre os dados utilizados para treiná-los. Shokri et al. (2017) mostraram que ataques de inferência de pertencimento podem estimar se determinado registro participou da base de treinamento. Fredrikson, Jha e Ristenpart (2015) demonstraram ataques de inversão capazes de explorar saídas de modelos para recuperar atributos sensíveis em certos contextos. Em modelos de linguagem, Carlini et al. (2021) extraíram sequências memorizadas do conjunto de treinamento do GPT-2, incluindo informações pessoalmente identificáveis presentes em dados públicos. Esses trabalhos não significam que todo modelo necessariamente exponha seus registros, mas demonstram que "treinar e descartar a base" não encerra automaticamente o risco de privacidade.

Esse ponto é particularmente relevante para anonimização. A LGPD exclui, em regra, do conceito de dado pessoal o dado anonimizado, salvo quando o processo puder ser revertido com meios próprios ou com esforços razoáveis, considerando fatores objetivos como custo e tempo necessários e tecnologias disponíveis (BRASIL, 2018). Pseudonimizar identificadores ou remover nome e CPF não equivale, portanto, a tornar um conjunto definitivamente anônimo. Bases ricas podem permitir reidentificação por combinação de atributos; modelos podem memorizar exemplos; e avanços técnicos podem alterar aquilo que é razoavelmente reversível.

A própria agenda regulatória da ANPD para 2025-2026 mantém anonimização e pseudonimização, inteligência artificial, tratamento de alto risco e dados biométricos entre os temas de desenvolvimento normativo. Em julho de 2026, o primeiro relatório parcial do Sandbox Regulatório em IA apontou, inclusive, necessidade de aperfeiçoar protocolos relacionados a dados sintéticos; em agosto, a metodologia de testagem passou a exigir análise de arquitetura, fluxos de dados, controles de segurança, tratamento de dados pessoais, transparência e produção de evidências ao longo de ciclos sucessivos (ANPD, 2026b). O movimento regulatório confirma que técnicas de proteção não devem ser tratadas como rótulos, mas como medidas cuja eficácia precisa ser demonstrada.

Privacidade desde a concepção ganha, nesse cenário, significado concreto. Não basta publicar um aviso de privacidade depois que o sistema está pronto. Decisões sobre arquitetura, origem dos dados, retenção, acesso, treinamento, possibilidade de

exclusão e mecanismos de segurança precisam ser consideradas antes da implantação. Um modelo que depende estruturalmente de coleta excessiva ou cuja documentação não permite distinguir fontes legítimas de fontes problemáticas cria um passivo difícil de corrigir posteriormente.

Relatórios de impacto cumprem função semelhante. A LGPD prevê a possibilidade de a autoridade determinar sua elaboração e define o relatório de impacto à proteção de dados pessoais como documentação do controlador que descreve processos capazes de gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação (BRASIL, 2018). Mantelero (2018) propõe que avaliações em contextos de IA e big data ultrapassem a segurança da informação e considerem efeitos sobre direitos humanos, valores sociais e impactos coletivos. Essa ampliação é especialmente adequada quando um sistema classifica populações inteiras e pode produzir desvantagens que nenhum titular consegue perceber isoladamente.

O tratamento de crianças e adolescentes merece cautela ainda maior. A LGPD determina que seu tratamento ocorra em seu melhor interesse, e a fiscalização da ANPD sobre treinamento de modelos em plataformas digitais reforçou que não é suficiente importar práticas pensadas para adultos. Tanto no caso Meta quanto na medida preventiva dirigida à X. Corp, a Autoridade exigiu salvaguardas específicas para impedir treinamento de IA generativa com dados de contas de menores enquanto questões regulatórias permaneciam sob análise (ANPD, 2024b; ANPD, 2024d). O aumento do valor econômico de conteúdos digitais não reduz o dever de proteção de grupos vulneráveis.

Treinar um modelo, em suma, não é um espaço juridicamente vazio entre coleta e produto final. É uma cadeia de operações em que finalidade, hipótese legal, minimização, segurança, proteção de grupos vulneráveis e capacidade de demonstrar conformidade precisam acompanhar o desenvolvimento. Quanto mais geral e poderoso o sistema, menos defensável se torna a ideia de que ninguém precisa saber exatamente de onde vieram os dados ou como foram utilizados.

Decisão automatizada precisa continuar contestável

Mesmo um tratamento de dados formalmente regular pode produzir problema jurídico no momento em que a informação é transformada em decisão. A LGPD reconhece essa diferença ao dedicar regra específica às decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem interesses dos

titulares. O art. 20 da LGPD assegura o direito de solicitar revisão dessas decisões, inclusive das destinadas a definir perfis pessoais, profissionais, de consumo e de crédito ou aspectos da personalidade (BRASIL, 2018). A norma desloca a proteção do banco de dados para o efeito: não importa apenas saber quais informações foram coletadas, mas compreender como elas participaram de uma decisão relevante.

O dispositivo precisa ser lido com precisão. A redação vigente garante o direito de solicitar revisão, mas não contém a expressão originalmente prevista que obrigava essa revisão a ser realizada por pessoa natural. A retirada desse trecho durante o processo legislativo não elimina a necessidade de contestabilidade, tampouco permite concluir que qualquer "revisão automática" satisfaça o direito em todas as situações. O problema jurídico é funcional: o mecanismo oferecido precisa ser capaz de reexaminar a decisão, corrigir erro e fornecer ao titular uma via real de questionamento.

A LGPD também determina que o controlador forneça, quando solicitado, informações claras e adequadas a respeito dos critérios e procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial. Se essas informações não forem fornecidas sob essa justificativa, a ANPD pode realizar auditoria para verificação de aspectos discriminatórios do tratamento automatizado (BRASIL, 2018). A lei, portanto, não escolhe entre transparência absoluta e segredo empresarial absoluto. Reconhece ambos e cria um mecanismo institucional para que o sigilo não inviabilize completamente o controle.

Isso ajuda a separar explicabilidade de exposição do código-fonte. Em muitos casos, o titular não precisa conhecer milhões de parâmetros internos para contestar uma decisão. Precisa saber quais categorias de informação foram relevantes, que objetivo orientava o processo, que critérios gerais influenciaram o resultado, quais fontes estavam envolvidas e como apresentar elementos capazes de corrigi-lo. Uma explicação tecnicamente detalhada pode ser juridicamente inútil se não permitir ação; uma informação mais simples pode ser adequada quando mostra de forma fiel o caminho para contestação.

A dificuldade aumenta quando decisões não são formalmente "unicamente automatizadas". Sistemas de IA frequentemente operam como filtros, rankings e recomendações que moldam aquilo que o decisor humano verá. Um recrutador pode receber apenas os dez currículos mais bem classificados; um analista de crédito pode enxergar primeiro a nota de risco; um servidor público pode ser direcionado para casos priorizados por algoritmo. A assinatura final é humana, mas a tecnologia já organizou

o espaço de decisão. Reduzir a proteção jurídica à existência física de uma pessoa no fim do processo criaria um incentivo à supervisão apenas simbólica.

A ANPD vem tratando exatamente dessas zonas de incerteza. A Tomada de Subsídios sobre inteligência artificial e revisão de decisões automatizadas, cujos resultados foram publicados em 2025, reuniu contribuições sobre princípios, hipóteses legais, direitos dos titulares, boas práticas e governança. Em 2026, a Agência manteve a transparência algorítmica como eixo de seu sandbox e publicou metodologia de testagem voltada a explicabilidade, fluxos de dados, controles de segurança, riscos e produção de evidências (ANPD, 2025a; ANPD, 2026a; ANPD, 2026b). O fato de esses temas permanecerem em construção regulatória é importante: não convém apresentar como regra fechada aquilo que ainda está sendo operacionalizado.

Contestabilidade também depende da qualidade das inferências. Um modelo pode usar dados corretos e, ainda assim, produzir uma conclusão inadequada porque a correlação não representa bem aquele indivíduo. Wachter e Mittelstadt (2019) denominam de alto risco as inferências capazes de afetar privacidade ou reputação e que apresentam baixa verificabilidade ou são utilizadas em decisões importantes. Magrani e Miranda (2025) trazem essa discussão para o contexto brasileiro e sustentam que proteção de dados e livre desenvolvimento da personalidade oferecem fundamento para exigir razoabilidade das inferências. A ideia é valiosa porque impede que uma probabilidade estatística seja tratada como descrição incontestável da pessoa.

Esse problema se conecta diretamente ao princípio da não discriminação da LGPD. Tratamentos não podem ser realizados para fins discriminatórios ilícitos ou abusivos (BRASIL, 2018). A proibição não se limita a modelos que utilizem explicitamente raça, gênero ou outra categoria protegida. Um sistema pode reproduzir desigualdades por meio de variáveis substitutas, amostras desequilibradas ou rótulos históricos. Por isso, governança de dados e avaliação algorítmica precisam observar desempenho entre grupos e consequências concretas, não apenas acurácia média.

A responsabilização exige ainda que a organização saiba reconstruir o próprio processo. Se um titular contesta uma negativa de crédito ou exclusão de oportunidade, o controlador precisa identificar que versão do modelo operou, quais dados estavam associados à pessoa, que lógica geral foi aplicada e se houve intervenção humana. Sem rastreabilidade, o direito à revisão corre o risco de se tornar formalidade.

A documentação, nesse sentido, não serve apenas ao regulador; é condição para que a própria instituição consiga explicar e corrigir suas decisões.

O princípio da responsabilização e prestação de contas da LGPD exige demonstração de medidas eficazes capazes de comprovar observância das normas de proteção de dados (BRASIL, 2018). Em IA, esse dever tende a ganhar densidade: inventário de dados, registro de finalidades, avaliação de base legal, testes de risco, documentação de versões, canais para exercício de direitos, monitoramento de incidentes e critérios para suspender sistemas deixam de ser burocracia periférica. São a infraestrutura jurídica que permite utilizar automação sem transformar complexidade técnica em zona de irresponsabilidade.

Deepfakes mostraram que dados podem ser utilizados para reconstruir a identidade de alguém. Aqui, a análise revela que dados também podem ser reutilizados para treinar modelos, produzir inferências e orientar decisões que o titular talvez jamais perceba. Em ambos os casos, a proteção exige acompanhar o percurso completo da informação, e não apenas o momento em que o dano se torna visível.

A questão seguinte surge justamente quando esse percurso produz prejuízo. Se um sistema treinado com dados inadequadamente selecionados nega uma oportunidade, recomenda uma conduta errada ou causa perda econômica, saber que houve tratamento de dados é apenas o início. Será necessário identificar quem possuía dever de cuidado, qual conduta contribuiu para o dano, como se estabelece o nexo causal e até onde desenvolvedores, fornecedores, operadores e usuários podem responder. A proteção de dados mostra como o sistema conheceu e classificou a pessoa; a responsabilidade civil perguntará quem deve responder quando essa classificação produz dano.

CAPÍTULO 4

RESPONSABILIDADE CIVIL POR DANOS DECORRENTES DE SISTEMAS DE INTELIGÊNCIA ARTIFICIAL

O dano não é causado pela IA em abstrato

Dizer que "a inteligência artificial causou um dano" pode ser uma descrição conveniente, mas é uma explicação juridicamente incompleta. Antes que um sistema produza uma saída capaz de lesar alguém, existe uma cadeia de escolhas: uma finalidade foi definida, dados foram selecionados, um modelo foi desenvolvido ou adquirido, parâmetros foram configurados, uma interface foi desenhada, o sistema foi integrado a determinado serviço e alguém decidiu quanto peso atribuir ao resultado. A responsabilidade civil começa justamente quando se abandona a imagem da máquina como agente isolado e se reconstrói essa cadeia de decisões humanas e organizacionais.

Essa reconstrução é importante porque o ordenamento brasileiro já possui regimes capazes de responder a grande parte dos danos associados à inteligência artificial. O Código Civil disciplina o ato ilícito, o abuso de direito, a obrigação de reparar e, em determinadas hipóteses, a responsabilidade independentemente de culpa. O Código de Defesa do Consumidor estabelece responsabilidade por defeitos de produtos e serviços e vincula o conceito de defeito à segurança que legitimamente se pode esperar. Nenhuma dessas normas exige que o dano seja produzido por uma tecnologia simples. A novidade da IA está menos na ausência de regras e mais na dificuldade de identificar qual regra incide, quem ocupava posição de controle e como demonstrar o nexo entre falha e prejuízo (BRASIL, 1990; BRASIL, 2002; TEPEDINO; SILVA, 2019).

Por isso, a primeira pergunta não deve ser se a responsabilidade é "da IA", mas qual relação jurídica existia. Um sistema oferecido ao consumidor como parte de um serviço financeiro, de saúde, educação, transporte ou comércio eletrônico pode submeter-se ao regime consumerista. Uma ferramenta desenvolvida internamente por empresa e utilizada em relação que não seja de consumo pode exigir análise pelo Código Civil ou por legislação setorial. Um sistema empregado pelo Estado envolve ainda deveres próprios de legalidade e atuação administrativa. O simples fato de existir um algoritmo não substitui essa qualificação jurídica.

Tampouco existe razão para transformar a autonomia funcional do sistema em personalidade jurídica. Modelos podem gerar resultados não redigidos previamente

por qualquer programador, adaptar-se a padrões e produzir comportamentos difíceis de antecipar em detalhe. Isso não significa que se tornem titulares autônomos de deveres patrimoniais. Tepedino e Silva (2019) alertam contra soluções assistemáticas que, diante da novidade tecnológica, abandonem a dogmática da responsabilidade civil. Em lugar de criar um sujeito eletrônico para receber a culpa, é mais útil perguntar quem introduziu o risco, quem podia preveni-lo e quem obteve benefício econômico da atividade.

Esse ponto também afasta a ideia de que imprevisibilidade técnica equivale automaticamente a ausência de responsabilidade. Em aprendizado de máquina, pode ser impossível antecipar cada saída concreta, mas isso é diferente de desconhecer que o sistema possui limites, taxas de erro, condições de uso e riscos de falha. Um fornecedor que sabe que o modelo perde desempenho fora de determinado contexto não pode transformar essa limitação em surpresa depois do dano. O que importa não é se alguém previu exatamente o resultado, mas se o tipo de risco era conhecido ou deveria ter sido gerenciado, documentado e comunicado.

A culpa continua relevante em muitas situações. Desenvolver sem testes minimamente adequados, ignorar alertas conhecidos, não corrigir vulnerabilidade grave, utilizar modelo fora da finalidade prevista ou substituir análise humana indispensável por confiança cega podem configurar violações de deveres de cuidado, a depender do caso. Mas a inteligência artificial também aparece em relações nas quais a responsabilidade objetiva já resulta da lei. Por isso, tentar escolher previamente entre "culpa" e "risco" para toda IA produz mais confusão do que segurança. Fonseca (2021) e Melo (2024) convergem ao mostrar que a heterogeneidade das aplicações impede solução uniforme e exige articulação com regimes já existentes.

O parágrafo único do art. 927 do Código Civil permite responsabilidade independentemente de culpa quando a atividade normalmente desenvolvida pelo autor do dano implicar, por sua natureza, risco para direitos de terceiros. Isso não autoriza concluir que todo uso de IA seja atividade de risco. Um editor de texto com recurso de correção automática e um sistema que controla equipamento capaz de causar lesão física não apresentam a mesma exposição. A cláusula geral precisa ser aplicada a partir da atividade concreta e da intensidade do risco, não do prestígio ou do temor associado ao rótulo "inteligência artificial" (BRASIL, 2002; TEPEDINO; SILVA, 2019).

Essa diferenciação protege simultaneamente vítima e inovação. Se qualquer presença de IA gerasse responsabilidade objetiva universal, sistemas de baixo impacto receberiam o mesmo tratamento de aplicações perigosas. Se toda falha fosse tratada como imprevisível porque resultou de aprendizado automatizado, a complexidade tecnológica funcionaria como escudo contra reparação. O desafio é outro: escolher o critério de imputação coerente com o risco, o vínculo jurídico e a posição dos agentes.

A mesma cautela vale para o dano. Sistemas podem produzir prejuízos patrimoniais, lesões físicas, violações de honra, imagem e privacidade, discriminação, perda de oportunidade ou danos coletivos. A categoria tecnológica não determina a espécie de prejuízo. Monteiro e Carrá (2026), ao examinarem danos digitais, mostram que o ambiente virtual altera escala, persistência e formas de propagação da lesão, mas não elimina a necessidade de demonstrar juridicamente o interesse atingido. Responsabilidade civil não é punição abstrata por usar tecnologia arriscada; é resposta a dano imputável segundo determinado regime.

Defeito, risco e nexo em uma cadeia tecnológica

Nas relações de consumo, o conceito de defeito oferece uma ferramenta especialmente útil. Pelo CDC, produto ou serviço é defeituoso quando não oferece a segurança que legitimamente se espera, consideradas as circunstâncias de fornecimento, os riscos razoavelmente esperados e, conforme o caso, a época em que foi colocado em circulação. A responsabilidade objetiva não significa indenização automática. Continuam necessários dano, defeito e relação causal, e a lei prevê hipóteses de afastamento da responsabilidade. A vantagem do regime está em não exigir que o consumidor descubra e prove a culpa interna de uma cadeia produtiva à qual normalmente não tem acesso (BRASIL, 1990).

Em inteligência artificial, a noção de segurança precisa incluir mais do que ausência de pane. Um sistema pode funcionar exatamente como foi programado e ainda ser defeituoso se o modo de fornecimento cria expectativa de confiabilidade incompatível com seus limites. Pode haver defeito de concepção, de informação ou de integração: dados inadequados para o público-alvo, falta de barreiras para uso previsivelmente perigoso, interface que induz confiança excessiva, ausência de alerta sobre taxa de erro relevante, atualização que degrada desempenho ou conexão mal realizada com outro componente. A falha não precisa aparecer como tela quebrada ou código que deixa de executar.

Essa abordagem impede que o fornecedor se defenda apenas afirmando que "o algoritmo fez o que deveria". O que deve ser avaliado é se o produto ou serviço, como oferecido e inserido na prática, entregava a segurança legitimamente esperada. A nova Diretiva europeia sobre responsabilidade por produtos defeituosos segue lógica semelhante ao atualizar o regime para tecnologias digitais: considera software dentro da disciplina de produtos e admite que defeitos possam surgir após a colocação no mercado quando atualizações, serviços relacionados ou mecanismos de aprendizado permaneçam sob controle do fabricante (UNIÃO EUROPEIA, 2024b). O modelo europeu não se aplica ao Brasil, mas ilustra como a responsabilidade acompanha cada vez mais o ciclo de vida digital, e não apenas o instante da venda.

A dificuldade maior costuma surgir no nexo causal. Se um sistema de contratação recusa uma pessoa, por exemplo, o dano pode resultar do dado de entrada, do modelo, de regra adicional programada pela instituição, de uma decisão humana posterior ou da combinação desses fatores. Em aplicações generativas, o conteúdo ofensivo pode depender do modelo de base, de uma adaptação feita por terceiro, do comando do usuário e dos filtros instalados na interface. Falar em "causa algorítmica" sem separar essas etapas apenas desloca o problema.

A opacidade pode agravar essa dificuldade porque evidências técnicas relevantes ficam sob domínio do fornecedor: logs, versões, parâmetros, documentação de testes e registros de atualização. Não se deve, porém, transformar opacidade em presunção automática de defeito. O caminho mais consistente é aproximar dever probatório da disponibilidade da informação. A experiência europeia recente é ilustrativa: a Diretiva (UE) 2024/2853 criou mecanismos de divulgação de evidências e presunções refutáveis em hipóteses nas quais a complexidade técnica ou científica torna excessivamente difícil provar defeito ou causalidade. A opção procura reduzir assimetria informacional sem dispensar a vítima de apresentar base plausível para sua pretensão (UNIÃO EUROPEIA, 2024b).

O direito brasileiro já conhece instrumentos semelhantes de redistribuição do ônus probatório, especialmente no sistema consumerista e no Código de Processo Civil, ainda que a aplicação dependa do caso concreto. Em IA, isso reforça a importância de preservação de registros. Se uma empresa não sabe qual versão do modelo estava ativa, quais atualizações haviam sido aplicadas ou quais variáveis alimentaram determinado resultado, a própria ausência de governança pode tornar a reconstrução do dano mais difícil. Documentar não é apenas boa prática técnica; pode ser condição para demonstrar que o dever de cuidado foi cumprido.

A jurisprudência sobre serviços digitais ajuda a compreender esse limite sem que seja preciso fingir que já existem dezenas de precedentes específicos sobre IA. No REsp 2.222.059/SP, julgado em 2025, o STJ reconheceu defeito na prestação de serviço de instituição de pagamento diante de fraude por falsa central de atendimento, destacando deveres de segurança e prevenção de operações suspeitas. A decisão não envolvia inteligência artificial, mas mostra uma lógica aplicável a sistemas automatizados: quem oferece infraestrutura digital não responde porque toda fraude é inevitavelmente sua, mas pode responder quando o risco se materializa justamente em ponto que deveria integrar sua arquitetura de segurança (STJ, 2025).

A conclusão inversa também é possível. No REsp 2.164.130/SP, julgado em fevereiro de 2026, a Terceira Turma afastou a responsabilidade da plataforma de pagamento pela não entrega de produto comprado em site de terceiro porque o serviço de pagamento havia sido processado sem falha e o evento lesivo ocorreu fora daquela cadeia de fornecimento específica (STJ, 2026a). O precedente, igualmente alheio à IA, é útil porque rejeita responsabilidade por mera proximidade tecnológica. Participar de alguma etapa da transação não torna um agente garantidor universal de tudo o que acontece depois.

A Súmula 479 do STJ fornece o outro lado da equação ao estabelecer responsabilidade objetiva das instituições financeiras por fortuito interno relacionado a fraudes e delitos praticados por terceiros no âmbito das operações bancárias (STJ, 2012). A expressão "fortuito interno" é relevante para IA porque impede que atos de terceiros sejam usados automaticamente como causa de exclusão. Se manipulações, ataques, usos abusivos ou erros humanos fazem parte do risco previsível da atividade e deveriam ser mitigados, a intervenção de terceiro pode não romper o nexo. Se o evento é externo à esfera de controle e não há falha do serviço, a conclusão pode ser diferente.

Essa análise deve alcançar o uso humano do sistema. Uma instituição pode alegar que a decisão final foi de um profissional; o profissional pode alegar que confiou na recomendação tecnológica; o fornecedor pode afirmar que sua ferramenta era apenas "auxiliar". Nenhuma dessas fórmulas resolve o caso sem investigar a prática real. Se o desenho da interface induzia aceitação automática, a supervisão humana pode ser apenas nominal. Se o profissional tinha informação clara sobre o erro e decidiu ignorá-la, sua conduta pode ganhar peso causal. Se a ferramenta foi utilizada fora dos limites explicitamente informados e contra instruções razoáveis, o usuário pode assumir parcela relevante do risco.

Por isso, causalidade e imputação devem ser analisadas conjuntamente, mas não confundidas. Saber que determinado componente contribuiu para o resultado não responde, sozinho, quem deve reparar. É preciso verificar deveres jurídicos, posição na cadeia, capacidade de controle, previsibilidade do risco e regime aplicável. A complexidade do sistema não elimina esses elementos; apenas exige investigação mais disciplinada.

Responsabilidade distribuída não pode significar responsabilidade diluída

Sistemas de inteligência artificial raramente chegam ao usuário pelas mãos de um único agente. Um desenvolvedor cria o modelo de base; outro ator fornece dados; uma empresa adapta o sistema; um integrador conecta-o à infraestrutura local; a organização usuária define procedimentos; profissionais tomam decisões; plataformas de nuvem mantêm parte da operação. Essa fragmentação cria risco jurídico evidente: cada participante pode apontar para o anterior ou para o seguinte até que a vítima fique sem interlocutor responsável.

A solução não está em declarar todos solidariamente responsáveis em qualquer hipótese. Solidariedade depende de fundamento jurídico. No CDC, regras próprias disciplinam integrantes da cadeia de fornecimento e hipóteses de responsabilidade; em outras relações, será necessário demonstrar participação juridicamente relevante no evento danoso. Uma empresa que simplesmente fornece infraestrutura genérica não ocupa a mesma posição de quem escolhe o modelo, define sua finalidade e o oferece como parte essencial do serviço. Responsabilidade distribuída deve acompanhar distribuição efetiva de controle e deveres.

O desenvolvedor pode responder por falhas relacionadas ao projeto, treinamento, documentação ou segurança quando elas se conectem ao dano e ao regime aplicável. O integrador pode criar defeito ao combinar corretamente dois componentes isolados de maneira insegura. A instituição usuária pode falhar ao empregar ferramenta inadequada para sua população, não treinar profissionais ou ignorar alertas de deterioração. O operador pode utilizar deliberadamente o sistema fora da finalidade indicada. Em muitos casos, não haverá uma causa única, mas contribuições diferentes. A tarefa do Direito é tornar essas contribuições juridicamente visíveis.

Isso exige contratos de tecnologia melhores, mas contratos não podem funcionar como mecanismo para transferir à vítima todos os riscos. Cláusulas entre fornecedor e instituição podem distribuir internamente obrigações de atualização, auditoria,

documentação, confidencialidade e indenização regressiva. Elas ajudam a definir quem deveria fazer o quê, mas não necessariamente alteram direitos de terceiro protegido por normas cogentes. A cadeia pode organizar regressos entre seus participantes depois da reparação sem transformar a vítima em responsável por decifrar acordos empresariais aos quais nunca teve acesso.

Melo (2024) sustenta que as cláusulas gerais do Código Civil e do CDC já oferecem base importante para tratar responsabilidade por IA, argumento que merece atenção justamente porque evita uma substituição precipitada do sistema existente. Isso não significa que legislação específica seja inútil. Regras novas podem melhorar documentação, gestão de riscos, acesso à prova e deveres de governança. Mas a inovação regulatória é mais consistente quando complementa critérios de imputação, e não quando cria uma responsabilidade completamente desconectada das categorias civis já consolidadas.

O debate legislativo brasileiro mostra essa tensão. O PL nº 2.338/2023 permanece em análise na Câmara dos Deputados em agosto de 2026 e, portanto, não pode ser usado como se fosse lei vigente. A tramitação reúne propostas distintas sobre responsabilidade, risco, fornecedores e operadores, enquanto o texto aprovado pelo Senado parte de uma estrutura regulatória mais ampla que convive com Código Civil, CDC e legislação especial (CÂMARA DOS DEPUTADOS, 2026a; CÂMARA DOS DEPUTADOS, 2026b). Para uma obra comprometida com precisão, o ponto decisivo é simples: o sistema brasileiro atual continua sendo construído principalmente a partir das regras existentes, complementadas por normas setoriais e pela interpretação jurisprudencial.

A comparação europeia também revela uma evolução interessante. A União Europeia aprovou em 2024 uma nova Diretiva de Responsabilidade por Produtos Defeituosos que inclui expressamente a realidade de software, atualizações e aprendizado de máquina e procura enfrentar assimetrias de prova em sistemas tecnicamente complexos (UNIÃO EUROPEIA, 2024b). O movimento é menos uma tentativa de responsabilizar "a IA" e mais uma atualização de quem integra a cadeia econômica e de quais informações devem estar disponíveis quando um produto digital causa dano. Essa é uma lição útil para qualquer regime: quanto maior a complexidade, maior a importância de rastreabilidade e definição prévia de papéis.

Mecanismos de seguro também podem ganhar espaço, sobretudo em aplicações de alto impacto. Salles e Costa (2023) discutem a securitização como forma de

diluição social de riscos difíceis de individualizar. O seguro, contudo, não substitui responsabilidade nem governança. Se mal desenhado, pode apenas precificar a falha e enfraquecer incentivos de prevenção. Sua utilidade está em complementar um sistema que continue exigindo padrões de segurança, documentação, atualização e possibilidade de regresso contra quem efetivamente contribuiu para o dano.

Há uma consequência institucional importante: responsabilidade civil deve influenciar o desenho do sistema antes do acidente. Saber que será necessário explicar por que determinada versão foi usada incentiva registro de versões; saber que falhas previsíveis de segurança podem ser imputadas incentiva testes; saber que instruções vagas não afastam defeito incentiva documentação clara; saber que supervisão meramente formal pode não romper o nexo incentiva desenho de interfaces que realmente permitam discordância humana. A responsabilidade cumpre, assim, função reparatória, mas também preventiva.

Essa função preventiva não pode ser confundida com promessa de risco zero. Nenhuma tecnologia complexa elimina integralmente falhas, e o Direito não deve converter toda inovação malsucedida em ilícito. Um resultado adverso pode ocorrer mesmo quando produto, serviço e usuário atuaram dentro de padrões razoáveis. O desafio é diferenciar dano inevitável de dano produzido por risco mal gerenciado. Essa distinção exige prova, contexto e prudência, não presunções morais sobre tecnologia.

No fim, a pergunta relevante não é "quem paga quando a IA erra?", mas quais decisões humanas e institucionais fizeram com que aquele erro alcançasse a vítima. Às vezes a resposta estará no fornecedor; em outras, no operador, na instituição, em vários agentes ou em nenhum deles, se faltar defeito, ilicitude ou nexo juridicamente relevante. A responsabilidade civil não deve escolher um culpado simbólico. Deve reconstruir a arquitetura real do dano.

A partir daí, porém, muda o critério jurídico de imputação. A responsabilidade civil pode distribuir riscos e reparar prejuízos sem exigir, em todos os casos, culpa moral individual. O Direito Penal opera sob exigências muito mais estritas. Quando sistemas de IA são utilizados em fraudes, invasões, perseguições, manipulações ou outros ilícitos, não basta dizer que determinada tecnologia participou do evento. Será necessário identificar conduta humana, tipicidade, dolo ou culpa quando admitida e critérios pessoais de imputação. O foco se desloca, portanto, da reparação do dano para a atribuição penal de responsabilidade.

CAPÍTULO 5

INTELIGÊNCIA ARTIFICIAL, CRIMES DIGITAIS E RESPONSABILIDADE PENAL

A tecnologia pode ampliar o crime, mas não substitui a conduta

A passagem da responsabilidade civil para a responsabilidade penal exige uma mudança de linguagem. No capítulo anterior, o problema era determinar em que condições um dano poderia ser imputado a fornecedores, instituições ou usuários. No Direito Penal, a pergunta é mais estreita e mais exigente. Não basta identificar prejuízo, risco ou falha de segurança. É necessário verificar se houve uma conduta humana típica, antijurídica e culpável, respeitando legalidade, pessoalidade da pena e proibição de responsabilidade objetiva. A inteligência artificial pode ampliar a velocidade, a escala ou a sofisticação de uma prática criminosa, mas não dispensa esses filtros (BRASIL, 1940; FORNASIER, 2022; ARAÚJO; SANTOS, 2026).

Essa premissa evita uma expressão sedutora e imprecisa: "crime cometido pela IA". Um modelo pode gerar mensagem fraudulenta, encontrar vulnerabilidade, produzir voz clonada ou automatizar interações com centenas de vítimas. Ainda assim, o sistema não se converte, por esse fato, em sujeito penal dotado de consciência da ilicitude e culpabilidade. O Direito brasileiro admite responsabilidade penal de pessoas jurídicas em hipóteses constitucional e legalmente delimitadas, como ocorre no campo ambiental, mas essa construção não autoriza transferir por analogia o estatuto do ente coletivo para algoritmos ou modelos de linguagem (STF, 2013; FORNASIER, 2022).

O princípio da legalidade é decisivo justamente quando a tecnologia muda mais rápido que a legislação. Não há crime sem lei anterior que o defina, e uma conduta socialmente reprovável não pode ser transformada em infração penal por semelhança com outra já tipificada. Isso não significa que cada ferramenta de IA exija um novo tipo penal. Na maioria das situações, a tecnologia funciona como meio de execução de crimes existentes. Uma voz sintetizada para induzir alguém a transferir dinheiro, por exemplo, pode integrar uma fraude eletrônica se estiverem presentes os elementos do estelionato e a utilização de contato telefônico ou outro meio fraudulento análogo previstos no art. 171, § 2º-A, do Código Penal. A Lei nº 15.397/2026 atualizou esse dispositivo sem criar um "crime de clonagem de voz": o núcleo continua sendo obter vantagem ilícita mediante indução ou manutenção da vítima em erro (BRASIL, 1940; BRASIL, 2026a).

A diferença é importante porque o mesmo recurso técnico pode participar de condutas juridicamente opostas. Um sintetizador de voz utilizado para acessibilidade, dublagem autorizada ou produção artística é lícito; empregado para se passar por familiar sequestrado e exigir dinheiro, pode funcionar como instrumento de extorsão ou fraude, conforme a dinâmica concreta. O Direito Penal não pune a capacidade tecnológica. Pune a conduta descrita em lei, praticada com o elemento subjetivo exigido pelo tipo.

O mesmo raciocínio vale para invasões de sistemas. Ferramentas de IA podem ajudar a identificar falhas, produzir códigos ou adaptar ataques, mas o art. 154-A do Código Penal exige invasão de dispositivo informático de uso alheio com finalidade específica de obter, adulterar ou destruir dados ou informações sem autorização, ou de instalar vulnerabilidades para obtenção de vantagem ilícita. Um sistema de segurança que encontra vulnerabilidades em ambiente autorizado e um agente que usa automação para explorar clandestinamente a mesma falha podem executar operações tecnicamente semelhantes, mas ocupam posições penais distintas porque autorização, finalidade e contexto modificam o fato juridicamente relevante (BRASIL, 1940).

Essa capacidade de aumentar escala é uma das razões pelas quais a criminalidade digital vem incorporando IA generativa a esquemas já conhecidos. A Organização Internacional de Polícia Criminal (INTERPOL) registrou em 2026 o uso crescente de deepfakes, identidades sintéticas e ferramentas automatizadas em fraudes transnacionais, enquanto o Escritório das Nações Unidas sobre Drogas e Crime (UNODC) descreveu a incorporação de IA a golpes de engenharia social, falsificação de identidade e outras práticas cibernéticas. Esses documentos não demonstram que a IA criou uma espécie penal nova; mostram que reduziu barreiras operacionais, tornou a personalização mais barata e permitiu que grupos criminosos reproduzissem abordagens em escala maior (INTERPOL, 2026; UNODC, 2025).

O legislador brasileiro começou, ao mesmo tempo, a reconhecer situações em que o uso de IA merece relevância penal expressa. A Lei nº 15.123/2025 acrescentou causa de aumento de metade ao crime de violência psicológica contra a mulher quando praticado mediante inteligência artificial ou outro recurso tecnológico que altere imagem ou som da vítima. Aqui, a tecnologia não cria o crime-base; agrava uma forma de execução que pode intensificar humilhação, manipulação, isolamento ou dano emocional (BRASIL, 2025a).

A atualização mais significativa, até o fechamento deste capítulo, ocorreu com a Lei nº 15.487, de 6 de agosto de 2026. A norma reformulou dispositivos do Estatuto da Criança e do Adolescente (ECA) para enfrentar violência sexual também em ambiente digital e com uso de inteligência artificial. O art. 241-C do ECA passou a abranger expressamente a simulação da participação de criança ou adolescente em conteúdo de violência sexual mediante adulteração, montagem ou modificação de representação visual com recurso tecnológico que altere imagem ou voz, inclusive IA. O art. 241-E do ECA também passou a considerar, para os crimes do ECA, representações que envolvam criança ou adolescente real ou fictício, ainda que produzidas, manipuladas ou geradas por tecnologias digitais, quando presentes os elementos sexuais definidos pela lei (BRASIL, 2026b).

A mesma lei aumentou a pena de determinadas condutas de aliciamento quando o agente utiliza IA, deepfake, filtros ou outro recurso para alterar a própria imagem e voz e fazer-se passar por criança ou adolescente com finalidade sexual. O ponto jurídico é relevante porque a norma não pune "usar IA com menores" de forma genérica. Ela descreve finalidades e condutas específicas. A precisão legislativa é o que permite incorporar novas técnicas sem abandonar a estrutura garantista do Direito Penal (BRASIL, 2026b).

Esses exemplos revelam duas estratégias legislativas distintas. Em alguns crimes, a tecnologia cabe como meio dentro de tipos já suficientemente abertos, como fraude por contato telefônico ou invasão de dispositivo. Em outros, o legislador identifica um risco novo ou uma insuficiência interpretativa e menciona expressamente IA, deepfake ou conteúdo sintético. A escolha entre uma estratégia e outra não pode ser feita apenas pela novidade da ferramenta. Deve depender da existência real de lacuna de proteção e dos limites impostos pela legalidade.

Entre autoria, participação e omissão

A dificuldade aumenta quando várias pessoas participam do ciclo tecnológico. Um usuário pode solicitar ao modelo uma mensagem enganosa; um desenvolvedor pode ter criado a ferramenta para finalidade legítima; uma empresa pode oferecer a infraestrutura; outro agente pode adaptar o sistema para remover salvaguardas; terceiros podem fornecer dados ou contas utilizadas no golpe. A proximidade técnica com o instrumento não basta para transformar todos em autores ou partícipes. O art. 29 do Código Penal exige concurso para o crime e vincula a pena à culpabilidade de cada participante (BRASIL, 1940).

Isso significa que o desenvolvedor de um modelo de propósito geral não se torna partícipe de todo crime praticado por um usuário apenas porque sua tecnologia foi utilizada. Ferramentas de linguagem, edição de imagem ou programação têm incontáveis usos lícitos. Para imputação penal de participação, é preciso demonstrar contribuição relevante para o fato e o elemento subjetivo exigido. Saber abstratamente que um produto pode ser abusado é diferente de aderir a um plano criminoso, adaptar deliberadamente a ferramenta para viabilizá-lo ou prestar auxílio consciente a uma prática concreta (ARAÚJO; SANTOS, 2026; FORNASIER, 2022).

No sentido oposto, também seria inadequado supor que a distância física ou a automação excluem participação. Quem configura um sistema para disparar milhares de mensagens fraudulentas, fornece listas de vítimas, instala mecanismo de clonagem de voz e acompanha os resultados pode concorrer para o crime mesmo sem conversar diretamente com nenhuma vítima. A IA muda a forma material da contribuição, não a necessidade de demonstrar que o agente concorreu conscientemente para o fato.

As hipóteses omissivas são ainda mais delicadas. O art. 13, § 2º, do Código Penal atribui relevância penal à omissão quando o omitente devia e podia agir para evitar o resultado, dentro das situações legais de posição de garantidor. Isso impede uma conclusão frequente e perigosa: a de que todo diretor, programador ou responsável técnico responde penalmente por qualquer resultado ilícito gerado por um sistema sob sua esfera organizacional. A posição hierárquica, sozinha, não cria automaticamente dever penal de impedir todos os danos possíveis.

Colombo (2025), ao examinar criminal compliance apoiado por inteligência artificial, mostra justamente que falhas ou decisões inesperadas de sistemas internos não autorizam atribuição automática de responsabilidade penal aos garantes da organização. É necessário reconstruir deveres concretos de vigilância, capacidade de intervenção, conhecimento do risco e relação entre a omissão e o resultado. A automação pode alterar a forma de exercer o dever de controle, mas não elimina as exigências dogmáticas da omissão imprópria.

A mesma prudência vale para a culpa. O parágrafo único do art. 18 do Código Penal estabelece que, salvo previsão expressa, ninguém pode ser punido por fato previsto como crime senão quando o pratica dolosamente. Portanto, a afirmação de que um programador "foi negligente" não basta, por si, para criar responsabilidade penal. É necessário que exista modalidade culposa prevista para o delito aplicável, além de violação do dever objetivo de cuidado, previsibilidade e nexo causal. Erro de

engenharia, falha de teste ou documentação inadequada podem gerar consequências civis, administrativas ou contratuais sem necessariamente configurar crime (BRASIL, 1940).

Esse ponto diferencia de modo nítido responsabilidade penal e governança tecnológica. Uma organização pode ter deveres amplos de segurança, monitoramento e conformidade cuja violação justifique sanções administrativas ou indenização. No campo penal, porém, a passagem da falha organizacional para a pena individual exige fundamento adicional. O sistema não pode usar o vocabulário da "sociedade de risco" para presumir culpabilidade. A pena continua vinculada a fatos atribuíveis a pessoas segundo critérios normativos definidos (COLOMBO, 2025).

Há situações em que a arquitetura da ferramenta se aproxima mais diretamente do plano criminoso. Um serviço criado especificamente para gerar documentos falsos, burlar verificações biométricas, produzir material sexual ilícito ou automatizar ataques pode apresentar indícios muito diferentes de uma plataforma genérica posteriormente desviada por usuários. Mesmo assim, a análise deve recair sobre condutas demonstráveis: finalidade do serviço, conhecimento dos administradores, mecanismos oferecidos, mensagens de incentivo, adaptação a pedidos ilícitos e eventual participação econômica no esquema. Não se pune uma "má tecnologia"; pune-se a contribuição humana juridicamente comprovada.

O problema da previsibilidade também precisa ser tratado com cautela. Sistemas de aprendizado podem produzir saídas inesperadas, mas imprevisibilidade técnica não é sinônimo de ausência de autoria humana nem prova de dolo. Se um usuário solicita reiteradamente conteúdo criminoso, contorna barreiras e escolhe utilizar a saída, sua ação posterior pode ser clara mesmo que o modelo tenha gerado detalhes que ele não antecipava. Para o desenvolvedor, em contraste, a existência de comportamentos emergentes pode exigir deveres de segurança, mas não autoriza presumir adesão subjetiva a cada conteúdo ilícito que apareça.

Essa distinção protege o Direito Penal contra dois excessos simétricos. O primeiro seria permitir que a expressão "foi a IA" se tornasse uma forma de ocultar decisões humanas deliberadas. O segundo seria atribuir a pessoas tecnicamente próximas do sistema uma responsabilidade penal objetiva por resultados que não quiseram, não previram e para os quais não possuíam dever penal específico de prevenção. Nenhuma das soluções é compatível com um sistema de culpabilidade.

Legalidade antes da ansiedade punitiva

A velocidade das mudanças tecnológicas produz pressão por novas criminalizações. Deepfakes, clones de voz, agentes autônomos e ferramentas de geração de código fazem surgir rapidamente condutas que parecem não caber com conforto em categorias tradicionais. A resposta penal, entretanto, não deve partir da sensação de novidade. O primeiro passo é verificar se o bem jurídico já está protegido por tipo existente e se o novo recurso altera apenas o meio de execução. Criar tipos penais sobrepostos pode aumentar incerteza sem melhorar proteção.

A experiência brasileira recente mostra essa busca por precisão. Em vez de criar um delito geral de "uso criminoso de inteligência artificial", o legislador escolheu intervenções pontuais: agravou violência psicológica contra a mulher quando a IA altera imagem ou som; atualizou fraude eletrônica e delitos patrimoniais; e reformulou o ECA para alcançar de maneira expressa determinadas formas de violência sexual sintética e aliciamento digital (BRASIL, 2025a; BRASIL, 2026a; BRASIL, 2026b). Essa técnica não resolve todos os conflitos, mas preserva uma vantagem: mantém a resposta vinculada a bens jurídicos e condutas definidos.

Também existem áreas nas quais a tentação de analogia deve ser contida. Conteúdos sexuais sintéticos envolvendo adultos, por exemplo, podem tocar honra, violência psicológica, perseguição, ameaça, constrangimento ou outros crimes dependendo do caso, mas não se deve assumir que qualquer dispositivo criado para fotografias ou gravações reais abrange automaticamente uma imagem inteiramente gerada. Quando faltar correspondência típica, a lacuna não pode ser preenchida contra o acusado por interpretação extensiva ilimitada. O desconforto social pode justificar debate legislativo; não substitui lei penal anterior.

Essa disciplina é especialmente importante diante de modelos generativos porque eles tornam barata a produção de variações. Um mesmo agente pode gerar milhares de mensagens, rostos ou documentos em poucos minutos. A escala pode agravar dano, demonstrar organização, influenciar dosimetria ou integrar circunstâncias específicas, mas quantidade tecnológica não muda por si só o núcleo do crime. É preciso analisar cada estrutura normativa: número de vítimas, continuidade delitiva, concurso de crimes, organização criminosa ou outras consequências dependem de requisitos próprios, e não da simples presença de automação.

O caráter transnacional acrescenta outra dificuldade. Ferramenta, servidor, agente e vítima podem estar em países diferentes. A cooperação policial internacional

vem tratando fraudes com IA como fenômeno cada vez mais conectado a redes de engenharia social, lavagem de dinheiro e identidades sintéticas (INTERPOL, 2026). Isso aumenta a importância de preservação de registros, identificação de infraestrutura e cooperação entre jurisdições, mas não altera o princípio básico de que competência e aplicação da lei penal dependem das regras de territorialidade e extraterritorialidade previstas no ordenamento.

A discussão sobre autonomia algorítmica também precisa ser recolocada em seu tamanho real. Parte da literatura pergunta se sistemas suficientemente autônomos deveriam receber alguma forma de personalidade ou sanção própria. No Direito Penal brasileiro atual, essa hipótese permanece especulativa. A pena pressupõe censura jurídica dirigida a sujeito capaz de ocupar posição normativa no ordenamento. Punir um modelo com prisão é obviamente impossível; impor-lhe multa significa, na prática, atingir patrimônio de alguém que o possui ou opera. Criar ficção de culpabilidade algorítmica pode mais esconder do que esclarecer quem detinha controle e benefício (FORNASIER, 2022; CHAVES JUNIOR; BERZAGUI, 2021).

A comparação com pessoas jurídicas não resolve o problema. O Supremo Tribunal Federal (STF) reconheceu, no Recurso Extraordinário (RE) 548.181, a possibilidade de persecução penal de pessoa jurídica por crime ambiental sem exigir imputação simultânea a pessoa física, com fundamento constitucional e legal específicos. Essa construção resulta de uma opção normativa expressa e de desenvolvimento jurisprudencial próprio; não oferece autorização para criar, por analogia, responsabilidade penal de "pessoas algorítmicas" (STF, 2013).

O caminho mais consistente, no presente, é manter a inteligência artificial dentro da análise do fato humano. A tecnologia pode ser instrumento, objeto material, mecanismo de ocultação, meio de execução, fonte de risco ou componente de uma organização criminosa. Pode também tornar mais difícil provar quem digitou um comando, quem controlou uma conta ou quem produziu determinado arquivo. Mas essas dificuldades são, antes de tudo, problemas de autoria, prova e investigação, não razões para abandonar tipicidade e culpabilidade.

Isso não significa passividade legislativa. Novos tipos ou causas de aumento podem ser necessários quando técnicas sintéticas criam formas de lesão que escapam aos tipos existentes ou quando a definição anterior produz incerteza incompatível com legalidade. A Lei nº 15.487/2026 é exemplo de intervenção dirigida: definiu expressamente como certas representações digitais, inclusive geradas por IA, se

inserem na tutela penal de crianças e adolescentes. A qualidade da criminalização depende justamente dessa precisão, e não de expressões amplas como "uso indevido de IA" sem definição suficiente.

O Direito Penal deve, portanto, resistir a duas narrativas fáceis. A primeira afirma que as leis atuais nada podem fazer porque a tecnologia é nova. A segunda sustenta que qualquer consequência danosa pode ser penalmente imputada ao usuário, desenvolvedor ou empresa porque alguém "deveria ter previsto" o comportamento da máquina. Entre essas posições existe um trabalho dogmático mais difícil: definir o fato, localizar a contribuição humana, identificar o tipo penal, provar o elemento subjetivo e respeitar os limites da imputação.

Essa exigência desloca a análise para a prova. Quando a IA é usada para fraudar, ameaçar, aliciar ou invadir, o processo penal precisa reconstruir autoria e materialidade em um ambiente no qual áudios podem ser clonados, imagens podem ser sintetizadas, conversas podem ser alteradas e arquivos genuínos podem ser denunciados como falsos. O mesmo conjunto de tecnologias que amplia certas práticas criminosas também modifica o valor e a vulnerabilidade da evidência digital. Depois de perguntar quem pode responder penalmente, torna-se indispensável demonstrar como sabemos que o fato aconteceu e por que o arquivo apresentado ao processo merece confiança.

CAPÍTULO 6

PROVA DIGITAL E CONTEÚDOS PRODUZIDOS OU MANIPULADOS POR INTELIGÊNCIA ARTIFICIAL

Parecer verdadeiro não é o mesmo que ser autêntico

A prova digital sempre exigiu cautela, mas a inteligência artificial generativa alterou o ponto de partida dessa cautela. Durante anos, a principal preocupação consistia em saber se um arquivo havia sido editado depois de criado, se uma conversa estava incompleta ou se uma fotografia fora retirada de contexto. Hoje é possível produzir, desde a origem, áudio, vídeo, imagem e texto que jamais corresponderam ao acontecimento que aparentam registrar. A questão probatória deixa, portanto, de ser apenas "este arquivo foi alterado?" e passa a incluir outra pergunta: "o evento representado existiu como acontecimento real?".

Essa mudança exige separar conceitos que frequentemente aparecem misturados. Autenticidade diz respeito à correspondência entre o objeto apresentado e aquilo que ele afirma ser. Integridade refere-se à preservação do conteúdo, sem modificação indevida, ao longo de seu percurso. Proveniência procura reconstruir origem e histórico do arquivo: onde surgiu, por quais sistemas passou, que transformações sofreu e quem teve controle sobre ele. Um vídeo pode permanecer integralmente inalterado desde o momento em que foi recebido e, ainda assim, ser um deepfake criado antes da coleta. Nesse caso, a cadeia de custódia posterior é perfeita, mas a representação continua falsa.

O processo brasileiro já oferece ferramentas para tratar documentos eletrônicos sem presumir que o suporte digital seja inferior ou superior ao papel. No processo civil, o Código de Processo Civil (CPC) admite meios legais e moralmente legítimos de prova, disciplina a ata notarial e dedica regras próprias aos documentos eletrônicos, cujo valor deve ser apreciado pelo juiz em conjunto com os demais elementos e com possibilidade de impugnação pelas partes (BRASIL, 2015). A tecnologia não cria uma categoria probatória imune ao contraditório. Ao contrário: quanto mais fácil se torna produzir um documento convincente, mais relevante se torna demonstrar de onde ele veio e em que contexto deve ser interpretado.

A ata notarial é exemplo importante porque costuma ser apresentada como solução universal para prints de telas, páginas e conversas. Ela pode ser extremamente útil para documentar que determinado conteúdo estava visível ao tabelião em certo momento. Isso não significa, porém, que transforme automaticamente em verdadeiro

tudo o que aparece na tela. Se um vídeo sintético já estava publicado quando foi constatado, a ata pode comprovar sua existência e aparência naquela data; não comprova, por si só, que a cena filmada ocorreu. O instrumento fortalece a documentação do fato observado, mas não elimina a necessidade de distinguir existência do arquivo, autoria, integridade e veracidade do evento representado.

No processo penal, essa preocupação ganha intensidade porque a prova pode sustentar restrições severas de liberdade. A Lei nº 13.964/2019 incorporou ao Código de Processo Penal disciplina expressa da cadeia de custódia, definida como o conjunto de procedimentos empregados para manter e documentar a história cronológica do vestígio desde o reconhecimento até o descarte. Os arts. 158-A a 158-F do Código de Processo Penal estruturam etapas de rastreamento e preservação que procuram assegurar que o material examinado em juízo corresponda ao vestígio efetivamente coletado (BRASIL, 1941). Para dados digitais, essa "mesmidade" assume importância particular porque alterações podem ocorrer sem sinais perceptíveis a olho nu.

A inteligência artificial não torna essas regras obsoletas; torna mais clara a diferença entre integridade e verdade. Um código hash pode demonstrar que determinada cópia permanece matematicamente idêntica ao arquivo de referência. Ele não informa se o rosto mostrado no vídeo pertence de fato à pessoa, se a voz foi sintetizada antes da apreensão ou se a conversa foi fabricada no dispositivo de origem. A função do hash é decisiva, mas delimitada: proteger a comparação entre versões de um mesmo vestígio, e não certificar a realidade histórica representada por seu conteúdo.

Essa distinção impede duas respostas igualmente perigosas. A primeira seria continuar tratando imagem e áudio como provas naturalmente persuasivas, exigindo da parte contrária demonstração quase impossível de falsificação. A segunda seria considerar toda mídia digital suspeita apenas porque deepfakes existem. O processo não pode substituir ingenuidade por ceticismo absoluto. Se toda gravação puder ser neutralizada pela simples alegação de que "pode ter sido criada por IA", a tecnologia produzirá o mesmo efeito discutido no capítulo anterior sobre o chamado dividendo do mentiroso: evidências verdadeiras perderão força não porque foram refutadas, mas porque a possibilidade abstrata de fabricação passou a existir (CHESNEY; CITRON, 2019).

A resposta mais consistente é proporcional ao grau de controvérsia. Conteúdo cuja origem é reconhecida pelas partes, confirmado por testemunhos ou corroborado

por outros elementos pode dispensar investigação técnica extensa. Já uma gravação decisiva cuja autoria seja seriamente impugnada, com inconsistências de metadados ou ausência de fonte original, pode exigir perícia, comparação com arquivos de referência e exame da forma de obtenção. A presença de IA não elimina a valoração contextual da prova; apenas torna inadequado confiar exclusivamente na impressão sensorial de que algo "parece real".

Cadeia de custódia é a história da prova

A jurisprudência recente do Superior Tribunal de Justiça mostra que essa exigência de contexto já está sendo construída de maneira concreta. No Agravo Regimental no Habeas Corpus (AgRg no HC) nº 828.054/RN, julgado em 2024, a Quinta Turma considerou inadmissíveis dados de celular quando a extração se resumiu a consulta direta e capturas de tela, sem metodologia capaz de assegurar idoneidade e integridade. O problema não era a existência de mensagens de WhatsApp em si, mas a impossibilidade de demonstrar que aquilo que chegou ao processo correspondia, de maneira auditável, ao conteúdo originalmente coletado (STJ, 2024).

Esse entendimento não significa que todo vestígio digital sem formalidade perfeita seja automaticamente nulo. Em fevereiro de 2026, no AgRg no HC 1.014.212/ES, a Sexta Turma enfrentou situação em que havia dúvida razoável sobre a integridade de prints de conversas e imagens de videomonitoramento. A resposta foi exigir perícia complementar para permitir avaliação técnica e contraditório efetivo, e não simplesmente afirmar, de modo abstrato, que todo material digital era imprestável (STJ, 2026b). O critério é relevante: quando a confiabilidade não pode ser esclarecida pela documentação disponível, a solução deve buscar produzir informação verificável.

Poucos meses depois, a Corte Especial, no Inquérito 1.674/DF, reconheceu que o espelhamento de dados associado à função hash constitui mecanismo hábil para preservar integridade e auditabilidade da evidência imaterial. O julgamento destacou que a cadeia de custódia serve para documentar a regularidade do percurso probatório e que eventuais desconformidades devem ser examinadas concretamente, em cotejo com risco de adulteração e prejuízo, e não tratadas como nulidade automática em qualquer hipótese (STJ, 2026c). Essa posição evita formalismo vazio sem abandonar a exigência de rastreabilidade.

O ponto de equilíbrio aparece ainda com mais nitidez no Agravo em Recurso Especial (AREsp) nº 3.132.918/RS, julgado em junho de 2026. A Quinta Turma considerou inadmissíveis prints de mensagens desacompanhados do arquivo digital de

origem, porque essa ausência impedia verificar cadeia de custódia e mesmidade. No mesmo processo, contudo, os vídeos foram tratados de forma diferente: estavam disponíveis integralmente às partes, não havia demonstração concreta de manipulação e o Tribunal não identificou risco equivalente à credibilidade do material. A decisão mostra que "prova digital" não constitui bloco homogêneo; cada vestígio precisa ser analisado de acordo com as condições reais de preservação e impugnação (STJ, 2026d).

Esses precedentes fornecem uma orientação particularmente útil para a era dos conteúdos sintéticos. A prova confiável não depende de uma tecnologia milagrosa, mas de uma sequência documentada de atos. Sempre que possível, deve-se preservar o arquivo original ou a melhor fonte disponível, registrar o modo de aquisição, evitar intervenções sobre a mídia de origem, criar cópias de trabalho verificáveis e manter informação suficiente para que outro especialista possa repetir ou contestar a análise. Quanto mais decisivo for o vestígio e mais séria a controvérsia sobre sua autenticidade, maior deve ser a exigência de auditabilidade.

Também é necessário distinguir a evidência original de produtos derivados. Transcrever um áudio, melhorar contraste de uma imagem, reduzir ruído, ampliar resolução, resumir milhares de mensagens ou utilizar IA para identificar padrões pode ajudar a investigação e a compreensão do processo. Mas o resultado transformado não deve apagar a fonte. Uma transcrição é representação linguística de um áudio; um quadro "aprimorado" é resultado de processamento; um resumo produzido por modelo generativo é interpretação automatizada de um conjunto de documentos. Se o artefato derivado passa a ser tratado como se fosse o vestígio original, cria-se uma camada de inferência cuja origem pode desaparecer do debate.

Esse risco é maior quando o próprio método de aprimoramento é generativo. Ferramentas de super-resolução, reconstrução de imagem ou restauração de voz podem acrescentar padrões estatisticamente plausíveis que não estavam explicitamente presentes no registro original. A utilidade visual pode ser real, mas o produto não deve ser confundido com uma revelação neutra do conteúdo. O processo precisa saber o que foi coletado e o que foi criado posteriormente para facilitar visualização. A fronteira entre evidência e ilustração deve permanecer identificável.

Contraditório técnico também exige acesso suficiente ao material. Não basta dizer que uma perícia foi feita se a parte não consegue compreender minimamente quais arquivos foram examinados, qual metodologia foi aplicada e que limitações

existem. Isso não implica abertura irrestrita de ferramentas proprietárias ou exposição de dados protegidos; significa que a prova não pode ganhar autoridade apenas pela opacidade técnica de quem a produziu. Em questões decisivas, deve ser possível testar os pressupostos relevantes, oferecer assistente técnico quando cabível e confrontar conclusões periciais.

Por isso, a cadeia de custódia não deve ser compreendida como formulário burocrático. Ela cumpre função epistêmica e processual: reduz a distância entre o vestígio inicialmente encontrado e o objeto posteriormente valorado pelo juiz. Em ambiente digital, esse percurso pode envolver aparelho, nuvem, backup, exportação, ferramenta forense, banco de dados e múltiplas cópias. Cada etapa adiciona risco de perda de contexto. Documentá-las não torna a prova infalível; torna sua confiabilidade discutível de maneira racional.

Quando o falso pode nascer pronto

Deepfakes acrescentam um problema que a cadeia de custódia tradicional não resolve sozinha: a falsidade pode preceder o início da própria cadeia. Se uma gravação sintética é enviada a uma testemunha e o telefone é apreendido minutos depois, é perfeitamente possível preservar, com rigor absoluto, um arquivo que nunca registrou o evento que parece retratar. O processo deve então investigar não apenas o percurso do vestígio após a coleta, mas sua proveniência anterior.

A literatura jurídica internacional vem chamando atenção para esse deslocamento. Delfino (2023) sustenta que deepfakes pressionam os mecanismos tradicionais de autenticação porque a semelhança visual deixou de funcionar como indicativo suficiente de origem. A questão não é importar propostas específicas das regras probatórias norte-americanas para o Brasil, mas reconhecer o problema subjacente: quanto mais convincente a fabricação, maior a necessidade de fundamentos externos à própria aparência do conteúdo.

Esses fundamentos podem incluir arquivo original, metadados coerentes, histórico de dispositivo, registros de plataforma, assinaturas digitais, testemunhos sobre a captação, correspondência com outros eventos e mecanismos técnicos de proveniência. Nenhum elemento isolado deve ser tratado como certificado universal de verdade. Metadados podem ser apagados ou alterados; testemunhas podem se equivocar; assinaturas confirmam determinadas relações técnicas, não necessariamente todo o significado do conteúdo. A força probatória emerge da convergência entre sinais independentes.

Iniciativas de proveniência digital procuram fortalecer esse ecossistema. A Coalition for Content Provenance and Authenticity (C2PA) desenvolveu o padrão de mesmo nome, no qual credenciais criptograficamente vinculadas podem registrar informações sobre a origem e o histórico de transformações de imagens, vídeos, áudios e documentos. O objetivo é permitir que aplicações preservem e verifiquem uma trilha de proveniência desde a criação ou edição do conteúdo (C2PA, 2025). Esses mecanismos podem se tornar valiosos para o processo, principalmente quando incorporados já no dispositivo de captação.

Mas as credenciais também precisam ser interpretadas com cuidado. A presença de uma cadeia de proveniência válida fortalece certas afirmações sobre histórico do arquivo; não demonstra automaticamente que a cena representada é verdadeira em todos os seus aspectos. Da mesma forma, a ausência de credencial não prova falsidade. Bilhões de arquivos legítimos são produzidos sem sistemas de proveniência. Transformar um padrão técnico voluntário em presunção absoluta criaria nova forma de exclusão probatória e privilegiaria dispositivos e plataformas capazes de implementar a tecnologia.

A mesma cautela se aplica a detectores de deepfake. Eles podem auxiliar a perícia e indicar sinais incompatíveis com determinado processo de geração, mas não devem funcionar como oráculos. Modelos de detecção são treinados sobre conjuntos de exemplos e podem perder eficácia diante de técnicas, compressões e transformações diferentes daquelas observadas no desenvolvimento. Em processo judicial, o resultado "92% provável de ser sintético", por exemplo, não fala por si: será necessário compreender o que esse número significa, para qual população foi validado, quais erros são esperados e se o arquivo analisado estava dentro das condições de uso.

Essa exigência vale também quando uma parte simplesmente alega que a prova é deepfake. O surgimento da IA generativa não pode criar uma presunção de falsidade audiovisual. Uma impugnação séria deve apontar elementos capazes de gerar dúvida concreta ou demonstrar por que a origem não está suficientemente estabelecida. Caso contrário, qualquer pessoa poderia desqualificar gravação autêntica mediante uma palavra tecnológica. O contraditório precisa permitir contestar a prova, mas também precisa impedir que a possibilidade abstrata de falsificação substitua argumentos e evidências.

No processo civil, essa ponderação combina-se com a atipicidade dos meios de prova e a valoração motivada. Um print pode ter pouco valor isoladamente e ganhar

força quando acompanhado de dados de contexto, confirmação da contraparte, registros independentes ou outros documentos. Um arquivo tecnicamente sofisticado pode, ao contrário, perder valor se sua origem permanecer obscura. O CPC não exige que todo documento eletrônico seja submetido previamente a perícia; exige que autenticidade, acesso ao conteúdo e controvérsia sejam tratados de forma compatível com o caso concreto (BRASIL, 2015).

No processo penal, as exigências de confiabilidade e defesa assumem peso ainda maior. O fato de uma prova digital parecer espontânea ou ter sido recebida de terceiro não elimina a necessidade de avaliar licitude da obtenção, preservação e integridade. Ao mesmo tempo, a jurisprudência do STJ mostra que o exame não pode ser reduzido a lista mecânica de formalidades. A pergunta decisiva permanece: o material conserva condições suficientes para que sua identidade, integridade e significado sejam auditados, e a defesa teve oportunidade efetiva de impugná-lo? (STJ, 2026b; STJ, 2026c; STJ, 2026d).

A consequência mais profunda da IA generativa, portanto, não é o desaparecimento da prova digital. É o fim da confiança baseada apenas na aparência. O processo precisa deslocar parte de sua confiança da superfície do arquivo para a história que pode ser demonstrada sobre ele. Quem produziu? Em qual dispositivo? Quando? Como foi obtido? O que foi alterado? Existe original ou fonte comparável? Há elementos independentes que confirmam seu conteúdo? Essas perguntas sempre foram relevantes; agora deixaram de poder ser negligenciadas.

Ao mesmo tempo, não se deve imaginar que a verdade judicial se transformará em competição entre laboratórios forenses. Prova continua sendo atividade jurídica orientada por contraditório, distribuição de ônus, standards de decisão e motivação. A perícia informa; não decide. Um detector auxilia; não sentencia. Um hash demonstra identidade entre arquivos; não comprova autoria do fato. Uma credencial de proveniência fortalece rastreabilidade; não elimina interpretação. A qualidade do processo está justamente em combinar conhecimento técnico com critérios jurídicos de admissibilidade e valoração.

Quando a tecnologia deixa de apenas produzir, alterar ou auxiliar a análise de evidências e passa a classificar pessoas, pontuar riscos, recomendar resultados ou participar diretamente de decisões que afetam direitos, o problema jurídico muda novamente. Já não basta perguntar se a prova é autêntica. Será preciso perguntar como a decisão automatizada pode ser compreendida, contestada e atribuída a

alguém. A confiabilidade da evidência cede lugar, então, à responsabilidade pelo próprio processo decisório.

CAPÍTULO 7

DECISÕES AUTOMATIZADAS, TRANSPARÊNCIA E RESPONSABILIDADE ALGORÍTMICA

Uma assinatura humana não torna a decisão humana

A decisão automatizada é frequentemente imaginada como uma situação extrema: um sistema recebe dados, calcula um resultado e executa sozinho uma consequência sobre a pessoa. Esse cenário existe, mas representa apenas uma parte do problema. Em muitas relações contemporâneas, o algoritmo não "decide" formalmente. Ele classifica, pontua, ordena, recomenda ou sinaliza um risco, e uma pessoa aparece ao final para confirmar o resultado. A presença dessa etapa humana pode alterar o enquadramento jurídico, mas não deveria encerrar a análise. O que importa é compreender quanto poder decisório o sistema exerceu de fato.

A diferença é central para a Lei Geral de Proteção de Dados Pessoais. O art. 20 da LGPD assegura ao titular o direito de solicitar revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluindo decisões destinadas a definir perfis pessoais, profissionais, de consumo e de crédito ou aspectos de sua personalidade. O dispositivo também prevê o fornecimento de informações claras e adequadas sobre os critérios e procedimentos utilizados, resguardados os segredos comercial e industrial (BRASIL, 2018). A redação cria proteção concreta, mas também uma fronteira interpretativa difícil: quando uma decisão deixa de ser "unicamente" automatizada porque alguém participou do processo?

Uma resposta puramente formal seria insuficiente. Se o funcionário recebe apenas um score, não tem acesso aos fatores relevantes, trabalha sob metas que tornam inviável revisar cada caso e normalmente confirma a recomendação, sua presença pode ter pouca capacidade real de modificar o resultado. De modo inverso, um sistema pode oferecer uma recomendação forte e ainda assim existir intervenção humana substantiva quando o responsável possui competência, tempo, autoridade, dados independentes e liberdade prática para discordar. A distinção não está no número de cliques humanos, mas na qualidade da intervenção.

Essa questão aparece de forma particularmente clara no caso SCHUFA, julgado pelo Tribunal de Justiça da União Europeia (TJUE) em 2023. A empresa alemã produzia automaticamente uma pontuação de crédito que era transmitida a terceiros. O Tribunal entendeu, à luz do art. 22 do Regulamento Geral sobre a Proteção de

Dados da União Europeia (RGPD), que a produção desse score podia constituir decisão individual automatizada quando o terceiro se apoiasse fortemente nele para estabelecer, executar ou encerrar uma relação contratual (TJUE, 2023). O acórdão não vincula o ordenamento brasileiro e o texto europeu possui estrutura diferente da LGPD. Seu valor comparativo está em outro ponto: uma etapa intermediária pode exercer efeito decisório material mesmo que outra organização assine formalmente a decisão final.

Esse enfoque funcional impede uma estratégia de evasão simples. Bastaria inserir um empregado entre o algoritmo e o resultado para afastar qualquer proteção destinada às decisões automatizadas, ainda que esse empregado não tivesse poder real de revisão. O Direito precisa observar a arquitetura decisória como um todo: qual informação chega à pessoa responsável, com que peso, quais alternativas estão disponíveis, com que frequência há divergência e se a organização considera legítimo contrariar a saída do sistema. Uma supervisão que não pode produzir discordância é supervisão apenas nominal.

A então Autoridade Nacional de Proteção de Dados (ANPD) reconheceu que essa fronteira ainda demanda desenvolvimento interpretativo. A Tomada de Subsídios sobre inteligência artificial e revisão de decisões automatizadas reuniu posições divergentes sobre a extensão do art. 20 da LGPD, inclusive quanto a decisões intermediárias, necessidade de revisão humana, alcance da afetação de interesses e equilíbrio entre transparência e segredos empresariais (ANPD, 2025a). Em 2026, o tema permanece em andamento na agenda regulatória e voltou a ser debatido publicamente pela Agência, o que recomenda prudência diante de afirmações categóricas sobre uma matéria ainda em construção (ANPD, 2026a).

Também é importante separar decisão automatizada de simples automação operacional. Um sistema que distribui senhas por ordem de chegada, organiza arquivos ou converte formatos não ocupa necessariamente a mesma posição de um modelo que define prioridade de atendimento, risco de fraude, perfil de crédito ou adequação de candidato. A intensidade jurídica depende da capacidade de produzir efeitos sobre interesses, direitos e oportunidades. Quanto mais a saída tecnológica modifica a posição da pessoa, maior a necessidade de justificar por que aquele processo é aceitável.

Por isso, a discussão não deve se concentrar apenas em saber se a máquina "substituiu" alguém. Sistemas de apoio podem influenciar decisões humanas de

maneira profunda. Um alerta de risco apresentado em vermelho, uma recomendação exibida como primeira opção ou um score sem intervalo de incerteza pode conduzir comportamento mesmo quando o operador conserva formalmente liberdade de escolha. A interface, a ordem das informações e a confiança institucional depositada no sistema passam a fazer parte da própria arquitetura da decisão.

Responsabilidade algorítmica começa, então, antes da responsabilidade civil discutida no capítulo anterior. Não se trata ainda de perguntar quem indeniza depois do dano. Trata-se de exigir que organizações saibam explicar como distribuíram poder entre pessoas e sistemas: quem pode alterar parâmetros, quem monitora resultados, quem recebe reclamações, quem possui autoridade para suspender o uso e quais registros permitem reconstruir uma decisão individual. Sem essas respostas, a expressão "decisão assistida por IA" pode esconder mais do que esclarecer.

Explicar não é entregar o código-fonte

Transparência costuma aparecer como resposta imediata à opacidade algorítmica, mas o conceito pode significar coisas muito diferentes. Uma organização pode informar publicamente que utiliza inteligência artificial; pode divulgar quais dados e finalidades orientam o sistema; pode explicar a uma pessoa por que recebeu determinado resultado; pode permitir auditoria técnica; ou pode abrir integralmente o código-fonte. Todas essas medidas aumentam alguma forma de visibilidade, mas não são equivalentes. A pergunta jurídica é qual transparência é necessária para qual finalidade.

Para a pessoa afetada, a explicação precisa tornar possível compreender e contestar. Essa finalidade é diferente da documentação dirigida a desenvolvedores ou auditores. Informar que "o modelo utilizou 143 variáveis e uma arquitetura de gradient boosting" pode ser tecnicamente correto e juridicamente pouco útil. Da mesma forma, uma explicação excessivamente simplificada, como "seu perfil apresentou risco elevado", não permite verificar se o dado estava errado, se um critério foi mal aplicado ou se a decisão incorporou fator juridicamente inadequado. Explicar é fornecer informação proporcional ao tipo de decisão e ao que o destinatário precisa fazer com ela.

O art. 20, § 1º, da LGPD aponta justamente para informações claras e adequadas sobre critérios e procedimentos. A lei não estabelece um direito geral de acesso ao código-fonte, e a proteção de segredos comercial e industrial continua relevante. Isso não significa, contudo, que segredo empresarial possa funcionar como resposta

automática a qualquer pedido. O núcleo da transparência deve sobreviver: a pessoa precisa receber informação suficiente para compreender a lógica decisória em nível útil e exercer os direitos que o ordenamento lhe assegura (BRASIL, 2018; LIMA; SÁ, 2020).

Um desenvolvimento comparado recente ajuda a tornar essa ideia concreta. Em fevereiro de 2025, no caso Dun & Bradstreet Austria, o Tribunal de Justiça da União Europeia entendeu que informações significativas sobre a lógica de uma decisão automatizada devem permitir que a pessoa compreenda quais dados pessoais foram utilizados e de que maneira contribuíram para o resultado. O Tribunal observou que poderia ser apropriado mostrar em que medida uma variação dos dados levaria a resultado diferente e afirmou que a mera comunicação do algoritmo não constitui, por si, explicação concisa e inteligível. Também rejeitou a ideia de que segredo empresarial elimine, como regra, o direito de acesso; em caso de conflito, a ponderação cabe à autoridade ou ao tribunal competente (TJUE, 2025).

Novamente, não se trata de transplantar automaticamente o RGPD para o Brasil. O valor do precedente está na função atribuída à explicação: ela não existe para satisfazer curiosidade técnica, mas para permitir compreensão e impugnação. Essa perspectiva aproxima transparência de contestabilidade. Uma explicação que não ajuda a pessoa a verificar dados, identificar fatores relevantes ou questionar o resultado pode ser formalmente extensa e ainda assim falhar em sua finalidade jurídica.

A literatura brasileira tem avançado na mesma direção. Pádua e Lorenzetto (2024) relacionam explicabilidade à justificação das decisões estatais e à accountability, distinguindo a compreensão do funcionamento técnico de uma explicação juridicamente voltada aos fundamentos de determinada decisão. Godoy et al. (2026) também situam o direito à explicação como mecanismo de transparência, responsabilização e proteção de direitos fundamentais diante de decisões baseadas em machine learning. A convergência é importante porque impede reduzir o problema a uma disputa entre sistemas "explicáveis" e "caixas-pretas". Mesmo um modelo tecnicamente interpretável pode ser utilizado em processo juridicamente injustificável.

A explicação técnica também possui limites epistemológicos. Ferramentas de interpretabilidade podem indicar variáveis associadas à saída, aproximar importância relativa de atributos ou construir exemplos contrafactuais. Elas não necessariamente reproduzem um raciocínio causal da máquina nem transformam a saída em fundamento jurídico suficiente. Se um sistema de seleção associa baixa pontuação a

determinada trajetória profissional, saber que essa variável pesou no resultado não responde se o critério era lícito, proporcional, necessário ou discriminatório. O Direito continua precisando avaliar a regra social incorporada ao modelo.

É por isso que transparência deve existir em camadas. Antes da decisão, a organização pode informar que utiliza sistema automatizado, sua finalidade, categorias de dados e consequências esperadas. Depois de uma decisão individual relevante, deve existir informação capaz de permitir contestação. Em nível institucional, auditorias, relatórios, logs, documentação de versões e avaliações de impacto precisam sustentar controle por autoridades e atores especializados. Nenhuma dessas camadas substitui integralmente as demais.

O Poder Judiciário brasileiro oferece um exemplo normativo particularmente desenvolvido dessa arquitetura. A Resolução CNJ nº 615/2025 estabelece transparência, explicabilidade, contestabilidade, auditabilidade e supervisão humana efetiva como princípios do uso de IA. A norma admite que auditoria e monitoramento sejam realizados sem acesso irrestrito ao código-fonte, desde que existam mecanismos de transparência e controle suficientes; para soluções de alto risco, exige documentação, registros de operação, medidas de explicabilidade, avaliação de impacto e monitoramento contínuo (CNJ, 2025; CNJ, 2026). A opção é significativa: transparência não é sinônimo de abertura absoluta, mas também não pode ser substituída por confiança cega no fornecedor.

Esse modelo é especialmente relevante para decisões estatais porque eficiência não substitui motivação. Se um sistema auxilia o julgador a localizar precedentes, classificar demandas ou sugerir texto, sua contribuição deve permanecer subordinada à autoridade responsável e às garantias do processo. A Resolução 615 determina supervisão humana, informa usuários externos sobre o caráter consultivo das soluções e preserva a possibilidade de modificação do produto gerado pela IA (CNJ, 2025). A tecnologia pode apoiar a decisão; não pode servir como lugar oculto onde a justificativa desaparece.

Accountability é uma arquitetura, não uma declaração

Responsabilidade algorítmica não se constrói acrescentando ao contrato uma frase de que "o usuário é responsável pela decisão final". Também não nasce de um selo genérico de "IA responsável". Accountability exige que a organização consiga demonstrar, de maneira verificável, quem decidiu o quê durante o ciclo de vida do

sistema, quais riscos foram identificados, que controles foram implementados e como uma pessoa afetada pode provocar revisão quando algo dá errado.

A documentação é parte dessa arquitetura. Registrar versão do modelo, parâmetros relevantes, data da decisão, fontes de dados aplicáveis, nível de intervenção humana e alterações posteriores permite reconstruir o processo sem depender da memória dos envolvidos. Logs não são, por si, transparência para o cidadão, mas tornam auditoria e responsabilização possíveis. Sem registro, uma instituição pode não conseguir responder se determinado resultado foi produzido antes ou depois de uma atualização, se havia alerta conhecido de desempenho ou se a saída foi modificada por um operador.

A supervisão humana precisa ser projetada com a mesma seriedade. Não basta designar uma pessoa responsável; ela precisa possuir competência para interpretar o resultado, informação sobre limitações, tempo para analisar casos relevantes e autoridade para ignorar, reverter ou interromper o sistema. A própria Resolução CNJ nº 615/2025 exige supervisão humana efetiva e proporcional ao risco e prevê que sistemas judiciais permitam modificação do produto gerado pela IA. A norma também reconhece riscos de automação e prevê capacitação contínua dos usuários (CNJ, 2025).

Essa preocupação aparece igualmente no Regulamento Europeu de Inteligência Artificial. O texto consolidado exige, para sistemas de alto risco, transparência suficiente para que o responsável pela implantação possa interpretar a saída e utilizá-la adequadamente, além de medidas de supervisão humana voltadas a reconhecer limitações, evitar confiança automática e permitir desconsiderar, reverter ou interromper o resultado do sistema. O art. 86 do AI Act prevê ainda, em determinadas decisões baseadas na saída de sistemas de alto risco do Anexo III, direito a explicações claras e significativas sobre o papel da IA e os principais elementos da decisão (UNIÃO EUROPEIA, 2024a).

A cronologia europeia precisa, contudo, ser apresentada corretamente. O Regulamento tornou-se aplicável de forma geral em 2 de agosto de 2026, mas o Regulamento (UE) 2026/1744 adiou a aplicação das seções centrais do regime de alto risco do Capítulo III: para sistemas enquadrados pelo Anexo III, a data passou a ser 2 de dezembro de 2027; para sistemas de alto risco incorporados a produtos do Anexo I, 2 de agosto de 2028 (UNIÃO EUROPEIA, 2026). Assim, os dispositivos europeus são úteis como arquitetura regulatória comparada, mas não devem ser apresentados

como se todo o regime de alto risco já estivesse plenamente operacional em agosto de 2026.

Uma boa governança também precisa prever o que acontece quando humano e sistema discordam. Se a organização mede desempenho apenas pela adesão do operador à recomendação, cria incentivo para automatizar a concordância. Se toda divergência precisa ser justificada, mas a concordância não, o desenho institucional transforma o algoritmo em padrão presumidamente correto. Em aplicações de impacto relevante, seria mais saudável registrar padrões de divergência e investigar se determinados grupos, contextos ou tipos de caso concentram erros. A discordância humana pode ser ruído; pode também ser sinal de falha sistemática.

Da mesma forma, a possibilidade de recurso precisa levar a algum lugar diferente. Uma pessoa que contesta uma decisão automatizada não deveria receber apenas a mesma saída executada novamente sobre os mesmos dados, sem oportunidade de corrigir informação, acrescentar contexto ou submeter o caso a avaliação capaz de reconsiderar critérios relevantes. Revisão implica abertura para mudança. Sem isso, o procedimento apenas reproduz a decisão sob outra etiqueta.

A Agência Nacional de Proteção de Dados vem tratando exatamente desses pontos ao desenvolver parâmetros interpretativos para o art. 20 da LGPD. A agenda regulatória vigente mantém inteligência artificial em andamento e inclui direitos dos titulares, princípios, hipóteses legais, boas práticas e governança entre os eixos do projeto. Em 2026, a Agência também iniciou a fase de testagem de seu sandbox regulatório com foco em transparência algorítmica, explicabilidade, riscos, incidentes e planos de descontinuidade (ANPD, 2026a; ANPD, 2026b). O cenário brasileiro, portanto, não é de vazio regulatório, mas de densificação progressiva de deveres já ancorados na proteção de dados e em normas setoriais.

Essa evolução sugere uma mudança importante na própria ideia de transparência. O objetivo não é tornar cada cidadão especialista em modelos estatísticos. É impedir que decisões relevantes se tornem incontestáveis porque os critérios foram distribuídos entre software, contratos, fornecedores e equipes técnicas. A pessoa deve conseguir saber que houve participação algorítmica quando isso for juridicamente relevante, compreender os principais fatores e procedimentos, corrigir dados incorretos, apresentar elementos que o sistema não considerou e alcançar um responsável capaz de modificar o resultado.

Responsabilidade algorítmica também possui dimensão coletiva. Um indivíduo pode obter explicação satisfatória sobre seu caso e ainda existir um sistema que produz diferenças sistemáticas entre grupos. Por isso, transparência individual precisa conviver com monitoramento agregado, auditorias, avaliações de impacto e controle regulatório. O caso concreto mostra como uma pessoa foi afetada; a governança revela se o mesmo problema está sendo repetido em escala.

No fundo, decisões automatizadas colocam o Direito diante de um problema conhecido sob uma nova arquitetura. Sempre que alguém exerce poder sobre interesses de outra pessoa, existem perguntas sobre razão, procedimento, prova, revisão e responsabilidade. A IA amplia a velocidade e a distância entre quem define os critérios e quem sofre os efeitos, mas não elimina essas exigências. Uma decisão legítima não precisa revelar cada operação matemática interna; precisa permanecer vinculada a pessoas e instituições capazes de justificar, contestar, corrigir e responder por seu uso.

As mesmas exigências de transparência e contestabilidade ganham outra densidade quando decisões algorítmicas entram na vida cotidiana por meio do consumo. Preços personalizados, ofertas, recomendações, bloqueios antifraude, concessão de crédito, atendimento automatizado e publicidade dirigida influenciam o que consumidores veem, pagam e conseguem contratar. Nesse ambiente, a automação passa a conviver com vulnerabilidade, dever de informação, práticas abusivas e responsabilidade do fornecedor.

CAPÍTULO 8

IA NAS RELAÇÕES DE CONSUMO E PROTEÇÃO DO CONSUMIDOR

Personalizar não significa poder tratar consumidores como mercados isolados

A inteligência artificial entrou nas relações de consumo de forma silenciosa. Antes de aparecer como chatbot ou assistente virtual, já estava presente em mecanismos de recomendação, detecção de fraude, classificação de risco, segmentação de publicidade, ordenação de ofertas e definição de quais consumidores seriam considerados mais ou menos interessantes para determinado produto. O resultado é uma mudança na própria experiência de mercado: duas pessoas podem acessar a mesma plataforma e encontrar produtos, condições, prioridades e mensagens diferentes sem perceber que o ambiente foi organizado a partir de informações inferidas sobre elas.

Essa personalização não é, por si, incompatível com o Direito do Consumidor. Recomendar um produto compatível com preferências anteriores pode reduzir custos de busca; sistemas antifraude podem proteger patrimônio; análise automatizada pode acelerar concessão responsável de crédito; atendimento digital pode ampliar disponibilidade. A questão jurídica começa quando a personalização altera informação, oportunidade ou condição contratual de modo que o consumidor não consegue compreender, comparar ou contestar. O CDC continua partindo de uma premissa que a sofisticação tecnológica não elimina: a vulnerabilidade do consumidor diante do fornecedor e o direito a informação adequada, liberdade de escolha e proteção contra métodos comerciais desleais (BRASIL, 1990).

O credit scoring oferece um exemplo brasileiro anterior à atual geração de IA, mas ainda extremamente útil. No REsp 1.419.697/RS, julgado sob o rito dos repetitivos, o Superior Tribunal de Justiça reconheceu a licitude da avaliação estatística de risco de crédito e afastou a necessidade de consentimento prévio para a simples utilização do método. Ao mesmo tempo, afirmou que o consumidor pode exigir esclarecimentos sobre as fontes dos dados e as informações pessoais valoradas, devendo ser preservadas privacidade e máxima transparência. O uso de informações excessivas, sensíveis, incorretas ou desatualizadas pode caracterizar abuso e gerar responsabilidade (STJ, 2014). A Súmula 550 consolidou essa orientação (STJ, 2015).

A importância desse precedente não está em presumir que modelos contemporâneos de aprendizado de máquina sejam idênticos aos sistemas estatísticos avaliados em 2014. Está em um princípio que permanece válido: segredo empresarial pode proteger a metodologia, mas não transformar a pessoa avaliada em objeto completamente opaco de classificação. A Lei nº 12.414/2011 reforça essa lógica ao disciplinar bancos de dados de crédito, proibir informações excessivas e sensíveis em determinados contextos e assegurar conhecimento dos principais elementos considerados na análise de risco, resguardado o segredo empresarial (BRASIL, 2011a).

A LGPD acrescenta outra camada quando a pontuação deriva de tratamento automatizado de dados pessoais. Nesse contexto, o art. 20 da LGPD permite solicitar revisão de decisões tomadas unicamente com base em tratamento automatizado que afetem interesses e requerer informações sobre critérios e procedimentos, observados segredos comercial e industrial (BRASIL, 2018). Em uma relação de consumo, esses direitos não substituem o CDC; convivem com ele. A própria Agência Nacional de Proteção de Dados informa que direitos relacionados a tratamento de dados podem ser exercidos também perante órgãos de defesa do consumidor quando o tratamento ocorre dentro de relação de consumo (ANPD, s.d.).

O risco aumenta quando pontuações deixam de utilizar apenas histórico financeiro e passam a incorporar sinais comportamentais pouco intuitivos. Horário de acesso, padrão de navegação, dispositivo utilizado, localização, redes de relacionamento e outros indicadores podem ser estatisticamente associados a probabilidade de compra, inadimplência ou fraude. Mesmo quando determinada variável não é juridicamente proibida, sua relevância precisa ser examinada à luz de finalidade, necessidade, transparência e possível efeito discriminatório. O consumidor não pode corrigir um dado que sequer sabe estar sendo usado, nem contestar adequadamente uma inferência cuja existência lhe é invisível.

A personalização de preços torna essa assimetria ainda mais sensível. Preços diferentes não são automaticamente ilícitos: mercados convivem com variação por horário, demanda, estoque, fidelidade, canal, custos e condições objetivas de contratação. O problema surge quando características pessoais ou comportamentais são utilizadas para extrair maior disposição individual a pagar, restringir artificialmente ofertas ou discriminar consumidores sem justificativa legítima. Em 2018, o Departamento de Proteção e Defesa do Consumidor multou administrativamente a Decolar.com por práticas de geo pricing e geo blocking, após identificar diferenciação de preços e disponibilidade de acomodações segundo

localização geográfica do consumidor (MJSP, 2018). O caso não transforma toda precificação dinâmica em ilícito, mas mostra que personalização comercial encontra limites jurídicos.

A distinção entre preço dinâmico e preço personalizado é importante. No primeiro, a condição pode variar para todos em função de fatores do mercado, como demanda e disponibilidade. No segundo, a oferta pode ser adaptada a características ou inferências sobre uma pessoa ou grupo. Na prática, os dois mecanismos podem se combinar. O Direito do Consumidor não precisa proibir algoritmos de precificação para enfrentar abusos; precisa conseguir identificar quando a diferença decorre de critério legítimo e quando explora assimetrias, oculta informação relevante ou restringe oportunidades de maneira incompatível com igualdade nas contratações e boa-fé.

Também não se deve presumir que a recomendação mais visível seja a "melhor" para o consumidor. Sistemas de ranking podem ordenar produtos segundo relevância, margem comercial, pagamento por destaque ou probabilidade de conversão. Se esses critérios são indistinguíveis para quem compra, a arquitetura da plataforma passa a influenciar a escolha sem que a pessoa saiba se está vendo a opção mais adequada ou a mais rentável para o intermediário. A OCDE observa que ranking personalizado pode beneficiar consumidores ao aproximar ofertas de seus interesses, mas também pode explorar vieses de posição e ser otimizado para extrair maior disposição a pagar (OCDE, 2023).

A resposta jurídica, portanto, não está em exigir que todos vejam a mesma internet. Personalização pode ser legítima e útil. O limite aparece quando a diferença de tratamento é construída de forma incompatível com informação adequada, boa-fé, proteção de dados, não discriminação ou possibilidade de escolha. Em mercados algorítmicos, igualdade não significa necessariamente identidade absoluta de ofertas; significa que diferenças relevantes precisam ter fundamento defensável e não podem se esconder atrás da complexidade do sistema.

Quando a interface aprende a pressionar

A inteligência artificial pode influenciar o consumidor antes mesmo de qualquer decisão automatizada sobre crédito ou preço. Interfaces digitais registram cliques, abandonos, tempo de permanência, rolagem, respostas a descontos e milhares de pequenas interações que permitem testar quais mensagens aumentam a probabilidade de compra. Isso transforma a arquitetura de escolha em objeto central da proteção

consumerista. A pergunta deixa de ser apenas se uma afirmação publicitária é verdadeira e passa a incluir como opções, botões, avisos e obstáculos foram organizados para conduzir comportamento.

A literatura internacional utiliza a expressão dark commercial patterns para práticas de design digital que direcionam, enganam, coagem ou manipulam consumidores em escolhas que frequentemente não correspondem a seus melhores interesses. Entre os exemplos estão informação escondida, dificuldade artificial de cancelamento, falsa urgência, opções pré-selecionadas, obstáculos sucessivos para recusar coleta de dados e assimetria visual entre aceitar e rejeitar. A OCDE documenta evidências de prevalência e dano dessas práticas e destaca que personalização pode tornar a manipulação ainda mais difícil de reconhecer (OCDE, 2022; OCDE, 2023).

O CDC não utiliza essa terminologia, mas possui instrumentos que alcançam grande parte do fenômeno. São direitos básicos a informação adequada e clara, liberdade de escolha e proteção contra publicidade enganosa ou abusiva e métodos comerciais coercitivos ou desleais. O art. 37 do CDC proíbe publicidade enganosa e abusiva; o art. 39 do mesmo Código, entre outras práticas, veda prevalecer-se da fraqueza ou ignorância do consumidor em razão de idade, saúde, conhecimento ou condição social para impingir produtos ou serviços (BRASIL, 1990). O design digital não cria uma zona livre desses deveres.

A dificuldade é que uma interface manipulativa pode não conter uma frase objetivamente falsa. Um botão de adesão enorme e colorido ao lado de uma recusa quase invisível; um cancelamento escondido em várias telas; um contador de urgência que reinicia; ou uma recomendação apresentada como neutra quando corresponde a publicidade paga podem alterar escolhas sem recorrer à mentira tradicional. Nesses casos, a análise precisa olhar para o conjunto da prática comercial, e não apenas para a literalidade das palavras exibidas.

A personalização aumenta a preocupação porque a plataforma pode aprender quais pressões funcionam melhor para pessoas diferentes. Um consumidor pode responder a escassez; outro, a medo de perder benefício; outro, a parcelamento aparentemente pequeno. A OCDE observa que o ambiente digital pode transformar vulnerabilidade em condição situacional: cansaço, urgência, baixa familiaridade tecnológica, dificuldade financeira ou forte carga emocional podem tornar qualquer consumidor mais suscetível em determinados momentos (OCDE, 2023). Isso amplia

a ideia tradicional de vulnerabilidade sem apagar grupos que já merecem proteção reforçada.

O crédito ao consumidor mostra como o ordenamento brasileiro já reage a esse tipo de exploração. A Lei nº 14.181/2021, voltada à prevenção e ao tratamento do superendividamento, incorporou ao CDC deveres de informação e crédito responsável e proibiu, entre outras condutas, ocultar ou dificultar a compreensão sobre ônus e riscos e assediar ou pressionar consumidores para contratar crédito, especialmente idosos, analfabetos, pessoas doentes ou em situação de vulnerabilidade agravada (BRASIL, 2021b). Um sistema capaz de identificar precisamente quem está financeiramente pressionado não recebe autorização para explorar essa fragilidade; ao contrário, a capacidade de segmentação torna o dever de cuidado ainda mais relevante.

Publicidade dirigida exige a mesma atenção. O fato de uma mensagem ser exibida apenas para determinado perfil não reduz os deveres de identificação, veracidade e não abusividade. Pelo contrário, campanhas microsegmentadas podem dificultar fiscalização pública porque diferentes consumidores recebem anúncios distintos e o conteúdo pode desaparecer rapidamente. Em 2026, a Senacon firmou acordo com o Google Brasil para reforçar verificação de anunciantes de produtos e serviços financeiros, em resposta a anúncios de falsas corretoras, empréstimos fraudulentos e outras modalidades de golpe. A iniciativa não cria um regime geral de responsabilidade por IA, mas demonstra que publicidade digital passou a exigir mecanismos de rastreabilidade dos anunciantes e prevenção de fraude (MJSP, 2026).

A proteção de crianças e adolescentes tornou esse movimento ainda mais explícito. A Lei nº 15.211/2025, denominada Estatuto Digital da Criança e do Adolescente, proíbe técnicas de perfilamento para direcionamento de publicidade comercial a crianças e adolescentes e também veda, para essa finalidade, análise emocional e determinadas tecnologias imersivas. A lei impede ainda a criação de perfis comportamentais de crianças e adolescentes para direcionamento publicitário (BRASIL, 2025b). O Decreto nº 12.880/2026 regulamentou aspectos desse regime e vinculou expressamente a publicidade que explora deficiência de julgamento e experiência da criança ao conceito de publicidade abusiva do CDC (BRASIL, 2026c).

Esse exemplo é importante porque mostra que transparência nem sempre basta. Informar que uma criança está vendo publicidade personalizada não transforma automaticamente o perfilamento em prática aceitável quando a própria lei o proíbe. Da mesma forma, um aviso de “oferta personalizada” não legitima discriminação ou

pressão abusiva sobre adultos. A informação é uma ferramenta central do Direito do Consumidor, mas existem situações em que o problema está na própria prática, e não apenas em sua invisibilidade.

A IA generativa acrescenta outro risco ao ambiente publicitário: produção em escala de textos, imagens, depoimentos e personagens altamente persuasivos. O fato de uma peça ter sido produzida por IA não a torna enganosa, assim como uma publicidade escrita por pessoa não é verdadeira por esse motivo. O critério continua sendo o efeito da comunicação sobre o consumidor, a identificabilidade de sua natureza publicitária e a correção das afirmações sobre produto ou serviço. Quando avatares, vozes sintéticas ou depoimentos artificiais são apresentados de modo capaz de criar falsa impressão de experiência real, endosso ou autoridade, os princípios tradicionais de identificação e veracidade ganham nova aplicação.

A proteção consumerista não deveria, portanto, buscar uma interface "neutra" em sentido impossível. Todo design organiza escolhas. A questão é impedir que assimetria informacional e conhecimento comportamental sejam usados para remover autonomia de forma desleal. Se a empresa conhece muito mais sobre a vulnerabilidade momentânea do consumidor do que ele próprio percebe e utiliza esse conhecimento para pressioná-lo, o problema deixa de ser mera eficiência de marketing e passa a envolver a própria lealdade da prática comercial.

Automatizar atendimento não pode automatizar o desamparo

A relação de consumo não termina quando o contrato é celebrado. Sistemas de inteligência artificial também filtram reclamações, respondem dúvidas, identificam fraudes, priorizam chamados, autorizam estornos e bloqueiam transações. Esses usos podem tornar o atendimento mais rápido e disponível, mas criam um risco particular: a automação que facilita a contratação pode se transformar em barreira quando o consumidor deseja cancelar, contestar ou corrigir um erro.

Chatbots ilustram bem a ambivalência. Um assistente capaz de resolver imediatamente uma dúvida simples reduz espera e custo para todos. O problema aparece quando o sistema não compreende situações excepcionais, repete respostas padronizadas ou impede que a pessoa alcance alguém com autoridade para revisar a demanda. Em setores regulados pelo Poder Executivo federal, o Decreto nº 11.034/2022 disciplina o Serviço de Atendimento ao Consumidor (SAC) e exige, entre outros pontos, acesso gratuito, canais integrados e atendimento telefônico humano por período mínimo diário definido na norma (BRASIL, 2022). A

automação pode integrar o SAC; não apaga os deveres de efetividade e acesso estabelecidos para esses serviços.

A mesma lógica deve orientar bloqueios antifraude. Modelos capazes de identificar transações atípicas protegem consumidores e fornecedores, mas falsos positivos podem impedir pagamento, bloquear conta ou exigir procedimentos adicionais justamente em situações legítimas. A proteção contra fraude não autoriza um sistema de contestação incapaz de explicar minimamente o bloqueio ou de corrigir erro. Quando o modelo produz consequência relevante, precisa existir caminho real para revisão, especialmente porque o consumidor raramente possui acesso aos dados técnicos que geraram o alerta.

Fraudes praticadas por terceiros também mostram que IA não altera automaticamente a distribuição de responsabilidade. Como visto no Capítulo 4, o STJ distingue riscos que integram a esfera de segurança do serviço daqueles externos à cadeia de fornecimento. A existência de tecnologia sofisticada não transforma fornecedor em segurador universal, mas tampouco permite classificar como "culpa do golpista" riscos previsíveis que deveriam ter sido mitigados. Quanto mais uma empresa automatiza autenticação, prevenção e atendimento, maior a importância de demonstrar que esses mecanismos funcionavam de forma compatível com o risco da atividade.

As plataformas digitais ocupam posição particular quando o golpe chega ao consumidor por publicidade paga. O Decreto nº 12.975/2026 atualizou a regulamentação do Marco Civil da Internet e passou a impor deveres relacionados à prevenção e mitigação de riscos, fraudes digitais e anúncios ilícitos. Entre as novas regras, provedores que oferecem ferramentas pagas de anúncio ou impulsionamento devem conservar informações relativas aos anúncios e respectivos anunciantes, e há previsão de indisponibilização, após notificação, de publicidade enganosa, abusiva ou fraudulenta nos termos do CDC (BRASIL, 2026d). O regime não elimina a necessidade de analisar o caso concreto, mas aumenta a rastreabilidade de uma cadeia publicitária que antes podia ser muito difícil de reconstruir.

Essa atualização ajuda a compreender que proteção do consumidor em ambientes algorítmicos não depende apenas da relação bilateral entre comprador e vendedor. Marketplaces, redes sociais, motores de busca, sistemas de pagamento e intermediários de publicidade podem ocupar funções distintas. Nem todos respondem pelo mesmo fato, mas cada um pode possuir deveres próprios relacionados àquilo que controla.

Uma plataforma que apenas hospeda determinada informação não está na mesma posição daquela que vende segmentação, verifica anunciantes, recomenda conteúdo e recebe remuneração pelo impulsionamento.

O dever de informação também muda de forma. Não seria razoável exigir uma etiqueta "feito por IA" em qualquer interação automatizada irrelevante. Mas quando o consumidor acredita estar conversando com pessoa, recebendo aconselhamento individual ou obtendo avaliação humana e essa impressão influencia sua confiança, ocultar a automação pode tornar-se juridicamente relevante. A transparência deve acompanhar a expectativa criada e a consequência da interação. Um robô que informa horário de funcionamento exige menos explicação do que um sistema que recomenda produto financeiro ou decide se uma contestação será acolhida.

Da mesma forma, mecanismos de recurso precisam ser proporcionais ao impacto. Não faz sentido exigir revisão humana complexa para cada recomendação de filme. Mas crédito recusado, conta bloqueada, pagamento negado, contrato cancelado ou reclamação encerrada automaticamente afetam interesses concretos e podem justificar canais mais robustos de contestação. O desafio é evitar dois extremos: transformar qualquer automação em procedimento burocrático inviável ou permitir que decisões relevantes sejam protegidas por uma sequência infinita de respostas automáticas.

A vulnerabilidade do consumidor fornece o critério de leitura. Ela não significa incapacidade permanente nem autoriza paternalismo indiscriminado. Significa reconhecer que fornecedores normalmente controlam tecnologia, dados, desenho da interface e documentação, enquanto o consumidor enxerga apenas a superfície do serviço. A IA aumenta essa assimetria porque permite inferir preferências e testar comportamentos em escala. Consequentemente, boa-fé, informação, segurança e possibilidade de contestação tornam-se ainda mais importantes, e não menos.

A proteção eficiente também não pode depender exclusivamente da capacidade individual de perceber manipulação. Se um sistema é projetado para explorar vieses cognitivos, personalizar pressão ou dificultar cancelamento, não é razoável transferir todo o ônus para que cada consumidor "leia com atenção". O Direito do Consumidor nasceu precisamente porque liberdade contratual formal não elimina desigualdade material entre as partes. A inteligência artificial apenas torna essa premissa mais visível.

No fim, o critério não é ser a favor ou contra personalização. Sistemas inteligentes podem ampliar acesso, proteger contra fraude e tornar serviços melhores. O limite aparece quando a vantagem do fornecedor é construída pela opacidade, pela

exploração de vulnerabilidade ou pela impossibilidade prática de correção. Uma relação de consumo tecnologicamente avançada continua precisando ser uma relação juridicamente leal.

A mesma assimetria muda de natureza quando o sistema deixa de ser empregado por um fornecedor privado e passa às mãos do Estado. Em relações de consumo, o ordenamento reconhece a vulnerabilidade diante do mercado; quando o poder público automatiza triagens, fiscalização, distribuição de recursos ou decisões administrativas, entram em primeiro plano legalidade, motivação, impessoalidade, transparência, devido processo e os limites específicos do exercício automatizado da autoridade estatal.

CAPÍTULO 9

USO DE INTELIGÊNCIA ARTIFICIAL PELO PODER PÚBLICO E SEUS LIMITES

Eficiência não substitui competência nem motivação

O poder público tem razões legítimas para incorporar inteligência artificial. Órgãos administrativos trabalham com grandes volumes de processos, fiscalização, benefícios, arrecadação, compras, atendimento e políticas públicas que exigem análise contínua de informação. Sistemas de IA podem auxiliar triagens, localizar inconsistências, priorizar casos, resumir documentos, detectar padrões e apoiar decisões. O ganho de eficiência pode ser relevante, mas a natureza estatal da atividade altera o problema jurídico: quando a tecnologia participa do exercício de autoridade pública, o cidadão não está apenas diante de um serviço digital. Está diante de uma decisão ou prática vinculada à Constituição Federal e a deveres próprios da Administração.

O art. 37 da Constituição Federal submete a Administração Pública aos princípios da legalidade, impessoalidade, moralidade, publicidade e eficiência. A presença conjunta desses princípios é importante porque impede que eficiência seja convertida em justificativa autônoma para qualquer automação. Um sistema mais rápido não é juridicamente superior se opera fora da competência do órgão, aplica critérios incompatíveis com a lei, cria discriminações, oculta fundamentos ou inviabiliza controle. A tecnologia pode melhorar a execução da função administrativa; não redefine sozinha os limites da função pública (BRASIL, 1988).

Essa conclusão aparece de forma ainda mais concreta na Lei nº 9.784/1999. No âmbito federal, os atos administrativos que neguem, limitem ou afetem direitos ou interesses, imponham deveres, decidam concursos, recursos ou outras situações previstas no art. 50 da Lei nº 9.784/1999 devem ser motivados com indicação dos fatos e dos fundamentos jurídicos. A motivação deve ser explícita, clara e congruente (BRASIL, 1999). Se um sistema produz uma pontuação de risco ou uma recomendação que passa a orientar a decisão, a autoridade não pode substituir a motivação por expressões como "resultado do algoritmo" ou "classificação automática". A saída tecnológica pode integrar a instrução; ela não elimina o dever de explicar por que a consequência jurídica é adequada ao caso.

O mesmo raciocínio vale quando a IA é utilizada apenas como apoio. Uma minuta produzida por modelo generativo pode organizar informação e sugerir

linguagem, mas a autoridade continua responsável pelos fatos, pelo enquadramento jurídico e pela decisão assinada. Uma recomendação estatística pode ajudar a priorizar fiscalização, mas a Administração precisa saber quais limites e riscos acompanham o modelo. Ribeiro (2024), ao estudar inteligência artificial e decisão administrativa, destaca justamente que integração tecnológica e princípio da motivação precisam caminhar juntos: explicabilidade e transparência importam porque a decisão administrativa deve continuar sujeita a controle.

A automação também desloca escolhas para etapas anteriores ao ato final. Quando uma equipe define quais dados serão usados, qual variável receberá maior peso, qual limiar transformará um caso em "alto risco" e quais situações serão encaminhadas para investigação, ela está estruturando parte da política pública. A discricionariedade não desaparece; pode migrar do servidor que decide individualmente para o desenho do sistema. Carvalho (2026) chama atenção para desafios jurídicos que ultrapassam a decisão visível e alcançam gestão de recursos tecnológicos, compras públicas, capacidade institucional e tensões com autodeterminação e democracia. Essa mudança exige que escolhas técnicas relevantes sejam tratadas como escolhas administrativas sujeitas a justificativa e controle.

A Lei nº 14.129/2021 reforça essa leitura. A Lei do Governo Digital incentiva a transformação digital, a simplificação e o uso de tecnologia para otimizar processos, mas coloca ao lado desses objetivos a transparência na execução dos serviços públicos, a participação social, a prestação de contas, a linguagem clara e o controle do tratamento de dados pessoais. Também prevê que plataformas de governo digital ofereçam ferramentas de transparência e controle sobre dados, incluindo fonte, finalidade e compartilhamento em hipóteses abrangidas pela lei (BRASIL, 2021a). Digitalizar não significa reduzir garantias; significa prestar o serviço por outro meio preservando direitos.

Essa cautela é particularmente importante em decisões de massa. Benefícios sociais, tributos, licenças, sanções, seleção pública e fiscalização podem envolver milhares ou milhões de pessoas. A automação oferece consistência e escala, mas também pode reproduzir um mesmo erro em grande número de casos. Um equívoco humano isolado pode atingir uma pessoa; um parâmetro incorreto aplicado em lote pode atingir uma população inteira. Por isso, o ganho de escala precisa ser acompanhado por capacidade equivalente de detectar, corrigir e interromper erros.

Também não se deve confundir padronização com neutralidade. Regras administrativas podem legitimamente reduzir variações injustificadas entre casos semelhantes. Porém, quando o sistema incorpora dados históricos, proxies ou critérios não transparentes, a padronização pode apenas tornar uniforme um tratamento inadequado. A impessoalidade administrativa exige critérios juridicamente pertinentes, não ausência de pessoas na decisão. Um algoritmo pode ser tão impessoal quanto uma tabela e, ao mesmo tempo, produzir distinções incompatíveis com igualdade e finalidade pública.

O ponto central, portanto, é de atribuição. A Administração não pode transferir competência jurídica a um modelo como se ele se tornasse novo centro autônomo de autoridade. O sistema é instrumento inserido em uma estrutura institucional. Mesmo quando executa etapas de forma automática, a decisão precisa permanecer atribuível a órgão ou autoridade responsável, com base normativa identificável e possibilidade de revisão. A tecnologia pode mediar o exercício do poder; não pode tornar indeterminado quem o exerce.

O Estado que classifica também precisa ser auditável

A transparência assume significado diferente quando o algoritmo pertence ao Estado. Em relações privadas, segredo empresarial pode limitar certas formas de acesso. Na Administração, a publicidade é princípio constitucional e a Lei de Acesso à Informação (LAI) estabelece publicidade como regra e sigilo como exceção. Mais especificamente, o art. 7º, § 3º, da LAI assegura acesso aos documentos e informações utilizados como fundamento da tomada de decisão e do ato administrativo após a edição do ato (BRASIL, 2011b). Isso cria um ponto de partida importante para sistemas algorítmicos: se dados, relatórios, regras ou resultados técnicos influenciaram a decisão pública, não podem desaparecer da história administrativa simplesmente porque foram processados por software.

Isso não significa que todo código-fonte deva ser aberto em qualquer circunstância. A própria LAI preserva hipóteses legais de sigilo e segredos protegidos, e sistemas públicos podem conter informações cuja divulgação comprometa segurança, fiscalização ou direitos de terceiros. A questão é mais precisa: a Administração precisa preservar informação suficiente para permitir controle da legalidade da decisão. Um contrato com fornecedor não deveria tornar impossível conhecer critérios, dados relevantes, versão do sistema, registros de operação ou limites do modelo quando essas informações forem necessárias para explicar um ato que afetou alguém.

A proteção de dados adiciona outra camada de responsabilidade. O tratamento pelo poder público possui disciplina específica na LGPD e deve estar conectado à finalidade pública, à persecução do interesse público e às competências legais do órgão. O Guia Orientativo da ANPD sobre tratamento de dados pessoais pelo Poder Público ressalta a necessidade de identificar base jurídica adequada, finalidade, compartilhamento e observância dos princípios da lei (ANPD, 2023). Quando dados de diferentes bases são reunidos para criar perfis de risco, a pergunta não é apenas se a integração é tecnicamente possível, mas se ela é necessária, proporcional e juridicamente vinculada à finalidade pública.

O caso holandês SyRI tornou-se referência comparada justamente por expor esse risco. O sistema era utilizado para combater fraude em benefícios, tributos e outras áreas mediante combinação e análise de dados públicos. Em 2020, a Corte Distrital de Haia concluiu que o regime não alcançava o equilíbrio exigido pelo art. 8 da Convenção Europeia de Direitos Humanos e destacou insuficiência de transparência e controle sobre o modelo de risco. A decisão não proibiu o uso estatal de algoritmos para combater fraude; considerou inadequadas as salvaguardas do sistema concreto (RECHTBANK DEN HAAG, 2020). Essa distinção importa: finalidade legítima não torna automaticamente proporcional todo meio tecnológico escolhido.

Fraude e fiscalização são áreas especialmente sensíveis porque o sistema pode transformar correlações em suspeita institucional. Se determinado endereço, padrão de consumo ou conjunto de vínculos aumenta uma pontuação de risco, pessoas podem ser selecionadas para investigação sem compreender por quê. Um modelo pode ser útil para direcionar recursos escassos, mas precisa ser validado para o contexto, monitorado quanto a erros e examinado quanto a impactos desiguais. A classificação algorítmica não é prova de irregularidade; é, no máximo, um instrumento de priorização cuja qualidade e limites devem ser conhecidos por quem o utiliza.

O problema também aparece em políticas distributivas. Sistemas podem ordenar pedidos, estimar vulnerabilidade, identificar elegibilidade ou antecipar demanda por serviços. Nessas situações, erros não produzem apenas inconveniência administrativa; podem afetar renda, saúde, moradia, educação ou acesso a direitos. Quanto mais essencial o serviço, menor a tolerância a procedimentos que não ofereçam explicação e correção efetivas. Faleiros Júnior e Oliveira (2026) ressaltam, em perspectiva de Direito Administrativo brasileiro, que decisões estatais de alto impacto assistidas por IA exigem mecanismos reforçados de transparência, auditabilidade e controle jurisdicional para preservar devido processo.

A contestação precisa alcançar não apenas o dado bruto, mas também a inferência relevante. Se o cidadão comprova que sua renda estava registrada incorretamente, a correção é relativamente direta. A situação é diferente quando o sistema conclui que ele apresenta "alto risco" a partir de combinação de dezenas de sinais. Nesse caso, a Administração precisa possuir condições de explicar quais elementos foram relevantes em nível suficiente para que a pessoa possa apontar erro, contexto ausente ou critério inadequado. Uma revisão que apenas repete o cálculo com os mesmos dados não responde ao problema.

O contraditório administrativo, por isso, deve ser pensado também para infraestruturas automatizadas. A Lei nº 9.784/1999 assegura ao administrado ciência da tramitação, acesso aos autos, apresentação de alegações e documentos antes da decisão e conhecimento das decisões proferidas (BRASIL, 1999). A IA não transforma essas garantias em incompatíveis com eficiência. Ao contrário, sistemas bem desenhados podem registrar razões, organizar documentos e facilitar revisão. O risco nasce quando o processo é construído para que o cidadão encontre apenas uma interface que informa o resultado sem revelar caminho de contestação.

O Poder Judiciário oferece um exemplo setorial de como esses deveres podem ser traduzidos em governança. A Resolução CNJ nº 615/2025, atualizada em 2026, estabelece diretrizes de transparência, explicabilidade, auditabilidade, supervisão humana e gestão de risco para IA judicial. Ela não se aplica indistintamente a toda a Administração Pública, mas demonstra que órgãos estatais podem estabelecer requisitos mais específicos para sistemas que participam de atividades sensíveis (CNJ, 2025; CNJ, 2026). A lição mais ampla é que governança pública precisa acompanhar o impacto institucional da ferramenta.

Em suma, transparência estatal não exige tornar cada sistema totalmente público em todos os seus detalhes, mas exige impedir que a técnica crie uma área imune a controle. O cidadão pode não ter direito a todos os parâmetros internos de um sistema, mas deve conseguir identificar que houve uso relevante de IA, quais dados e critérios principais influenciaram a decisão, quem responde por ela e como contestá-la. Sem isso, a informatização deixa de ser meio de governo e passa a dificultar o próprio controle democrático do poder.

Governança começa antes da contratação do algoritmo

Grande parte dos riscos de IA no setor público é definida antes de o sistema entrar em produção. O órgão escolhe se precisa realmente de inteligência artificial,

quais dados estarão disponíveis, que problema pretende resolver, como medirá sucesso, quais erros serão aceitáveis e se desenvolverá solução própria ou adquirirá tecnologia de terceiro. Essas decisões são menos visíveis que o resultado final, mas condicionam quase tudo o que virá depois. Governança, portanto, não pode começar quando surge a primeira reclamação.

A própria política federal recente caminha nessa direção. O Plano Brasileiro de Inteligência Artificial, publicado em versão final em 2025 pelo Ministério da Ciência, Tecnologia e Inovação (MCTI) e pelo Centro de Gestão e Estudos Estratégicos (CGEE), inclui ações voltadas à experimentação de projetos no governo, monitoramento da adoção e capacitação de servidores públicos, dentro de uma agenda mais ampla de infraestrutura, pesquisa e inovação (MCTI; CGEE, 2025). O plano não cria, por si, direitos subjetivos nem substitui normas administrativas, mas demonstra que o uso de IA no Estado deixou de ser iniciativa episódica de alguns órgãos e passou a integrar planejamento público estruturado.

No Ministério da Gestão e da Inovação em Serviços Públicos (MGI), a Estratégia de Inteligência Artificial 2025-2027 adota princípios de direitos humanos, dignidade, equidade, transparência, explicabilidade e auditabilidade. É importante registrar o limite jurídico do documento: trata-se de estratégia interna do MGI, não de norma geral aplicável automaticamente a toda a Administração. Ainda assim, ela evidencia uma tendência relevante de institucionalizar capacidade, papéis, processos e critérios de governança antes da expansão do uso de IA (MGI, 2025).

O GuIA Unificado de Inteligência Artificial para o Setor Público, desenvolvido pela Secretaria de Governo Digital (SGD) no âmbito do Governo Digital, segue lógica de ciclo de vida. A ferramenta organiza projetos desde prospecção e estruturação até experimentação, implementação, sustentação e desativação, incluindo verificações de proteção de dados, avaliação de impacto ético, definição de responsabilidades, métricas, segurança, monitoramento e critérios de encerramento (SGD, s.d.). O valor desse modelo está em lembrar que uma solução pública não termina quando é entregue pelo fornecedor. Modelos mudam, dados mudam, desempenho pode degradar e finalidades podem se expandir.

A contratação pública é um dos pontos mais delicados. Se o órgão adquire um sistema sem prever acesso a logs, documentação, métricas de desempenho, notificações de atualização, regras de retenção de dados e condições de auditoria, pode descobrir posteriormente que não possui informação suficiente para explicar suas

próprias decisões. Dependência tecnológica não é apenas problema econômico; pode transformar-se em déficit de accountability. Um fornecedor privado pode operar a tecnologia, mas a Administração continua precisando demonstrar por que o sistema é adequado à finalidade pública e como controla seu uso.

Isso exige atenção especial a atualizações. Serviços de IA em nuvem podem mudar sem que o usuário perceba a dimensão da alteração. Uma nova versão pode melhorar desempenho médio e piorar um grupo específico; filtros podem ser modificados; fontes podem ser atualizadas; comportamentos generativos podem mudar. Em sistemas relevantes para decisões públicas, mudanças significativas precisam ser documentadas e, conforme o risco, revalidadas. Não é razoável que a Administração continue atribuindo efeitos jurídicos a um modelo cuja versão ou comportamento já não consegue identificar.

A capacitação dos servidores é igualmente parte da legalidade prática. Um agente público que não compreende limitações básicas de modelos generativos pode incorporar informação inventada em parecer, revelar dado sigiloso em ferramenta externa ou confiar excessivamente em uma recomendação. O Governo Digital já orienta servidores a não inserir informações sigilosas, confidenciais ou dados pessoais em ferramentas públicas de IA generativa sem ambiente institucional adequado e destaca a necessidade de validação humana diante de alucinações e vieses (SGD; SERPRO, s.d.). A orientação é administrativa, não nova fonte de responsabilidade civil ou penal, mas traduz riscos concretos que precisam ser prevenidos.

A governança também deve incluir desativação. Sistemas públicos tendem a permanecer porque foram integrados a rotinas, treinamentos e contratos. Contudo, um modelo que perdeu desempenho, tornou-se incompatível com nova legislação ou deixou de oferecer vantagem proporcional ao risco precisa poder ser retirado. Desativar exige preservar registros necessários, garantir continuidade do serviço e impedir que dependências técnicas tornem a decisão impossível. Um sistema responsável é também um sistema do qual o Estado consegue sair.

O controle externo e social completa essa arquitetura. A Lei de Acesso à Informação, os órgãos de controle, a proteção de dados, o Judiciário e mecanismos de participação podem funcionar como camadas complementares. A Controladoria-Geral da União incluiu entre ações do Plano de Integridade e Combate à Corrupção 2025-2027 o fomento a estudos e boas práticas sobre transparência algorítmica no Poder Executivo Federal, com horizonte de implementação até 2026 (CGU, 2025).

Esse tipo de agenda mostra que controle algorítmico não é apenas tema de equipes de tecnologia; tornou-se questão de integridade pública.

A ausência, em agosto de 2026, de uma lei geral brasileira de inteligência artificial em vigor não transforma esse cenário em vazio jurídico. A Constituição Federal, o processo administrativo, o acesso à informação, a proteção de dados, o governo digital e as normas setoriais já limitam o uso estatal da tecnologia. O futuro marco geral poderá acrescentar obrigações e uniformizar conceitos, mas não cria do zero deveres de legalidade, motivação, proporcionalidade, transparência e responsabilidade que já pertencem ao Direito Público.

A adoção responsável de IA pelo Estado depende, enfim, de uma pergunta simples: o órgão continuaria capaz de defender a decisão se o sistema fosse colocado sob escrutínio? Se não consegue indicar a finalidade, os dados relevantes, a autoridade responsável, os critérios de validação, a forma de revisão e os limites da ferramenta, o problema não é apenas técnico. É institucional. A eficiência administrativa só é legítima quando permanece compatível com o modo democrático de exercer poder.

É nesse ponto que as exigências examinadas ao longo do livro convergem. Deepfakes, dados, responsabilidade civil e penal, prova, decisões automatizadas, consumo e poder público parecem problemas distintos, mas todos dependem da mesma capacidade institucional: transformar princípios e regras em mecanismos de governança que acompanhem tecnologias em rápida mudança. A questão transversal, portanto, não é encontrar uma fórmula universal para "regular a IA", e sim identificar quais princípios, instituições e procedimentos permitem inovar sem converter velocidade tecnológica em perda de direitos.

CAPÍTULO 10

ÉTICA, REGULAÇÃO E PERSPECTIVAS JURÍDICAS PARA A INTELIGÊNCIA ARTIFICIAL

Princípios sem mecanismos viram promessa

Ao longo desta obra, a inteligência artificial apareceu menos como uma tecnologia isolada do que como uma força capaz de modificar relações jurídicas já conhecidas. Ela permitiu sintetizar rostos e vozes, inferir características que uma pessoa nunca declarou, distribuir decisões por cadeias difíceis de reconstruir, ampliar fraudes, colocar em dúvida a autenticidade de provas, orientar escolhas de consumidores e apoiar o exercício de autoridade pública. Esses problemas parecem diferentes porque pertencem a ramos distintos do Direito. Ainda assim, todos convergem para uma questão comum: como impedir que complexidade tecnológica se transforme em perda de capacidade de identificar quem exerce poder, com base em quê e sob quais limites.

Essa pergunta explica por que ética e regulação não podem ser tratadas como capítulos separados da realidade jurídica. A ética ajuda a formular valores que devem orientar desenvolvimento e uso; a regulação define deveres, procedimentos, proibições e consequências; a governança organiza como tudo isso será executado dentro das instituições. Quando essas dimensões se afastam, surgem dois riscos simétricos. De um lado, princípios elevados podem se transformar em linguagem reputacional sem capacidade de alterar práticas. De outro, regras detalhadas podem se tornar rapidamente inadequadas se ignorarem a forma como sistemas são desenvolvidos, atualizados e utilizados.

A proliferação de declarações éticas sobre inteligência artificial mostrou uma convergência internacional importante, mas também seus limites. Jobin, Ienca e Vayena (2019), ao analisarem dezenas de documentos de ética em IA, identificaram recorrência de princípios como transparência, justiça, não maleficência, responsabilidade e privacidade, ao mesmo tempo em que observaram divergências significativas quanto à interpretação e à implementação desses valores. A existência de vocabulário comum, portanto, não significa consenso sobre o que deve ser feito diante de um sistema concreto.

Mittelstadt (2019) aprofunda exatamente esse problema ao sustentar que princípios, sozinhos, não garantem uma inteligência artificial ética. Valores abstratos precisam ser traduzidos em decisões institucionais, critérios de projeto,

responsabilidades e mecanismos de avaliação. Dizer que um modelo deve ser "justo" não informa qual conceito de justiça será utilizado, quais diferenças entre grupos serão toleradas ou quem decidirá quando objetivos legítimos entram em conflito. Dizer que deve existir "supervisão humana" não esclarece se a pessoa possui informação, tempo e autoridade para discordar. O princípio começa a produzir efeito apenas quando encontra um procedimento.

Isso não diminui a importância dos marcos éticos. A Recomendação da UNESCO sobre Ética da Inteligência Artificial, adotada em 2021, organiza sua abordagem a partir da dignidade humana, dos direitos humanos, da diversidade, da inclusão e da proteção ambiental, além de associar princípios a áreas concretas de ação pública (UNESCO, 2021). Os Princípios de IA da OCDE, atualizados em 2024, também relacionam direitos humanos, transparência, robustez, segurança e accountability a mecanismos de supervisão, rastreabilidade e gestão contínua de riscos (OCDE, 2024). A relevância desses instrumentos está justamente em aproximar valores de práticas que possam ser observadas e avaliadas.

O mesmo movimento aparece em estruturas técnicas de gestão. O AI Risk Management Framework do NIST organiza a gestão de riscos nas funções governar, mapear, medir e gerenciar e trata o risco como algo que deve ser acompanhado ao longo do ciclo de vida, e não verificado uma única vez antes da implantação (NIST, 2023). Como se trata de instrumento voluntário dos Estados Unidos, ele não cria obrigação jurídica no Brasil. Seu valor para esta discussão é metodológico: responsabilidade não pode depender apenas da intenção ética de quem desenvolve um sistema; precisa de processos capazes de identificar, registrar, medir e corrigir riscos.

Os nove capítulos anteriores mostraram a mesma conclusão por caminhos diferentes. Deepfakes exigem mais que um apelo genérico à honestidade: precisam de proteção da personalidade, regras de publicidade, meios de remoção e mecanismos de autenticação. Proteção de dados exige mais que consentimento formal: precisa de finalidade, necessidade, segurança e possibilidade de revisão. Responsabilidade civil depende de reconstruir controle e causalidade; responsabilidade penal exige imputação pessoal; prova digital exige preservação e contraditório; decisões automatizadas exigem explicação e contestabilidade. Em todos esses campos, valores só se tornam jurídicos quando podem orientar conduta, produzir prova e sustentar controle.

É por isso que "IA responsável" não deveria funcionar como selo autoatribuído. Uma organização não demonstra responsabilidade porque publica princípios; demonstra quando consegue mostrar quais sistemas utiliza, quem responde por eles, quais riscos identificou, que dados e finalidades são relevantes, como testa desempenho, o que registra, como recebe contestação e em quais condições está disposta a suspender ou abandonar a ferramenta. A ética fornece direção. A governança fornece evidência de que essa direção realmente influencia decisões.

Regular por risco exige um piso de direitos

A regulação por risco tornou-se uma das respostas mais influentes à heterogeneidade da inteligência artificial. A ideia é intuitiva: um recurso de correção ortográfica não deveria suportar os mesmos deveres de um sistema utilizado para selecionar candidatos, controlar infraestrutura crítica ou apoiar decisões judiciais. Quanto maior a capacidade de causar impacto grave e difícil de reparar, maiores devem ser as exigências de documentação, supervisão, validação e controle. O problema não está na lógica, mas em imaginar que uma classificação inicial possa resolver sozinha o significado jurídico de um sistema.

Risco depende de contexto. Um modelo de propósito geral pode ser pouco relevante em uma aplicação e decisivo em outra. Um sistema inicialmente desenvolvido para organizar informação pode, depois de integrado a um fluxo institucional, passar a influenciar benefícios, crédito ou fiscalização. Uma ferramenta de baixo impacto individual pode alcançar milhões de pessoas e produzir efeito agregado importante. Por isso, classificação de risco precisa acompanhar finalidade, população atingida, escala, reversibilidade do dano e grau de dependência do resultado. Sistemas mudam de posição regulatória quando seu uso muda.

O Regulamento Europeu de Inteligência Artificial representa o exemplo mais desenvolvido de arquitetura horizontal baseada em risco. O AI Act combina práticas proibidas, deveres de transparência, regras para modelos de propósito geral e requisitos específicos para sistemas classificados como de alto risco (UNIÃO EUROPEIA, 2024a). O Regulamento tornou-se aplicável, em termos gerais, em 2 de agosto de 2026. O Regulamento (UE) 2026/1744, porém, deslocou a aplicação das Seções 1, 2 e 3 do Capítulo III para 2 de dezembro de 2027 nos sistemas de alto risco classificados pelo art. 6º, nº 2, do Regulamento (UE) 2024/1689 e pelo Anexo III, e para 2 de agosto de 2028 nos sistemas enquadrados pelo art. 6º, nº 1, do Regulamento (UE) 2024/1689 e pelo Anexo I (UNIÃO EUROPEIA, 2024a; UNIÃO EUROPEIA,

2026). A própria transição regulatória mostra que governar tecnologia exige capacidade de ajustar implementação sem abandonar objetivos de proteção.

Uma regulação por risco, contudo, precisa de um piso que não desapareça quando o sistema recebe rótulo de baixo risco. Direitos da personalidade, igualdade, proteção de dados, liberdade de expressão, devido processo, defesa do consumidor e responsabilidade não nascem da classificação administrativa de um modelo. Eles continuam incidindo quando os pressupostos jurídicos estiverem presentes. Um conteúdo sintético aparentemente trivial pode violar imagem; um sistema comercial de recomendação pode praticar publicidade enganosa; uma ferramenta de baixa complexidade pode tratar dados de forma ilícita. Classificação de risco organiza deveres adicionais, mas não substitui o ordenamento.

Essa observação é especialmente importante no Brasil. Em 25 de agosto de 2026, o Projeto de Lei nº 2.338/2023 continua aguardando parecer na Comissão Especial da Câmara dos Deputados. O texto recebido do Senado propõe uma estrutura geral baseada em centralidade da pessoa humana, direitos, gestão de riscos e deveres para diferentes agentes, mas permanece projeto legislativo, não lei em vigor (CÂMARA DOS DEPUTADOS, 2026a; CÂMARA DOS DEPUTADOS, 2026b). O debate sobre um marco geral é relevante; apresentá-lo como se já tivesse encerrado a questão seria juridicamente incorreto.

A ausência de uma lei geral vigente tampouco significa ausência de regulação. A Constituição Federal, a LGPD, o Código Civil, o Código de Defesa do Consumidor, o Direito Penal, a legislação processual, as normas eleitorais, as regras de governo digital e a regulação setorial já alcançam diferentes usos de IA. O CNJ estabeleceu governança própria para soluções utilizadas no Judiciário; a proteção de dados vem produzindo parâmetros para treinamento, decisões automatizadas e transparência; o Direito Eleitoral já disciplinou conteúdos sintéticos em contexto de campanha. O sistema jurídico brasileiro é fragmentado, mas não vazio.

Essa fragmentação pode até ser funcional quando diferentes setores apresentam riscos próprios. Um sistema de apoio clínico, um algoritmo de crédito e uma ferramenta judicial não deveriam ser governados apenas por uma regra horizontal genérica. Autoridades e normas setoriais possuem conhecimento sobre consequências específicas, deveres profissionais e padrões de segurança de cada campo. O desafio é impedir que regulação horizontal e setorial produzam lacunas, contradições ou

duplicação impossível de cumprir. Coordenação institucional passa a ser parte da qualidade regulatória.

No plano internacional, a Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial, Direitos Humanos, Democracia e Estado de Direito oferece uma perspectiva complementar. Aberta à assinatura em 2024, ela organiza a discussão ao redor de direitos humanos, democracia e Estado de Direito, e não de uma arquitetura tecnológica específica (CONSELHO DA EUROPA, 2024). Essa escolha é relevante porque tecnologias mudam mais rapidamente que valores constitucionais. Um regime durável precisa ser capaz de reconhecer novas formas de exercício de poder sem reescrever, a cada modelo lançado, a proteção básica da pessoa.

A regulação também precisa preservar espaço para experimentação responsável. Exigir prova impossível de risco zero antes de qualquer inovação pode bloquear benefícios sociais relevantes e favorecer apenas organizações grandes o suficiente para suportar altos custos de conformidade. O extremo oposto — liberar tudo até que o dano ocorra — transforma usuários e cidadãos em sujeitos involuntários de experimentos. Entre esses polos, ambientes controlados, testes limitados, monitoramento e exigências proporcionais podem produzir evidência antes de expansão.

O sandbox regulatório da Agência Nacional de Proteção de Dados (ANPD) ilustra essa estratégia. Em 19 de agosto de 2026, a ANPD publicou a metodologia de testagem de seu projeto-piloto em IA e proteção de dados, com procedimentos de supervisão, monitoramento e avaliação voltados à transparência, explicabilidade, proteção de dados e gestão de riscos (ANPD, 2026b). Um sandbox não suspende direitos nem funciona como licença para violar a lei. Sua função regulatória é permitir aprendizado em ambiente supervisionado, produzindo evidências que ajudem a calibrar regras e práticas.

O Plano Brasileiro de Inteligência Artificial segue lógica semelhante ao combinar incentivo à inovação, infraestrutura, formação e projetos de experimentação no governo com ações de monitoramento da adoção de IA (MCTI; CGEE, 2025). O plano não substitui legislação nem resolve conflitos jurídicos, mas mostra que política pública de IA precisa trabalhar simultaneamente com capacidade tecnológica e capacidade institucional. Um país que apenas regula o risco sem desenvolver competência técnica pode depender de modelos e padrões definidos fora de sua

realidade; um país que apenas acelera adoção sem governança pode ampliar vulnerabilidades.

Regular por risco, portanto, não significa escolher entre inovação e direitos. Significa reconhecer que diferentes aplicações exigem intensidades diferentes de controle, enquanto certos compromissos permanecem constantes: uma pessoa não deveria perder proteção de sua identidade porque a falsificação foi automatizada; não deveria ficar sem reparação porque a cadeia técnica é complexa; não deveria ser condenada por uma máquina; não deveria ter uma prova aceita apenas porque parece verdadeira; e não deveria ficar impossibilitada de contestar uma decisão porque seus fundamentos estão espalhados por um modelo proprietário.

Governança é o lugar onde a regra prova que funciona

A pergunta decisiva para o futuro jurídico da inteligência artificial não é quantos princípios ou quantas leis serão publicados, mas se instituições conseguirão transformar esses comandos em prática cotidiana. Governança é o ponto em que contrato, desenho técnico, treinamento, documentação, auditoria, supervisão e recurso deixam de ser questões separadas. É ali que se define se a pessoa protegida pelo Direito encontrará efetivamente alguém capaz de responder quando o sistema produz um resultado errado ou injusto.

Uma boa governança começa antes da aquisição ou desenvolvimento. A organização precisa saber qual problema pretende resolver e por que IA é necessária. Muitas falhas surgem porque tecnologia é escolhida antes de a finalidade estar clara. Se um método simples atende ao objetivo com menor risco, não existe mérito jurídico em utilizar modelo mais complexo apenas porque é mais moderno. Adequação inclui a escolha de não automatizar quando o ganho não justifica perda de controle, custo, exposição de dados ou dificuldade de explicação.

Depois vem a definição de papéis. Quem seleciona dados? Quem valida desempenho? Quem autoriza implantação? Quem recebe alertas? Quem pode alterar parâmetros? Quem verifica impacto sobre grupos diferentes? Quem responde à contestação? Quem possui autoridade para interromper o sistema? Essas perguntas apareceram ao longo de toda a obra porque responsabilidade distribuída não pode se tornar responsabilidade inexistente. Quanto mais fornecedores e componentes participam, maior a necessidade de registrar fronteiras de obrigação e canais de cooperação.

Documentação precisa acompanhar essa arquitetura. Um sistema juridicamente governável deve permitir saber qual versão estava em operação, que finalidade justificava o uso, quais limites eram conhecidos e que alterações ocorreram. Logs, avaliações de impacto, registros de testes e histórico de atualizações não servem apenas à engenharia. Eles sustentam explicação, auditoria, prova, defesa e reparação. Sem rastreabilidade, direitos podem existir no papel e se tornar difíceis de exercer justamente porque ninguém consegue reconstruir o que aconteceu.

Monitoramento é igualmente indispensável porque conformidade não é estado permanente. Dados mudam, populações mudam, comportamentos se adaptam e modelos são atualizados. Um sistema que cumpria determinada finalidade pode ser reutilizado em contexto mais sensível. Uma ferramenta que apresentava desempenho aceitável pode degradar. Um risco inicialmente improvável pode se tornar recorrente depois que agentes aprendem a explorar o sistema. O NIST e os Princípios da OCDE convergem nesse ponto ao tratar gestão de risco como atividade contínua durante o ciclo de vida (NIST, 2023; OCDE, 2024).

A possibilidade de intervenção precisa ser real. Governança responsável pressupõe que sistemas possam ser corrigidos, limitados, desativados ou substituídos. A ideia de supervisão humana perde sentido se ninguém consegue interromper a operação quando surge comportamento inadequado. O mesmo vale para regulação: exigir avaliação e monitoramento sem prever consequências quando o risco se materializa transforma controle em observação passiva. Uma organização responsável não é aquela que nunca encontra problemas; é aquela que possui capacidade e obrigação de reagir a eles.

Isso também significa aprender com contestação. Reclamações de consumidores, recursos administrativos, impugnações judiciais, incidentes de segurança, divergências de profissionais e auditorias não são apenas falhas a serem encerradas. São fontes de informação sobre como o sistema opera fora das condições de teste. Quando reclamações se repetem em determinado grupo ou contexto, o problema pode estar menos no indivíduo que "não entendeu" e mais na arquitetura que está produzindo efeitos sistemáticos.

A formação das pessoas que utilizam IA completa esse quadro. Não basta ensinar como escrever comandos ou operar interfaces. Letramento jurídico e algorítmico precisa incluir limites, risco de automação, proteção de dados, deveres de confidencialidade, incerteza, possibilidade de erro e responsabilidade sobre o uso. Um

profissional que desconhece que um modelo generativo pode inventar informações não consegue supervisioná-lo adequadamente; um gestor que não compreende viés ou mudança de contexto pode confiar em métricas que escondem dano.

A governança democrática exige ainda pluralidade de perspectivas. Quem desenvolve um sistema enxerga desempenho; quem compra pode priorizar custo e eficiência; quem utiliza percebe limitações operacionais; quem sofre a decisão percebe efeitos que métricas internas talvez não capturem. Participação não significa submeter cada atualização a plebiscito, mas reconhecer que avaliação de risco se torna mais completa quando inclui pessoas e grupos afetados. A tecnologia pode ser complexa; as consequências não são propriedade exclusiva de especialistas.

É nesse ponto que ética, regulação e Direito se reencontram. A ética pergunta que sociedade queremos construir com sistemas capazes de ampliar decisões e produzir representações. A regulação estabelece limites e mecanismos para reduzir riscos. O Direito oferece categorias para proteger direitos, atribuir deveres, produzir prova e reparar violações. A governança conecta essas dimensões à prática institucional. Nenhuma delas, isoladamente, é suficiente.

O futuro jurídico da inteligência artificial provavelmente será menos definido por uma grande resposta única do que por essa combinação de camadas. Normas gerais conviverão com legislação preexistente, regras setoriais, padrões técnicos, decisões judiciais, instrumentos internacionais, contratos e procedimentos internos. Isso pode parecer fragmentado, mas a complexidade reflete a própria diversidade dos usos. O que precisa permanecer coerente não é a forma de todas as regras, e sim o núcleo de proteção que elas procuram realizar.

Esse núcleo pode ser reconhecido nas questões que atravessaram o livro. Uma pessoa deve conservar proteção sobre sua identidade quando rosto e voz podem ser sintetizados. Deve saber como seus dados são utilizados quando sistemas extraem inferências. Deve existir alguém juridicamente responsável quando a tecnologia causa dano. O Direito Penal deve preservar imputação pessoal. A prova deve continuar auditável quando a aparência já não garante autenticidade. Decisões automatizadas precisam poder ser compreendidas e contestadas. Consumidores não podem ter vulnerabilidades exploradas em escala. O Estado não pode trocar motivação por pontuação.

Nenhuma dessas exigências depende de imaginar que a tecnologia seja inimiga. Sistemas de inteligência artificial podem ampliar acesso a informação, melhorar

serviços, reduzir tarefas repetitivas, apoiar pesquisa, detectar fraude e ajudar profissionais a lidar com complexidade. O problema começa quando benefício técnico é utilizado como argumento para dispensar as condições jurídicas que tornam legítimo o exercício de poder. Inovação e direitos não ocupam lados opostos de uma mesma equação; inovação juridicamente sustentável é justamente aquela que consegue produzir benefício sem tornar direitos impraticáveis.

Por isso, o melhor Direito para a inteligência artificial talvez não seja o que pretenda prever todas as tecnologias que ainda surgirão. É o que preserva perguntas capazes de sobreviver à mudança tecnológica: quem está sendo afetado? Qual direito está em jogo? Quem definiu a finalidade? Quem possui controle? Que informação foi usada? Como o risco foi avaliado? Como a decisão pode ser contestada? Quem pode corrigir ou interromper o sistema? Quem responde quando o dano ocorre? Essas perguntas são mais duráveis do que qualquer arquitetura de modelo.

Ao final, o teste jurídico mais importante de um sistema inteligente não está em quão convincente ele parece nem em quantas tarefas consegue automatizar. Está em saber se, diante de erro, dano ou contestação, ainda conseguimos reconstruir quem decidiu, com base em quê, sob qual autoridade e por qual razão. Se essas respostas desaparecem dentro da tecnologia, o problema não é apenas de explicabilidade: é de legitimidade. A inteligência artificial pode transformar instrumentos, escalas e formas de decisão. O Direito continua responsável por impedir que essa transformação torne invisíveis as pessoas, os poderes e as consequências que permanecem por trás de cada sistema.

CONSIDERAÇÕES FINAIS

A análise desenvolvida nesta obra indica que a inteligência artificial não inaugura um território juridicamente vazio. O que ela faz é tensionar categorias já conhecidas ao ampliar escala, velocidade, opacidade e capacidade de produzir resultados que interferem em direitos. Identidade, privacidade, proteção de dados, liberdade de expressão, responsabilidade, prova, devido processo, consumo e exercício do poder público continuam sendo problemas jurídicos reconhecíveis. A novidade está em como esses problemas se reorganizam quando uma decisão pode resultar de cadeias técnicas complexas, quando conteúdos falsos alcançam aparência de autenticidade e quando a pessoa afetada encontra dificuldade para identificar quem definiu a finalidade, quais dados foram utilizados e quem possuía capacidade efetiva de evitar ou corrigir o dano. Por isso, a pergunta central não é se o Direito consegue acompanhar cada nova ferramenta, mas se consegue preservar condições de legitimidade quando a tecnologia redistribui informação, poder e responsabilidade.

Essa conclusão aparece com especial nitidez no debate sobre deepfakes e conteúdos sintéticos. A possibilidade de reproduzir rosto, voz e gestos com elevado grau de verossimilhança enfraquece uma antiga associação entre aparência e presença, mas não elimina as distinções jurídicas entre fraude, sátira, criação artística, informação, publicidade e violência. Uma resposta adequada não pode pressupor que todo conteúdo sintético seja ilícito, nem tratar consentimento como solução universal. Direitos da personalidade, liberdade de expressão, proteção de dados e integridade de processos coletivos precisam ser ponderados conforme contexto, finalidade, circulação e dano. A evolução recente das normas eleitorais brasileiras e das regras europeias de transparência mostra que o problema não se resolve apenas pela origem artificial do conteúdo: importa saber se o público deve ser informado, se existe finalidade ilícita, se a identidade de alguém foi instrumentalizada e quais mecanismos permitem interromper ou responsabilizar usos abusivos (TSE, 2026; UNIÃO EUROPEIA, 2024a).

No campo da proteção de dados, a inteligência artificial torna mais visível a distância entre coletar informação e produzir conhecimento sobre uma pessoa. Sistemas capazes de inferir preferências, riscos, condições e perfis demonstram que a proteção jurídica não pode se limitar ao dado fornecido diretamente pelo titular. A Lei Geral de Proteção de Dados Pessoais oferece instrumentos relevantes para enfrentar esse deslocamento ao regular tratamento, finalidade, necessidade, não

discriminação e decisões automatizadas, mas a aplicação do art. 20 da LGPD continua exigindo interpretação cuidadosa. A existência de participação humana formal, por exemplo, não deveria funcionar como artifício para esvaziar direitos quando a decisão material permanece determinada pelo sistema. A discussão regulatória conduzida pela ANPD confirma que revisão, transparência e contestabilidade ainda estão em processo de densificação no ordenamento brasileiro (ANPD, 2025a; ANPD, 2026a).

A responsabilidade civil também precisa evitar a tentação de transformar a própria inteligência artificial em sujeito abstrato do dano. Sistemas são concebidos, treinados, integrados, fornecidos, atualizados e utilizados por atores distintos, com graus diferentes de controle e acesso à informação. A tarefa jurídica é reconstruir essa cadeia sem diluir a responsabilidade entre tantos participantes que, ao final, ninguém responda. Código Civil, Código de Defesa do Consumidor e regimes setoriais já fornecem categorias para examinar ilícito, defeito, risco, nexo causal, dever de informação e distribuição do ônus probatório. A inovação legislativa pode aperfeiçoar documentação, acesso à prova e deveres de governança, mas não deve substituir a análise concreta da relação jurídica por uma fórmula genérica segundo a qual todo dano envolvendo IA teria a mesma natureza. Quanto mais distribuído for o processo técnico, mais importante se torna identificar quem controlava a finalidade, quem podia conhecer o risco e quem tinha condições reais de agir.

No Direito Penal, a prudência precisa ser ainda maior. A velocidade tecnológica não autoriza flexibilizar legalidade, pessoalidade da pena e culpabilidade. Em grande parte dos casos, a inteligência artificial modifica o meio de execução de condutas já tipificadas, sem criar por si mesma um novo centro de imputação. Clonagem de voz, automação de fraude, geração de código malicioso e produção sintética de imagens podem integrar crimes existentes quando seus elementos legais estiverem presentes. As alterações legislativas recentes demonstram que o ordenamento pode responder de forma pontual a novas formas de lesão sem recorrer a uma categoria indistinta de "crime de inteligência artificial". Esse caminho preserva a exigência de vincular punição a conduta, dolo ou culpa quando prevista, participação e dever jurídico de agir, evitando que proximidade técnica com um sistema seja confundida com responsabilidade penal automática.

A prova digital apresenta talvez uma das transformações mais profundas, porque sistemas generativos afetam não apenas a produção de evidências, mas a confiança social em sua aparência. Cadeia de custódia, integridade criptográfica e preservação de arquivos permanecem indispensáveis, porém não resolvem sozinhas o problema da

autenticidade. Um arquivo pode permanecer íntegro desde a apreensão e ainda conter uma representação sintética produzida antes dela. Por isso, a distinção entre integridade, autenticidade e veracidade precisa ser incorporada à prática processual. Metadados, origem, contexto, perícia, registros de geração, testemunhos e confronto com outras fontes passam a compor uma avaliação probatória mais sofisticada. A resposta jurídica não deve ser abandonar a prova audiovisual, mas abandonar a presunção de que aparência convincente dispensa verificação. Em um ambiente no qual falsificações plausíveis se tornam mais acessíveis, contraditório e auditabilidade ganham importância adicional.

O mesmo princípio vale para decisões automatizadas. Transparência não pode ser confundida com publicação de código-fonte, assim como supervisão humana não pode ser reduzida à presença nominal de uma pessoa ao final do fluxo. O que importa é assegurar informação suficiente para compreender critérios relevantes, identificar a finalidade, contestar a conclusão e permitir que alguém com autoridade efetiva reveja ou interrompa o processo. A experiência do Poder Judiciário brasileiro é ilustrativa ao associar uso de inteligência artificial a explicabilidade, contestabilidade, auditabilidade, avaliação de impacto e supervisão humana. O valor dessas exigências não está em presumir que toda automação seja perigosa, mas em reconhecer que decisões de maior impacto exigem salvaguardas proporcionais e documentação que permita reconstruir o caminho percorrido (CNJ, 2025; CNJ, 2026).

Nas relações de consumo, a personalização algorítmica amplia tanto conveniência quanto capacidade de explorar vulnerabilidades. Sistemas podem adaptar ofertas, preços, interfaces, atendimento e persuasão a perfis individuais em uma escala antes inviável. A assimetria informacional, portanto, não desaparece porque o consumidor recebe uma experiência aparentemente personalizada; em alguns casos, ela pode se aprofundar. O Código de Defesa do Consumidor e a LGPD precisam ser lidos de forma complementar quando dados pessoais alimentam práticas comerciais capazes de influenciar escolhas. Informação adequada, liberdade de decisão, proteção contra publicidade enganosa ou abusiva e vedação de práticas desleais continuam relevantes mesmo quando a pressão é exercida por uma interface adaptativa em vez de por um vendedor. A tecnologia muda a forma da interação, não elimina a vulnerabilidade que justifica a tutela.

Quando a inteligência artificial é utilizada pelo Estado, o nível de exigência deve ser ainda mais alto. Eficiência administrativa é valor legítimo, mas convive com legalidade, impessoalidade, moralidade, publicidade, motivação, contraditório e

devido processo. Um score ou recomendação algorítmica não pode substituir a razão jurídica que legitima um ato público. Se dados, regras ou saídas técnicas influenciaram uma decisão, a Administração precisa preservar informação suficiente para permitir controle, recurso e responsabilização. Isso exige atenção não apenas ao momento de uso, mas também à contratação: cláusulas de sigilo, dependência tecnológica e arquitetura proprietária não podem tornar impossível explicar um ato estatal. Governança pública responsável começa antes da implementação do sistema e inclui critérios de aquisição, testes, documentação, interoperabilidade, auditoria, monitoramento e plano de descontinuidade.

O cenário regulatório analisado no livro reforça a necessidade de trabalhar com precisão temporal. No momento de fechamento desta obra, o Projeto de Lei nº 2.338/2023 permanece em tramitação na Câmara dos Deputados e não pode ser apresentado como direito positivo vigente (CÂMARA DOS DEPUTADOS, 2026a). Ao mesmo tempo, o Brasil não se encontra em ausência de regulação: a Constituição Federal, a LGPD, o Código Civil, o Código de Defesa do Consumidor, a legislação penal e processual, as normas eleitorais, as regras administrativas e os atos setoriais já alcançam diferentes aplicações. A experiência europeia oferece outro modelo, estruturado por níveis de risco e deveres diferenciados, mas sua própria implementação demonstra que regulação tecnológica é processo dinâmico: o Regulamento (UE) 2026/1744 alterou o cronograma de determinadas obrigações de alto risco sem abandonar o desenho geral de proteção (UNIÃO EUROPEIA, 2024a; UNIÃO EUROPEIA, 2026).

Esse caráter dinâmico torna governança mais importante do que conformidade pontual. Um sistema que foi avaliado antes da implantação pode mudar com atualizações, novos dados, novos usos ou mudanças no perfil das pessoas afetadas. A documentação precisa acompanhar o ciclo de vida: finalidade, versão, fontes de dados, limitações conhecidas, testes, incidentes, reclamações e decisões de correção. Monitorar não significa presumir falha, mas reconhecer que desempenho e risco não são propriedades permanentes. Da mesma forma, possuir uma política escrita de supervisão não basta se ninguém tem tempo, informação ou autoridade para discordar do sistema. Governança efetiva exige capacidade de intervir. Os marcos internacionais examinados ao longo da obra convergem justamente na importância de gestão de riscos, rastreabilidade, supervisão e responsabilização, ainda que possuam natureza jurídica distinta (CONSELHO DA EUROPA, 2024; NIST, 2023; OCDE, 2024; UNESCO, 2021).

Há, por fim, uma dimensão profissional que não pode ser terceirizada. Juristas, magistrados, membros do Ministério Público, advogados, gestores, encarregados de dados, peritos e equipes de tecnologia precisam desenvolver linguagem comum suficiente para discutir limites e evidências sem transformar complexidade técnica em autoridade incontestável. Não é necessário que todo profissional do Direito se torne programador, mas é necessário compreender o bastante para perguntar sobre dados, finalidade, taxa de erro, contexto de validação, possibilidade de viés, logs, atualização e formas de contestação. Da mesma maneira, profissionais técnicos precisam compreender que eficiência, acurácia e desempenho médio não esgotam a legitimidade jurídica. O encontro entre essas competências é uma condição para que decisões tecnológicas possam ser analisadas por critérios públicos e não apenas por confiança no fornecedor ou no sistema.

A principal conclusão desta obra, portanto, é que o Direito da inteligência artificial não será construído apenas pela criação de novas leis. Ele será definido pela capacidade de aplicar princípios duradouros a arranjos técnicos mutáveis, identificar quando categorias existentes são suficientes e reconhecer com precisão quando novas obrigações são necessárias. A centralidade da pessoa humana só deixa de ser fórmula abstrata quando se traduz em direitos exercíveis: saber que uma decisão foi automatizada quando isso for relevante, compreender fundamentos em nível útil, contestar resultados, obter revisão, produzir prova, exigir reparação e identificar quem responde. Inovação juridicamente sustentável não é a que elimina toda incerteza, mas a que mantém caminhos de controle mesmo quando a tecnologia é complexa.

Ao final, permanece um critério simples e exigente. Quanto maior o poder de um sistema para classificar, recomendar, sintetizar, persuadir ou decidir, maior deve ser a capacidade de reconstruir sua inserção na relação jurídica. É preciso conseguir responder quem definiu a finalidade, que dados foram utilizados, qual norma autorizava a atuação, que riscos eram conhecidos, quais salvaguardas existiam, quem podia interromper o processo e a quem a pessoa afetada podia recorrer. Quando essas respostas continuam disponíveis, a inteligência artificial pode ser incorporada ao Direito sem dissolver responsabilidade. Quando desaparecem dentro da arquitetura tecnológica, o problema já não é apenas de opacidade computacional: é de legitimidade jurídica. O desafio contemporâneo não consiste em impedir que o Direito mude com a tecnologia, mas em assegurar que, durante essa mudança, direitos, razões e responsabilidades permaneçam visíveis.

REFERÊNCIAS

AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **Nota Técnica nº 54/2025/FIS/CGF/ANPD**. Processo nº 00261.002292/2025-19. Mapa de temas prioritários, com inclusão do tratamento de dados pessoais sensíveis - biometria e do uso de face e voz para criação de mídias por inteligência artificial. Brasília, DF: ANPD, 2025b. Disponível em: https://www.gov.br/anpd/pt-br/acesso-a-informacao/institucional/atos-normativos/regulamentacoes_anpd/nota-tecnica-no-54-2025-fis-cgf-anpd.pdf. Acesso em: 23 ago. 2026.

AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **Decisão automatizada é tema de workshop em Fórum da Internet no Brasil**. Brasília, DF, 2 jun. 2026a. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/decisao-automatizada-e-tema-de-workshop-em-forum-da-internet-no-brasil. Acesso em: 22 ago. 2026.

AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **ANPD publica metodologia de testagem do Sandbox Regulatório em IA e Proteção de Dados**. Brasília, DF, 19 ago. 2026b. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-metodologia-de-testagem-do-sandbox-regulatorio-em-inteligencia-artificial-e-protecao-de-dados. Acesso em: 23 ago. 2026.

ARAÚJO, Pedro; SANTOS, Cássio André Borges dos. Como o Direito Penal pode (ou não) reagir ao uso de inteligência artificial em práticas ilícitas. **Revista Vertentes do Direito**, v. 13, n. 1, p. 48-66, 2026. DOI: 10.20873/uft.2359-0106.2026.v13n1.p48-66.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **Tratamento de dados pessoais pelo Poder Público**. Guia orientativo, versão 2.0. Brasília, DF: ANPD, 2023. Disponível em: https://www.gov.br/anpd/pt-br/centrais-de-conteudo/materiais-educativos-e-publicacoes/guia-poder-publico-anpd-versao-final.pdf. Acesso em: 23 ago. 2026.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **Guia Orientativo: Hipóteses Legais de Tratamento de Dados Pessoais - Legítimo Interesse**. Versão 1.0. Brasília, DF: ANPD, 2024a. Disponível em: https://www.gov.br/anpd/pt-br/centrais-de-conteudo/materiais-educativos-e-publicacoes/guia_orientativo_hipoteses_legais_tratamento_de_dados_pessoais_legitimo_interesse. Acesso em: 23 ago. 2026.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **ANPD determina suspensão cautelar do tratamento de dados pessoais para treinamento da IA da Meta**. Brasília, DF, 2 jul. 2024b. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-determina-suspensao-cautelar-do-tratamento-de-dados-pessoais-para-treinamento-da-ia-da-meta. Acesso em: 23 ago. 2026.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **Meta cumpre exigências da ANPD e poderá retomar, com restrições, o uso de dados pessoais para treinamento de inteligência artificial**. Brasília, DF, 30 ago. 2024c. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/meta-cumpre-exigencias-da-anpd-e-podera-retomar-com-restricoes-o-uso-de-dados-pessoais-para-treinamento-de-inteligencia-artificial. Acesso em: 23 ago. 2026.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **ANPD determina adoção de medidas preventivas à X**. Corp. Brasília, DF, 17 dez. 2024d. Disponível em:

https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-determina-adocao-de-medidas-preventivas-a-x-corp. Acesso em: 23 ago. 2026.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **ANPD apresenta resultados da Tomada de Subsídios sobre Inteligência Artificial e Revisão de Decisões Automatizadas**. Brasília, DF, 19 maio 2025a. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-apresenta-resultados-da-tomada-de-subsidios-sobre-tratamento-automatizado-de-dados-pessoais/. Acesso em: 22 ago. 2026.

BRASIL. **Decreto-Lei nº 2.848, de 7 de dezembro de 1940**. Código Penal. Brasília, DF: Presidência da República, 1940. Texto compilado. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848.htm. Acesso em: 23 ago. 2026.

BRASIL. **Decreto-Lei nº 3.689, de 3 de outubro de 1941**. Código de Processo Penal. Brasília, DF: Presidência da República, 1941. Texto compilado. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del3689compilado.htm. Acesso em: 23 ago. 2026.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 1988.

BRASIL. **Lei nº 8.078, de 11 de setembro de 1990**. Dispõe sobre a proteção do consumidor e dá outras providências. Brasília, DF: Presidência da República, 1990.

BRASIL. **Lei nº 9.784, de 29 de janeiro de 1999**. Regula o processo administrativo no âmbito da Administração Pública Federal. Brasília, DF: Presidência da República, 1999. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9784.htm. Acesso em: 23 ago. 2026.

BRASIL. **Lei nº 10.406, de 10 de janeiro de 2002**. Institui o Código Civil. Brasília, DF: Presidência da República, 2002.

BRASIL. **Lei nº 12.414, de 9 de junho de 2011**. Disciplina a formação e consulta a bancos de dados com informações de adimplemento, de pessoas naturais ou de pessoas jurídicas, para formação de histórico de crédito. Brasília, DF: Presidência da República, 2011a. Texto compilado. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12414.htm. Acesso em: 23 ago. 2026.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011**. Regula o acesso a informações e dá outras providências (Lei de Acesso à Informação). Brasília, DF: Presidência da República, 2011b. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 23 ago. 2026.

BRASIL. **Lei nº 12.965, de 23 de abril de 2014**. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. Brasília, DF: Presidência da República, 2014.

BRASIL. **Lei nº 13.105, de 16 de março de 2015**. Código de Processo Civil. Brasília, DF: Presidência da República, 2015. Texto compilado. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105compilada.htm. Acesso em: 23 ago. 2026.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709compilado.htm. Acesso em: 22 ago. 2026.

BRASIL. **Lei nº 14.129, de 29 de março de 2021**. Dispõe sobre princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública. Brasília, DF: Presidência da República, 2021a. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/l14129.htm. Acesso em: 23 ago. 2026.

BRASIL. **Lei nº 14.181, de 1º de julho de 2021**. Altera o Código de Defesa do Consumidor e o Estatuto do Idoso para aperfeiçoar a disciplina do crédito ao consumidor e dispor sobre a prevenção e o tratamento do superendividamento. Brasília, DF: Presidência da República, 2021b. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/l14181.htm. Acesso em: 23 ago. 2026.

BRASIL. **Decreto nº 11.034, de 5 de abril de 2022**. Regulamenta a Lei nº 8.078, de 11 de setembro de 1990, para estabelecer diretrizes e normas sobre o Serviço de Atendimento ao Consumidor. Brasília, DF: Presidência da República, 2022. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2022/decreto/d11034.htm. Acesso em: 23 ago. 2026.

BRASIL. **Lei nº 15.123, de 24 de abril de 2025**. Altera o art. 147-B do Código Penal para estabelecer causa de aumento de pena no crime de violência psicológica contra a mulher quando praticado com uso de inteligência artificial ou de recurso tecnológico que altere imagem ou som da vítima. Brasília, DF: Presidência da República, 2025a. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/l15123.htm. Acesso em: 23 ago. 2026.

BRASIL. **Lei nº 15.211, de 17 de setembro de 2025**. Dispõe sobre a proteção de crianças e adolescentes em ambientes digitais (Estatuto Digital da Criança e do Adolescente). Brasília, DF: Presidência da República, 2025b. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/l15211.htm. Acesso em: 23 ago. 2026.

BRASIL. **Lei nº 15.397, de 30 de abril de 2026**. Altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), a fim de majorar as penas previstas para os crimes de furto, roubo, estelionato, receptação, receptação de animal e interrupção ou perturbação de serviço telegráfico, telefônico, informático, telemático ou de informação de utilidade pública, bem como para tipificar os crimes de receptação de animal doméstico e de fraude bancária. Brasília, DF: Presidência da República, 2026a. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2026/lei/l15397.htm. Acesso em: 23 ago. 2026.

BRASIL. **Lei nº 15.487, de 6 de agosto de 2026**. Institui medidas de enfrentamento e repressão ao crime de violência sexual contra criança ou adolescente, inclusive no ambiente digital e com uso de inteligência artificial; e altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), o Decreto-Lei nº 3.689, de 3 de outubro de 1941 (Código de Processo Penal), a Lei nº 8.069, de 13 de julho de 1990 (Estatuto da Criança e do Adolescente), a Lei nº 8.072, de 25 de julho de 1990 (Lei dos Crimes Hediondos), e a Lei nº 12.850, de 2 de agosto de 2013 (Lei de Combate ao Crime Organizado), para recrudescer o tratamento penal aos criminosos sexuais. Brasília, DF: Presidência da República, 2026b. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2026/lei/l15487.htm. Acesso em: 23 ago. 2026.

BRASIL. **Decreto nº 12.880, de 18 de março de 2026**. Regulamenta a Lei nº 15.211, de 17 de setembro de 2025, que dispõe sobre a proteção de crianças e adolescentes em ambientes digitais, e institui a Política Nacional de Promoção e Proteção dos Direitos da Criança e do Adolescente no Ambiente Digital. Brasília, DF: Presidência da República, 2026c. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2026/decreto/d12880.htm. Acesso em: 23 ago. 2026.

BRASIL. **Decreto nº 12.975, de 20 de maio de 2026**. Altera o Decreto nº 8.771, de 11 de maio de 2016, que regulamenta a Lei nº 12.965, de 23 de abril de 2014. Brasília, DF: Presidência da República, 2026d. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2026/decreto/d12975.htm. Acesso em: 23 ago. 2026.

CÂMARA DOS DEPUTADOS. **Projeto de Lei nº 2.338, de 2023: ficha de tramitação**. Brasília, DF, 2026a. Situação em 23 ago. 2026: aguardando parecer do relator na Comissão Especial. Disponível em: https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2487262. Acesso em: 23 ago. 2026.

CÂMARA DOS DEPUTADOS. **Projeto de Lei nº 2.338, de 2023: texto recebido do Senado Federal**. Brasília, DF, 2026b. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2868197&filename=Tramitacao-PL+2338%2F2023. Acesso em: 22 ago. 2026.

CARLINI, Nicholas et al. Extracting training data from large language models. In: 30th USENIX SECURITY SYMPOSIUM (USENIX Security 21), 2021. **Proceedings [...]**. Berkeley: USENIX Association, 2021. p. 2633-2650.

CARVALHO, Rodrigo Arruda de. A inteligência artificial na Administração Pública brasileira: avançando na identificação de desafios jurídicos. **Revista Eletrônica do CEJUR**, v. 4, n. 1, 2026. DOI: 10.5380/cejur.v4i1.99308.

CHAVES JUNIOR, Airto; BERZAGUI, Bruno. Máquinas autônomas e responsabilidade no âmbito do Direito Penal: uma análise a partir da teoria da ação significativa. **Revista Eletrônica Direito e Política**, v. 16, n. 3, p. 1131-1165, 2021. DOI: 10.14210/rdp.v16n3.p1131-1165.

CHESNEY, Robert; CITRON, Danielle Keats. Deep fakes: a looming challenge for privacy, democracy, and national security. **California Law Review**, v. 107, p. 1753-1820, 2019. DOI: 10.15779/Z38RV0D15J.

COALITION FOR CONTENT PROVENANCE AND AUTHENTICITY (C2PA). **Content Credentials Explained: Addressing Common Questions and Updates**. [S.l.]: C2PA Technical Working Group, set. 2025. Disponível em: https://c2pa.org/wp-content/uploads/sites/33/2025/10/content_credentials_wp_0925.pdf. Acesso em: 23 ago. 2026.

COLOMBO, Julia Werlang. Da responsabilidade penal atrelada ao criminal compliance "inteligente": a influência da inteligência artificial na responsabilidade penal individual por crimes omissivos impróprios. **Revista Brasileira de Ciências Criminais**, São Paulo, v. 33, n. 210, p. 121-152, set./out. 2025. DOI: 10.5281/zenodo.15928589.

COMISSÃO EUROPEIA. **AI Act: regulatory framework for artificial intelligence**. Brussels, 2026. Disponível em: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. Acesso em: 22 ago. 2026.

CONSELHO DA EUROPA. **Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No**. 225). Strasbourg, 2024. Disponível em: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence. Acesso em: 22 ago. 2026.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). **Resolução nº 615, de 11 de março de 2025**. Estabelece diretrizes para o desenvolvimento, utilização e governança de soluções desenvolvidas com recursos de inteligência artificial no Poder Judiciário. Brasília, DF: CNJ, 2025.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). **Resolução nº 674, de 25 de março de 2026**. Altera o art. 15 da Resolução CNJ nº 615/2025. Brasília, DF: CNJ, 2026.

CONTROLADORIA-GERAL DA UNIÃO (CGU). **Plano de Integridade e Combate à Corrupção 2025-2027**. Brasília, DF: CGU, 2025. Disponível em: https://www.gov.br/cgu/pt-br/plano-de-integridade-e-combate-a-corrupcao-2025-2027. Acesso em: 23 ago. 2026.

DELFINO, Rebecca. Deepfakes on trial: a call to expand the trial judge's gatekeeping role to protect legal proceedings from technological fakery. **Hastings Law Journal**, v. 74, n. 2, p. 293-348, 2023.

DONEDA, Danilo Cesar Maganhoto; MENDES, Laura Schertel; SOUZA, Carlos Affonso Pereira de; ANDRADE, Norberto Nuno Gomes de. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. **Pensar - Revista de Ciências Jurídicas**, v. 23, n. 4, p. 1-17, 2018. DOI: 10.5020/2317-2150.2018.8257.

FALEIROS JÚNIOR, José Luiz de Moura; OLIVEIRA, Leonardo Augusto de. Inteligência artificial, segurança nacional e devido processo legal: o caso Anthropic v. Department of Defense e suas implicações para o Direito Administrativo brasileiro. **Revista Digital de Direito Administrativo**, v. 13, n. 2, p. 30-52, 2026. DOI: 10.11606/issn.2319-0558.v13n2p30-52.

FLORIDI, Luciano et al. AI4People - An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. **Minds and Machines**, v. 28, p. 689-707, 2018. DOI: 10.1007/s11023-018-9482-5.

FONSECA, Aline Klayse dos Santos. Delineamentos jurídico-dogmáticos da inteligência artificial e seus impactos no instituto da responsabilidade civil. **Civilistica.com**, Rio de Janeiro, v. 10, n. 2, p. 1-36, 2021. Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/671. Acesso em: 23 ago. 2026.

FORNASIER, Mateus de Oliveira. Criminal liability and artificial intelligence: responsabilidade penal e inteligência artificial. **Revista Brasileira de Ciências Criminais**, São Paulo, v. 30, n. 187, p. 153-170, jan. 2022. Disponível em: https://publicacoes.ibccrim.org.br/index.php/RBCCRIM/article/view/63. Acesso em: 23 ago. 2026.

FREDRIKSON, Matt; JHA, Somesh; RISTENPART, Thomas. Model inversion attacks that exploit confidence information and basic countermeasures. In: 22nd ACM SIGSAC CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY, 2015. **Proceedings [...]**. New York: ACM, 2015. p. 1322-1333. DOI: 10.1145/2810103.2813677.

GODOY, Sandro Marcos; LACERDA, Luana Pereira; MOZANER, Victoria Cássia; SANTOS, Gabriel Guerra Miranda Muzeka dos. Explicabilidade como condição de legitimidade algorítmica. **Veredas do Direito**, v. 23, n. 5, e235463, 2026. DOI: 10.18623/rvd.v23.5463.

HOFFMANN-RIEM, Wolfgang. Inteligência artificial como oportunidade para a regulação jurídica. **Direito Público**, v. 16, n. 90, p. 11-38, 2019. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3756. Acesso em: 23 ago. 2026.

INTERPOL. **Global Financial Fraud Threat Assessment 2026**. Lyon: INTERPOL, 2026. Disponível em: https://www.interpol.int/content/download/24291/file/INTERPOL%20Global%20Financial%20Fraud%20Threat%20Assessment%202026.pdf. Acesso em: 23 ago. 2026.

JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. The global landscape of AI ethics guidelines. **Nature Machine Intelligence**, v. 1, p. 389-399, 2019. DOI: 10.1038/s42256-019-0088-2.

LIMA, Taisa Maria Macena de; SÁ, Maria de Fátima Freire de. Inteligência artificial e Lei Geral de Proteção de Dados Pessoais: o direito à explicação nas decisões automatizadas. **Revista Brasileira de Direito Civil - RBDCivil**, Belo Horizonte, v. 26, p. 227-246, out./dez. 2020. DOI: 10.33242/rbdc.2020.04.011.

MAGRANI, Eduardo; MIRANDA, Paulo Rodrigo de. The right to reasonable inferences in automated decision systems as an unfolding of the fundamental right to the protection of personal data in Brazil and beyond. **International Review of Law, Computers & Technology**, v. 39, n. 1, p. 75-94, 2025. DOI: 10.1080/13600869.2024.2351676.

MANTELERO, Alessandro. AI and Big Data: a blueprint for a human rights, social and ethical impact assessment. **Computer Law & Security Review**, v. 34, n. 4, p. 754-772, 2018. DOI: 10.1016/j.clsr.2018.05.017.

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

MENDES, Laura Schertel; FONSECA, Gabriel C. Soares da. Proteção de dados para além do consentimento: tendências contemporâneas de materialização. **Revista Estudos Institucionais**, v. 6, n. 2, p. 507-533, 2020. DOI: 10.21783/rei.v6i2.521.

MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação algorítmica: conceito, fundamento legal e tipologia. **Direito Público**, v. 16, n. 90, p. 39-64, 2019.

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO (MCTI); CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS (CGEE). **IA para o bem de todos: Plano Brasileiro de Inteligência Artificial**. Brasília, DF: MCTI; CGEE, 2025. 104 p. ISBN 978-65-5775-097-1.

MINISTÉRIO DA GESTÃO E DA INOVAÇÃO EM SERVIÇOS PÚBLICOS (MGI). **Estratégia de Inteligência Artificial (E-IA/MGI 2025-2027)**. Brasília, DF: MGI, 2025. Disponível em: https://www.gov.br/gestao/pt-br/central-de-conteudo/publicacoes/planos/e-ia/. Acesso em: 23 ago. 2026.

MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA (MJSP). **Decolar.com é multada por prática de geo pricing e geo blocking**. Brasília, DF, 18 jun. 2018. Disponível em:

https://www.gov.br/mj/pt-br/assuntos/noticias/collective-nitf-content-51. Acesso em: 23 ago. 2026.

MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA (MJSP). **Publicidade de produtos financeiros na internet terá novas medidas de segurança e verificação**. Brasília, DF, 17 jul. 2026. Disponível em: https://www.gov.br/mj/pt-br/assuntos/noticias-1/publicidade-de-produtos-financeiros-na-internet-tera-novas-medidas-de-seguranca-e-verificacao/. Acesso em: 23 ago. 2026.

MITTELSTADT, Brent. Principles alone cannot guarantee ethical AI. **Nature Machine Intelligence**, v. 1, p. 501-507, 2019. DOI: 10.1038/s42256-019-0114-4.

MITTELSTADT, Brent Daniel; ALLO, Patrick; TADDEO, Mariarosaria; WACHTER, Sandra; FLORIDI, Luciano. The ethics of algorithms: mapping the debate. **Big Data & Society**, v. 3, n. 2, 2016. DOI: 10.1177/2053951716679679.

MONTEIRO, Lívia Mariane Couto; CARRÁ, Bruno. Danos digitais: proteção aos direitos da personalidade e funções da responsabilidade civil no ambiente virtual. **Revista IBERC**, Belo Horizonte, v. 9, n. 2, p. 1-23, 2026. DOI: 10.37963/iberc.v9i2.355.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). **Artificial Intelligence Risk Management Framework (AI RMF 1.0)**. Gaithersburg, MD: NIST, 2023. DOI: 10.6028/NIST.AI.100-1.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Dark commercial patterns**. OECD Digital Economy Papers, n. 336. Paris: OECD Publishing, 2022. DOI: 10.1787/44f5e846-en.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Consumer vulnerability in the digital age**. OECD Digital Economy Papers, n. 355. Paris: OECD Publishing, 2023. DOI: 10.1787/4d013cc5-en.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **OECD AI Principles**. Paris: OECD, 2024. Disponível em: https://www.oecd.org/en/topics/ai-principles.html. Acesso em: 22 ago. 2026.

PÁDUA, Sérgio Rodrigo de; LORENZETTO, Bruno Meneses. O direito fundamental à explicabilidade da inteligência artificial utilizada em decisões estatais. **Revista da AGU**, v. 23, n. 2, 2024. DOI: 10.25109/2525-328X.v.23.n.02.2024.3480.

RECHTBANK DEN HAAG. **NJCM c.s**. v. Staat der Nederlanden (SyRI). Processo C-09-550982-HA ZA 18-388. ECLI:NL:RBDHA:2020:865. Julgado em 5 fev. 2020. Haia, Países Baixos, 2020.

RIBEIRO, Ludmilla Rocha Cunha. **Inteligência artificial e decisão administrativa: integração tecnológica e o princípio da motivação**. 2024. Dissertação (Mestrado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2024. DOI: 10.11606/D.2.2024.tde-09122024-172640.

SALLES, Raquel Bellini; COSTA, Thais Silva. A securitização dos danos causados por inteligência artificial. **Civilistica.com**, v. 12, n. 1, p. 1-32, 2023. DOI: 10.5281/zenodo.18881744.

SECRETARIA DE GOVERNO DIGITAL (SGD). **GuIA Unificado de Inteligência Artificial para o Setor Público**. Brasília, DF: Ministério da Gestão e da Inovação em Serviços Públicos, [s.d.]. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-

dados/inteligencia-artificial-1/publicacoes/guia-unificado-de-inteligencia-artificial-para-o-setor-publico-1/guia-unificado-de-inteligencia-artificial-para-o-setor-publico. Acesso em: 23 ago. 2026.

SECRETARIA DE GOVERNO DIGITAL (SGD); SERVIÇO FEDERAL DE PROCESSAMENTO DE DADOS (SERPRO). **Guia IA Generativa no Serviço Público**. Brasília, DF: Ministério da Gestão e da Inovação em Serviços Públicos, [s.d.]. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/inteligencia-artificial-1/publicacoes/guia-ia-generativa. Acesso em: 23 ago. 2026.

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

SHOKRI, Reza; STRONATI, Marco; SONG, Congzheng; SHMATIKOV, Vitaly. Membership inference attacks against machine learning models. In: 2017 IEEE SYMPOSIUM ON SECURITY AND PRIVACY, 2017. **Proceedings [...]**. Los Alamitos: IEEE, 2017. p. 3-18. DOI: 10.1109/SP.2017.41.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Súmula 403**. Independe de prova do prejuízo a indenização pela publicação não autorizada de imagem de pessoa com fins econômicos ou comerciais. Segunda Seção, julgado em 28 out. 2009, DJe 24 nov. 2009. Brasília, DF: STJ, 2009.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Súmula 479**. As instituições financeiras respondem objetivamente pelos danos gerados por fortuito interno relativo a fraudes e delitos praticados por terceiros no âmbito de operações bancárias. Segunda Seção, julgado em 27 jun. 2012, DJe 1 ago. 2012. Brasília, DF: STJ, 2012.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Recurso Especial nº 1.419.697/RS**. Segunda Seção. Rel. Min. Paulo de Tarso Sanseverino. Recurso repetitivo - Tema 710. Julgado em 12 nov. 2014. DJe 17 nov. 2014. Brasília, DF: STJ, 2014.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Súmula 550**. A utilização de escore de crédito dispensa o consentimento do consumidor, assegurado o direito a esclarecimentos sobre informações pessoais valoradas e fontes dos dados considerados no cálculo. Segunda Seção, julgado em 14 out. 2015. DJe 19 out. 2015. Brasília, DF: STJ, 2015.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Recurso Especial nº 1.630.851/SP**. Terceira Turma. Rel. Min. Paulo de Tarso Sanseverino. Julgado em 27 abr. 2017. DJe 22 jun. 2017. Brasília, DF: STJ, 2017.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Agravo Regimental no Habeas Corpus nº 828.054/RN**. Quinta Turma. Rel. Min. Joel Ilan Paciornik. Julgado em 23 abr. 2024. DJe 29 abr. 2024. Brasília, DF: STJ, 2024.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Recurso Especial nº 2.222.059/SP**. Terceira Turma. Rel. Min. Ricardo Villas Bôas Cueva. Julgado em 7 out. 2025. DJEN 13 out. 2025. Brasília, DF: STJ, 2025.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Recurso Especial nº 2.164.130/SP**. Terceira Turma. Rel. Min. Moura Ribeiro. Julgado em 9 fev. 2026. DJEN 13 fev. 2026. Brasília, DF: STJ, 2026a.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Agravo Regimental no Habeas Corpus nº 1.014.212/ES**. Sexta Turma. Rel. Min. Carlos Pires Brandão. Julgado em 10 fev. 2026. DJEN 20 fev. 2026. Brasília, DF: STJ, 2026b.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Inquérito nº 1.674/DF**. Corte Especial. Rel. Min. Nancy Andrighi. Julgado em 6 maio 2026. DJEN 14 maio 2026. Brasília, DF: STJ, 2026c.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Agravo em Recurso Especial nº 3.132.918/RS**. Quinta Turma. Rel. Min. Ribeiro Dantas. Julgado em 9 jun. 2026. DJEN 15 jun. 2026. Brasília, DF: STJ, 2026d.

SUPREMO TRIBUNAL FEDERAL (STF). **Recurso Extraordinário nº 548.181/PR**. Primeira Turma. Rel. Min. Rosa Weber. Julgado em 6 ago. 2013. DJe 30 out. 2014. Brasília, DF: STF, 2013.

SUPREMO TRIBUNAL FEDERAL (STF). **Ação Direta de Inconstitucionalidade nº 4.815/DF**. Rel. Min. Cármen Lúcia. Tribunal Pleno. Julgado em 10 jun. 2015. Acórdão publicado em 1 fev. 2016. Brasília, DF: STF, 2015.

SUPREMO TRIBUNAL FEDERAL (STF). **Temas 533 e 987 da repercussão geral: responsabilidade civil de provedores de aplicações de internet por conteúdos gerados por terceiros; interpretação do art**. 19 do Marco Civil da Internet. RE 1.057.258/MG e RE 1.037.396/SP. Julgamento conjunto de mérito concluído em 26 jun. 2025; tese do Tema 987 ajustada no julgamento dos embargos de declaração concluído em 17 jun. 2026. Brasília, DF: STF, 2026.

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

TRIBUNAL DE JUSTIÇA DA UNIÃO EUROPEIA (TJUE). **Acórdão de 7 de dezembro de 2023, processo C-634/21, SCHUFA Holding (Scoring), ECLI:EU:C:2023:957**. Luxemburgo: TJUE, 2023.

TRIBUNAL DE JUSTIÇA DA UNIÃO EUROPEIA (TJUE). **Acórdão de 27 de fevereiro de 2025, processo C-203/22, Dun & Bradstreet Austria, ECLI:EU:C:2025:117**. Luxemburgo: TJUE, 2025.

TRIBUNAL SUPERIOR ELEITORAL (TSE). **Resolução nº 23.610, de 18 de dezembro de 2019**. Dispõe sobre propaganda eleitoral. Texto consolidado com as alterações das Resoluções nº 23.732/2024 e nº 23.755/2026. Brasília, DF: TSE, 2026. Disponível em: https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019. Acesso em: 23 ago. 2026.

UNESCO. **Recommendation on the Ethics of Artificial Intelligence**. Paris: UNESCO, 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000380455. Acesso em: 22 ago. 2026.

UNIÃO EUROPEIA. **Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial (Regulamento da Inteligência Artificial)**. Jornal Oficial da União Europeia, 12 jul. 2024a.

UNIÃO EUROPEIA. **Diretiva (UE) 2024/2853 do Parlamento Europeu e do Conselho, de 23 de outubro de 2024, relativa à responsabilidade decorrente de produtos defeituosos e que revoga a Diretiva 85/374/CEE**. Jornal Oficial da União Europeia, 18 nov. 2024b.

UNIÃO EUROPEIA. **Regulamento (UE) 2026/1744 do Parlamento Europeu e do Conselho, de 8 de julho de 2026, que altera os Regulamentos (UE) 2024/1689, (UE) 2018/1139 e (UE) 2023/1230 quanto à simplificação da aplicação das regras harmonizadas sobre inteligência artificial**. Jornal Oficial da União Europeia, 24 jul. 2026.

UNITED NATIONS OFFICE ON DRUGS AND CRIME (UNODC). **Emerging threats: the intersection of criminal and technological innovation in the use of automation and AI**. Bangkok: UNODC Regional Office for Southeast Asia and the Pacific, 2025.

VACCARI, Cristian; CHADWICK, Andrew. Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. **Social Media + Society**, v. 6, n. 1, 2020. DOI: 10.1177/2056305120903408.

WACHTER, Sandra; MITTELSTADT, Brent. A right to reasonable inferences: re-thinking data protection law in the age of Big Data and AI. **Columbia Business Law Review**, v. 2019, n. 2, p. 494-620, 2019. DOI: 10.7916/cblr.v2019i2.3424.

SOBRE OS AUTORES

Dyego Girão Anjos

Advogado e professor universitário, mestrando em Estudos Jurídicos com ênfase em Direito Internacional pela Must University. Possui formação continuada em diferentes áreas do Direito e atua profissionalmente no Barros Advogados Associados e na FACCON. Seus interesses acadêmicos abrangem Direito Público, Tributário, Administrativo, Constitucional, Direitos Humanos, Direito Eleitoral e Teoria Geral do Direito.

Karla Lôbo Fagundes

Graduada em Direito, especialista em Direito Constitucional e Administrativo e em Direito e Processo do Trabalho, além de mestranda em Estudos Jurídicos com ênfase em Direito Internacional pela Must University. É servidora pública estadual do Tribunal de Justiça de Goiás, onde atua como Assistente de Juiz de Direito, e advogada licenciada. Seus interesses acadêmicos envolvem Direito Internacional, direitos fundamentais, governança, tecnologia e desafios jurídicos contemporâneos.

Alan David Souza Salustiano da Silva

Educador e articulador sociocultural, licenciado em Pedagogia pelo Instituto Educacional Aliança do Maranhão. Atua em consultoria vinculada ao Instituto Terra Solidária, com experiência em gestão pedagógica, formação educacional e desenvolvimento de projetos culturais e sociais. Sua trajetória articula educação, cultura e políticas públicas, incluindo docência multidisciplinar, coordenação pedagógica, economia criativa e práticas educativas inclusivas.

Madison Lopes Inocencio

Graduado em Administração pelo Centro Universitário São José (UniSãoJosé) e especialista em Gestão e Desenvolvimento de Pessoas nas Organizações pela Faculdade Venda Nova do Imigrante (FAVENI). Atua como Administrador na Universidade Federal do Triângulo Mineiro (UFTM), desenvolvendo atividades relacionadas à gestão e aos processos administrativos no âmbito de uma instituição pública federal. Sua formação e experiência profissional articulam Administração, Gestão de Pessoas e gestão institucional.

DIREITO NA ERA DA INTELIGÊNCIA ARTIFICIAL

Sistemas capazes de classificar pessoas, produzir textos, imagens e vozes sintéticas, apoiar decisões e operar sobre grandes volumes de dados passaram a interferir em direitos da personalidade, relações de consumo, atuação estatal, processos judiciais e estruturas de responsabilidade.

A obra parte de uma premissa central: a novidade tecnológica não suspende o Direito, mas altera as formas pelas quais direitos podem ser atingidos, provas podem ser produzidas, danos podem se propagar e decisões podem ser tomadas. Os capítulos articulam Constituição, proteção de dados, responsabilidade civil e penal, direito do consumidor, prova digital, transparência algorítmica, uso de IA pelo poder público e os marcos regulatórios em construção.

Destinada a estudantes, pesquisadores, profissionais do Direito, gestores públicos e leitores interessados nas transformações digitais, a obra propõe uma leitura crítica e aplicada da relação entre inteligência artificial e ordem jurídica.

AUTORIA

Dyego Girão Anjos
Karla Lôbo Fagundes
Alan David Souza Salustiano da Silva
Madison Lopes Inocencio

ORGANIZAÇÃO

Kátia Karolini Amaro El Alam

EDITORA MANUAL

ISBN: 978-6-58312-130-1

9 786583 121301