UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE DIREITO “PROF. JACY DE ASSIS”

BIANCA THAÍS SOUSA OLIVEIRA

ERROS DE SISTEMAS DE IA E A RESPONSABILIDADE DAS EMPRESAS DESENVOLVEDORAS NO ORDENAMENTO JURÍDICO BRASILEIRO

UBERLÂNDIA
2026

BIANCA THAÍS SOUSA OLIVEIRA

ERROS DE SISTEMAS DE IA E A RESPONSABILIDADE DAS EMPRESAS DESENVOLVEDORAS NO ORDENAMENTO JURÍDICO BRASILEIRO

Trabalho de Conclusão de Curso apresentado à Faculdade de Direito da Universidade Federal de Uberlândia como requisito parcial para a obtenção do título de bacharel em Direito.

Orientador: Prof. Dr. Almir Garcia Fernandes

UBERLÂNDIA

2026

BIANCA THAÍS SOUSA OLIVEIRA

ERROS DE SISTEMAS DE IA E A RESPONSABILIDADE DAS EMPRESAS DESENVOLVEDORAS NO ORDENAMENTO JURÍDICO BRASILEIRO

Trabalho de Conclusão de Curso apresentado à Faculdade de Direito da Universidade Federal de Uberlândia como requisito parcial para a obtenção do título de bacharel em Direito.

Uberlândia, 17 de julho de 2026.

Banca examinadora:

Prof. Dr. Almir Garcia Fernandes – Orientador

Universidade Federal de Uberlândia – UFU

Prof. Dr. Luiz César Machado de Macedo

Universidade Federal de Uberlândia – UFU

Dedico este trabalho aos meus quatro avós,
que não tiveram a oportunidade de concluir o
ensino fundamental, mas hoje veem a
“menininha” que criaram, sendo a primeira
neta a se formar.
Aos amigos que foram abrigo no meu
caminho: Cleo, Fernanda, Grazielle, Maciela,
Milena e Renato.
E ao Thiago, que chegou no fim da caminhada,
mas segurou a minha mão como se tivesse
estado comigo desde o começo desta jornada.

RESUMO

O presente trabalho tem como objetivo analisar a responsabilidade civil das empresas desenvolvedoras de sistemas de inteligência artificial generativa diante de erros, falhas e danos causados a usuários e terceiros no ordenamento jurídico brasileiro. Este trabalho parte da constatação de que a evolução dos meios de comunicação, da internet e dos algoritmos modificou profundamente a forma de produção e circulação da informação, culminando no surgimento de sistemas capazes de criar textos, imagens, áudios e conteúdos com aparência de veracidade. Nesse cenário, a opacidade algorítmica, a dificuldade de identificação do nexo causal e a multiplicidade de agentes envolvidos na criação, disponibilização e uso dessas tecnologias desafiam as categorias tradicionais da responsabilidade civil. O estudo examina a aplicação da responsabilidade objetiva, da teoria do risco-proveito, do Marco Civil da Internet, da Lei Geral de Proteção de Dados e da recente interpretação do Supremo Tribunal Federal sobre os Temas 987 e 533. Também são analisados casos paradigmáticos envolvendo sistemas como Grok, Microsoft Tay e Chai App "Eliza", além dos riscos da inteligência artificial à integridade democrática, especialmente no contexto eleitoral. Por fim, são discutidos mecanismos de contenção e regulação, com destaque para o Projeto de Lei n.º 2.338/2023, o Projeto de Lei n.º 2.630/2020 e a *accountability* algorítmica. Conclui-se que o Direito brasileiro já dispõe de instrumentos relevantes para a responsabilização e prevenção de danos, mas ainda necessita de aperfeiçoamento normativo específico para enfrentar os riscos próprios da inteligência artificial generativa, garantindo transparência, governança, reparação adequada e proteção à dignidade da pessoa humana.

Palavras-chave: inteligência artificial generativa; responsabilidade civil; opacidade algorítmica; direito digital; *accountability*.

ABSTRACT

This paper aims to analyze the civil liability of companies that develop generative artificial intelligence systems in cases of errors, failures, and damages caused to users and third parties under the Brazilian legal system. The research is based on the understanding that the evolution of the media, the internet, and algorithms has profoundly changed the way information is produced and disseminated, leading to the emergence of systems capable of creating texts, images, audio, and synthetic content with an appearance of truthfulness. In this context, algorithmic opacity, the difficulty of identifying causation, and the multiplicity of agents involved in the creation, provision, and use of these technologies challenge traditional categories of civil liability. The study examines the application of strict liability, the risk-profit theory, the Brazilian Civil Rights Framework for the Internet, the General Data Protection Law, and the recent interpretation of the Federal Supreme Court on the Themes 987 and 533. It also analyzes paradigmatic cases involving systems such as Grok, Microsoft Tay, and Chai App "Eliza", as well as the risks posed by artificial intelligence to democratic integrity, especially in the electoral context. Finally, the paper discusses containment and regulatory mechanisms, with emphasis on Bill No. 2,338/2023, Bill No. 2,630/2020, and algorithmic accountability. The research concludes that Brazilian law already provides relevant instruments for liability and damage prevention, but still requires specific regulatory improvement to address the particular risks of generative artificial intelligence, ensuring transparency, governance, adequate compensation, and protection of human dignity.

Keywords: generative artificial intelligence; civil liability; algorithmic opacity; digital law; accountability.

SUMÁRIO

1 INTRODUÇÃO

A expansão das tecnologias digitais modificou profundamente as formas de produção, circulação e consumo da informação. Se, anteriormente, a divulgação de fatos dependia predominantemente da atuação de profissionais e instituições responsáveis pela seleção e verificação do conteúdo, o desenvolvimento da inteligência artificial (IA) generativa possibilitou a criação automatizada de textos, imagens, áudios, vídeos e outros materiais com elevado grau de verossimilhança. Ferramentas como o ChatGPT, o Grok, o Gemini e outros sistemas baseados em inteligência artificial passaram a integrar atividades profissionais, acadêmicas, comerciais e sociais, produzindo impactos cada vez mais relevantes sobre as relações jurídicas[1].

Apesar dos benefícios relacionados à rapidez, à automação e à ampliação do acesso à informação, os sistemas de inteligência artificial generativa também podem produzir resultados incorretos, discriminatórios ou lesivos. Entre os principais riscos associados à utilização dessas tecnologias encontram-se a criação de informações falsas, a reprodução de preconceitos presentes nos dados utilizados para o treinamento dos modelos, a violação de direitos da personalidade, a utilização indevida de dados pessoais e a geração de conteúdos capazes de afetar a honra, a imagem, a privacidade e a segurança dos indivíduos.

A ocorrência desses danos apresenta dificuldades específicas para o Direito, sobretudo porque os sistemas de inteligência artificial funcionam mediante estruturas técnicas complexas, frequentemente caracterizadas pela denominada opacidade algorítmica. Em determinadas situações, nem mesmo os responsáveis pelo desenvolvimento da tecnologia conseguem explicar de maneira completa o percurso realizado pelo sistema até a produção de determinado resultado. Essa dificuldade de compreensão e rastreamento pode comprometer a identificação do nexo de causalidade entre o funcionamento do sistema e o dano experimentado pelo usuário.

Nesse contexto, o presente trabalho delimita-se à análise da responsabilidade civil das empresas desenvolvedoras de sistemas de inteligência artificial generativa no ordenamento jurídico brasileiro. Este trabalho busca examinar de que forma as normas atualmente existentes podem ser aplicadas aos danos provocados por erros ou falhas desses sistemas, considerando especialmente as dificuldades relacionadas à autonomia tecnológica, à opacidade algorítmica e à identificação dos agentes responsáveis ao longo da cadeia de desenvolvimento e fornecimento

1 Levantamento da consultoria beAnalytic aponta Microsoft, Google/Alphabet, OpenAI, NVIDIA e Meta como líderes do setor de inteligência artificial em 2026, concentrando juntas mais de 60% dos investimentos globais na área, o que evidencia a concentração econômica e tecnológica em torno do desenvolvimento, infraestrutura e disponibilização de sistemas de IA.

da tecnologia. Diante dessa delimitação, estabelece-se o seguinte problema de pesquisa: de que maneira o Direito Civil brasileiro pode responsabilizar as empresas desenvolvedoras pelos danos decorrentes de erros de sistemas de inteligência artificial generativa, especialmente quando a opacidade algorítmica dificulta a demonstração do nexo de causalidade e da culpa?

A relevância do tema decorre da crescente inserção da inteligência artificial nas relações sociais e econômicas e da possibilidade de que os danos provocados por esses sistemas atinjam um número expressivo de pessoas[2]. A complexidade técnica da tecnologia não pode, por si só, impedir o acesso das vítimas à reparação, tampouco servir como justificativa para a ausência de definição acerca das responsabilidades dos agentes que desenvolvem, fornecem ou exploram economicamente esses sistemas.

O trabalho também se justifica pela necessidade de verificar se os instrumentos jurídicos atualmente existentes são suficientes para enfrentar os conflitos decorrentes da inteligência artificial. O Código Civil, o Marco Civil da Internet e a Lei Geral de Proteção de Dados Pessoais apresentam normas que podem incidir sobre essas situações. Paralelamente, as discussões legislativas relacionadas à regulamentação da inteligência artificial, o Projeto de Lei (PL) n.º 2.338/2023 demonstra a necessidade de estabelecer deveres de transparência, prevenção, segurança, supervisão e reparação de danos. Dessa forma, o estudo mostra-se relevante tanto para a compreensão teórica da responsabilidade civil quanto para a busca de respostas jurídicas capazes de proteger os direitos fundamentais e a dignidade da pessoa humana diante das novas tecnologias.

O objetivo geral do trabalho consiste em analisar o regime de responsabilidade civil aplicável às empresas desenvolvedoras de sistemas de inteligência artificial generativa diante dos erros e danos causados aos usuários, considerando o ordenamento jurídico brasileiro e as dificuldades decorrentes da opacidade algorítmica. Como objetivos específicos, pretende-se examinar as principais falhas éticas e técnicas relacionadas a esses sistemas; analisar o nexo de causalidade entre a atuação da inteligência artificial e os danos produzidos; investigar a possibilidade de aplicação da responsabilidade civil objetiva às empresas desenvolvedoras; avaliar a suficiência das normas brasileiras para o enfrentamento dessas situações; e identificar mecanismos jurídicos de responsabilização e prevenção de danos.

Para alcançar esses objetivos, será utilizado o método dedutivo, partindo-se das teorias gerais da responsabilidade civil para a análise de sua aplicação aos danos decorrentes da

---

2 Levantamento da Value Capital Advisors, feito a pedido da Forbes Tech (2026), aponta que o número de *startups* brasileiras ativas no setor de Inteligência Artificial saltou de 352, em 2016, para 975 em 2025, um crescimento de quase 40% apenas na última meia década, com o Sudeste concentrando 71,18% das operações e o estado de São Paulo detendo isoladamente 56% do *market share* nacional.

inteligência artificial generativa. O trabalho terá abordagem qualitativa e será desenvolvido por meio de levantamento bibliográfico e documental, com o exame da doutrina especializada em Direito Civil, Direito Digital, bem como da legislação, de projetos legislativos, de documentos institucionais e da jurisprudência relacionada ao tema. Também serão analisados casos paradigmáticos envolvendo falhas de sistemas de inteligência artificial, com a finalidade de identificar os principais problemas jurídicos decorrentes da utilização dessas tecnologias.

O trabalho será organizado em quatro capítulos de desenvolvimento. O primeiro abordará a evolução das formas de disseminação da informação, desde os meios de comunicação tradicionais até o surgimento da inteligência artificial generativa. O segundo examinará os fundamentos da responsabilidade civil e sua aplicação ao ambiente digital, com atenção à responsabilidade objetiva, à teoria do risco e ao desafio da demonstração do nexo de causalidade em sistemas caracterizados pela opacidade algorítmica. O terceiro apresentará casos concretos e paradigmáticos envolvendo falhas de inteligência artificial e danos aos indivíduos e à sociedade. Por fim, o quarto capítulo analisará os mecanismos de contenção, prevenção e regulamentação da inteligência artificial no Brasil, buscando compreender como o avanço tecnológico pode ser conciliado com a proteção dos direitos fundamentais e com a efetiva reparação das vítimas.

2 A EVOLUÇÃO DA DISSEMINAÇÃO DA INFORMAÇÃO: DO JORNALISMO À INTELIGÊNCIA ARTIFICIAL

A disseminação da informação sempre exerceu papel central na organização da vida coletiva. A formação da opinião pública, o controle social dos atos estatais, a participação política e o próprio funcionamento das instituições jurídicas dependem da possibilidade de acesso a informações minimamente confiáveis. No Estado Democrático de Direito, a liberdade de manifestação, o acesso à informação e a liberdade de comunicação não constituem apenas interesses individuais, mas condições para que os cidadãos possam fiscalizar o poder, participar do debate público e tomar decisões conscientes. A Constituição Federal de 1988 reconhece essa importância ao proteger a liberdade de expressão e de informação, o direito de resposta e a inviolabilidade da honra, da imagem, da intimidade e da vida privada (Brasil, 1988).

Os meios pelos quais a informação é produzida e distribuída, contudo, não permaneceram estáticos. Da comunicação pública realizada em suportes físicos à imprensa de massa, da internet participativa às plataformas digitais e, mais recentemente, aos sistemas de inteligência artificial generativa, cada transformação tecnológica alterou não somente a velocidade e o alcance das informações, mas também os agentes responsáveis por selecionar, produzir e validar os conteúdos. Compreender essa trajetória é indispensável para examinar o problema central deste trabalho: a atribuição de responsabilidade quando informações ou conteúdos gerados por sistemas de inteligência artificial provocam danos.

Este capítulo apresenta essa evolução em três momentos. Inicialmente, analisa-se a mediação humana característica dos meios de comunicação tradicionais, marcada pela existência de editores, jornalistas e organizações identificáveis. Em seguida, examina-se a transição para o ambiente digital, no qual a participação dos usuários e a personalização algorítmica modificaram a circulação da informação. Por fim, aborda-se o advento da inteligência artificial generativa e a substituição parcial do filtro editorial humano por modelos probabilísticos capazes de produzir conteúdos sintéticos, preparando o terreno para a análise da responsabilidade civil desenvolvida no capítulo seguinte.

2.1 A mediação humana e a era dos meios de comunicação de massa

A relação entre informação e democracia pressupõe não apenas a existência de canais de comunicação, mas também a possibilidade de identificar quem produziu, selecionou e divulgou determinada mensagem. Antes da internet, a circulação pública de informações era predominantemente mediada por instituições e profissionais que ocupavam posições definidas na cadeia comunicacional. Jornais impressos, emissoras de rádio e televisão decidiam quais fatos mereciam publicidade, como seriam apresentados e quais fontes seriam consideradas

confiáveis. Embora esse sistema não fosse neutro nem imune a interesses políticos e econômicos, havia uma estrutura editorial visível, formada por repórteres, revisores, editores, diretores e empresas de comunicação.

Entre os registros históricos mais conhecidos de divulgação organizada de notícias encontra-se a Acta Diurna, instituída em Roma no século I a.C. para comunicar atos públicos, acontecimentos políticos e decisões governamentais. Apesar de seu caráter oficial e de sua distância em relação aos critérios contemporâneos de jornalismo, a experiência evidencia uma preocupação antiga com a circulação sistemática de fatos de interesse coletivo. Séculos depois, a prensa de tipos móveis associada a Johannes Gutenberg ampliou significativamente a capacidade de reprodução de textos, reduziu os custos da circulação escrita e contribuiu para a formação de uma esfera pública mais ampla. A multiplicação de livros, panfletos e periódicos transformou o conhecimento em conteúdo reproduzível em larga escala e forneceu bases materiais para o desenvolvimento da imprensa moderna (Eisenstein, 2013).

Nos séculos XIX e XX, o crescimento dos jornais de grande circulação, seguido pela expansão do rádio e da televisão, consolidou o modelo de comunicação de massa. A mensagem era produzida por um número relativamente restrito de organizações e transmitida simultaneamente a grandes públicos. Tratava-se, em regra, de uma comunicação vertical e predominantemente unilateral: poucos emissores falavam para muitos receptores, enquanto o público possuía possibilidades limitadas de resposta imediata ou de produção autônoma de conteúdos com alcance maior.

Nesse contexto desenvolveu-se o conceito de *gatekeeping*[3], empregado para descrever o processo pelo qual determinados profissionais controlavam os "portões" de entrada da informação no espaço público. No estudo clássico de White[4] (1950 *apud* Di Lauro; Baracuhy, 2025, p. 108), publicado em 1950, a seleção das notícias foi analisada a partir da atuação do *gatekeeper*, isto é, do agente responsável por controlar quais informações seriam transformadas em notícia. A teoria evidencia que a informação jornalística não chegava ao público de forma neutra ou integral, mas passava previamente por escolhas humanas e critérios editoriais de seleção.

---

3 No modelo clássico de *gatekeeping*, próprio dos meios de comunicação de massa, a seleção das notícias era condicionada pelas rotinas de produção, distribuição e consumo existentes no jornalismo impresso, radiofônico e televisivo. Como esses veículos não conseguiam divulgar todos os acontecimentos do dia, era necessário selecionar quais fatos seriam considerados mais relevantes para a audiência e quais conteúdos poderiam ocupar o espaço disponível nas publicações ou transmissões (Bruns, 2011, p. 118-119, tradução nossa).

4 WHITE, David Manning. The "gatekeeper": a case study in the selection of news. **Sage Journals**, v. 27, n. 4, p. 383-390, 1950. Disponível em: https://journals.sagepub.com/doi/10.1177/107769905002700403. Acesso em: 14 ago. 2026.

O *gatekeeping* humano desempenhava, portanto, uma dupla função. De um lado, restringia a pluralidade ao concentrar o poder de selecionar os acontecimentos considerados noticiáveis. De outro, criava instâncias profissionais de checagem, revisão e responsabilização. Antes da publicação, o conteúdo poderia ser confrontado com documentos, fontes e padrões éticos da atividade jornalística. Depois da divulgação, a existência de um veículo, de uma direção editorial e de profissionais identificáveis permitia localizar com maior precisão a origem do conteúdo e os sujeitos potencialmente responsáveis por eventual ilícito.

Essa estrutura possuía consequências jurídicas relevantes. Quando uma reportagem violava a honra, a imagem ou a privacidade de alguém, era possível direcionar o pedido de correção, direito de resposta ou reparação ao autor, ao editor ou à empresa de comunicação envolvida. A Constituição Federal assegura o direito de resposta proporcional ao agravo e a indenização por dano material, moral ou à imagem, enquanto o Código Civil estabelece o dever de reparar o dano decorrente de ato ilícito. Isso não significa que a responsabilização fosse simples em todos os casos, em que o jornalismo estivesse livre de abusos. Significa apenas que a cadeia de produção e divulgação era, em geral, mais delimitada e reconhecível.

O filtro editorial também estava relacionado à construção da chamada verdade jornalística, que não se confunde com uma verdade absoluta, mas resulta de procedimentos profissionais de apuração, confronto de versões, identificação de fontes e revisão. No campo jurídico, a confiança das informações é igualmente relevante porque a aplicação do Direito depende da reconstrução de fatos. Documentos, testemunhos, perícias e notícias podem influenciar investigações, decisões administrativas e processos judiciais. Quanto mais clara a origem de uma informação, maiores são as possibilidades de verificar sua autenticidade, contestá-la e atribuir responsabilidade por sua produção.

A mediação humana, contudo, não deve ser idealizada. Editores podem agir com parcialidade, veículos podem reproduzir preconceitos e estruturas empresariais podem privilegiar determinados interesses. Ainda assim, o sistema tradicional mantinha um elemento que se tornaria progressivamente mais difuso no ambiente digital: a presença de pessoas e organizações que exerciam controle prévio sobre a mensagem e assumiram publicamente a autoria editorial de sua divulgação.

A passagem para a internet modificou esse equilíbrio. O controle exercido por poucos veículos foi reduzido pela possibilidade de publicação direta por usuários, organizações e comunidades, especialmente a partir da consolidação da chamada Web 2.0, caracterizada por uma arquitetura de participação na qual os usuários deixam de ser apenas receptores e passam a produzir, compartilhar e reorganizar conteúdos no ambiente digital (O’Reilly, 2005). Nesse

mesmo sentido, Bruns (2003) observa que o ambiente online desloca o modelo clássico de *gatekeeping*, centrado na seleção editorial prévia das notícias, para práticas de *gatewatching*, nas quais a circulação informacional passa a depender também da observação, seleção, indicação e redistribuição colaborativa de conteúdos por diferentes atores conectados em rede. Assim, ao mesmo tempo em que essa abertura democratizou a expressão, ela enfraqueceu a centralidade do filtro profissional e multiplicou os pontos de produção de conteúdo, tornando autoria, alcance e responsabilidade mais difíceis de delimitar.

2.2 A transição para o digital e a ascensão das IAS generativas (LLMS)

A popularização da internet, especialmente a partir da década de 1990, representou uma ruptura com o modelo centralizado dos meios de comunicação de massa. Castells (2003) compreende a internet como uma tecnologia de comunicação própria da sociedade em rede, capaz de reorganizar fluxos informacionais, relações sociais e formas de interação no contexto da era da informação. Em sua fase inicial, posteriormente identificada como Web 1.0, grande parte das páginas funcionava de modo predominantemente estático, com menor participação direta dos usuários na produção e reorganização dos conteúdos. Nesse modelo, o usuário comum ocupava principalmente a posição de leitor ou navegador, ao passo que a Web 2.0 passou a ser associada a maior interatividade, comunicação bidirecional, redes sociais e produção colaborativa de conteúdo (Cormode; Krishnamurthy, 2008; O'Reilly, 2005).

O desenvolvimento posterior da chamada Web 2.0 alterou esse cenário ao transformar a rede em plataforma de participação. Blogs, fóruns, redes sociais, serviços de compartilhamento e sistemas colaborativos permitiram que usuários não apenas recebessem, mas também produzissem, comentassem, modificassem e distribuíssem conteúdos. O'Reilly (2005, tradução nossa)[5] descreveu essa transformação a partir da ideia de uma "arquitetura de participação", expressão utilizada pelo autor para demonstrar que, na Web 2.0, determinados serviços digitais se aperfeiçoam à medida que mais pessoas os utilizam e contribuem para seu funcionamento.

A ampliação da participação teve efeitos democráticos importantes. Indivíduos e grupos anteriormente afastados dos grandes meios de comunicação passaram a divulgar experiências, denúncias, opiniões e conhecimentos sem depender da autorização de um editor ou da infraestrutura de uma emissora. A produção informacional tornou-se mais plural e os acontecimentos locais puderam alcançar repercussão nacional ou internacional em poucos

5 No original: "the service automatically gets better the more people use it" e "architecture of participation" (O'Reilly, 2005, p. 2).

minutos. Entretanto, a mesma abertura que reduziu o monopólio dos veículos tradicionais também eliminou parte das barreiras editoriais que filtravam conteúdos falsos, ofensivos ou manipulados.

Essa transformação digital, no entanto, não se limita à mudança dos instrumentos de comunicação. Na perspectiva de Barcarollo (2021), a inteligência artificial não representa apenas uma inovação tecnológica, mas um fenômeno capaz de afetar a organização social e jurídica contemporânea. Por isso, o autor defende a necessidade de repensar categorias do Direito diante da transformação digital, da sociedade em rede e dos novos fluxos informacionais produzidos pela interação entre tecnologia, instituições e sujeitos.

O crescimento exponencial da quantidade de dados disponíveis produziu uma nova dificuldade: a sobrecarga informacional. Bawden e Robinson (2009) observam que o excesso de informação constitui uma das patologias do ambiente informacional contemporâneo, pois a ampliação do volume de dados disponíveis pode superar a capacidade humana de seleção, processamento e uso significativo da informação. Diante de um volume de publicações muito superior à capacidade humana de leitura, as plataformas passaram a utilizar algoritmos para classificar, selecionar e recomendar conteúdos. Em vez de uma edição igual para todos os leitores, cada usuário passou a receber uma sequência personalizada de informações, definida a partir de seu histórico de navegação, interações, localização, preferências e padrões de comportamento.

Na década de 2010, essa lógica consolidou-se nas redes sociais e nos mecanismos de busca. Di Lauro e Baracuhy (2025) observam que, nas redes sociais, os algoritmos atuam por meio de sistemas de recomendação e bases de dados, personalizando o conteúdo exibido aos usuários e criando filtros de informação conforme suas preferências e rastros digitais. Conforme explicam as autoras, o controle da visibilidade deixou de depender exclusivamente da decisão de um editor humano e passou a ser compartilhado com sistemas automatizados que ordenam feeds, selecionam notícias e recomendam conteúdos a partir de dados como interações, compartilhamentos, padrões de consumo e tempo de permanência diante das publicações. O estudo conclui que a informação, nesse ambiente, não é apenas disponibilizada: ela é selecionada, organizada e distribuída de maneira distinta para cada indivíduo, conforme lógicas algorítmicas de relevância e engajamento.

A personalização pode facilitar o acesso a conteúdos relevantes, mas também tende a limitar o contato com perspectivas divergentes. O conceito de “bolha de filtros”, desenvolvido por Pariser (2011 *apud* Di Lauro; Baracuhy, 2025), descreve situações em que sistemas de recomendação reforçam interesses e opiniões previamente demonstrados pelo usuário, criando

ambientes informacionais personalizados e potencialmente fechados à diversidade de pontos de vista. Essa dinâmica favorece a rápida disseminação de conteúdos emocionais e sensacionalistas, muitas vezes mais aptos a gerar interação do que informações complexas ou verificadas.

A substituição parcial do *gatekeeping* editorial pelo *gatekeeping* algorítmico alterou a natureza da mediação. Nos meios tradicionais, os critérios de seleção podiam ser questionados perante jornalistas, editores e empresas identificáveis. Nas plataformas digitais, parte significativa das decisões de visibilidade decorre de modelos matemáticos, segredos empresariais e sistemas atualizados. O usuário costuma perceber apenas o resultado final, a postagem recomendada, o vídeo impulsionado ou a notícia destacada, sem conhecer os critérios que determinaram sua exibição.

Nesse ponto, a discussão sobre a mediação algorítmica aproxima-se da ideia de opacidade dos sistemas digitais. Di Lauro e Baracuhy (2025) observam que, nas redes sociais, os algoritmos organizam e filtram dados, controlando fluxos de informação, mapeando preferências e definindo quais conteúdos aparecem aos usuários. Essa atuação evidencia que a circulação informacional nas plataformas digitais não é neutra, pois passa a ser condicionada por mecanismos automatizados de seleção, recomendação e visibilidade.

O Direito brasileiro começou a responder a essa nova configuração por meio de normas direcionadas ao uso da internet. O Marco Civil da Internet estabeleceu princípios, garantias, direitos e deveres para o ambiente digital, tratando de temas como liberdade de expressão, proteção da privacidade, guarda de registros e responsabilidade dos provedores. Ainda assim, o modelo normativo foi construído em um período no qual o principal debate envolvia conteúdos criados por pessoas ou distribuídos por plataformas. A inteligência artificial generativa acrescentou uma dificuldade distinta: o próprio sistema passou a produzir o conteúdo.

A evolução dos algoritmos de recomendação para os modelos generativos constitui um terceiro salto tecnológico. Sistemas tradicionais de recomendação selecionam e classificam materiais já existentes; sistemas generativos aprendem padrões em grandes conjuntos de dados e produzem novas saídas, como textos, imagens, áudios, vídeos e códigos. Segundo a Autoridade Nacional de Proteção de Dados (ANPD), os modelos generativos descrevem a formação dos dados em termos probabilísticos e podem gerar novos conteúdos por amostragem, enquanto os modelos de linguagem em larga escala são treinados em grandes volumes de textos para capturar relações sintáticas e semânticas (ANPD, 2024).

O aprendizado de máquina permitiu que sistemas computacionais identificassem padrões nos dados sem depender da programação manual de todas as regras possíveis. Com o

avanço do aprendizado profundo, redes são compostas por múltiplas camadas passaram a processar grandes bases e a realizar tarefas complexas. Nos modelos de linguagem em larga escala, a produção textual ocorre mediante o cálculo de probabilidades condicionais: considerando o texto já fornecido ou produzido, o sistema estima quais unidades linguísticas possuem maior probabilidade de aparecer em seguida (ANPD, 2024).

Essa explicação técnica é importante porque diferencia a geração de linguagem da compreensão humana. O modelo pode produzir textos gramaticalmente corretos, coerentes e formalmente sofisticados sem possuir consciência, intenção ou conhecimento do conteúdo. Ele reconhece e reproduz padrões presentes nos dados de treinamento e nas instruções recebidas, gerando uma resposta que parece adequada ao contexto. A aparência de compreensão decorre da qualidade da linguagem produzida, não da existência de um humano sobre verdade, ética ou relevância jurídica.

O lançamento público do ChatGPT, em 30 de novembro de 2022, contribuiu decisivamente para a popularização dessas ferramentas. A própria OpenAI apresentou o sistema como um modelo conversacional capaz de responder a perguntas e seguir instruções, mas reconheceu desde o início que ele poderia produzir respostas plausíveis, porém incorretas ou sem sentido. A facilidade de uso da interface tornou a IA generativa acessível a milhões de pessoas e acelerou sua incorporação em atividades acadêmicas, profissionais, comerciais e pessoais. Em dezembro de 2023, o Google anunciou o Gemini, apresentado como um modelo desenvolvido desde a origem para operar de modo multimodal, combinando texto, código, áudio, imagem e vídeo. A multimodalidade amplia a capacidade de produzir e interpretar diferentes formatos e demonstra que a IA generativa não se limita à redação automatizada, alcançando tarefas de criação visual, programação, análise documental e processamento de voz.

O ChatGPT e o Gemini são exemplos conhecidos de uma transformação mais ampla. A produção de conteúdo deixa de depender de uma redação humana individualizada e passa a resultar de uma interação entre bases de treinamento, arquitetura do modelo, parâmetros definidos pelos desenvolvedores, instruções fornecidas pelos usuários e mecanismos de segurança. Isso não significa que a atuação humana tenha desaparecido. Pessoas escolhem os dados, projetam o sistema, definem finalidades, formulam comandos e decidem como utilizar as respostas. Entretanto, o texto, a imagem ou o áudio final pode surgir sem que qualquer pessoa tenha elaborado diretamente cada uma de suas partes.

A mudança produz um novo problema de autoria e de controle. Quando um jornal publica uma informação, a mensagem pode ser associada a um repórter e a um editorial. Quando um usuário publica em uma rede social, a autoria inicial costuma estar vinculada ao seu perfil,

ainda que a plataforma determine o alcance. Já no conteúdo gerado por IA, a origem é distribuída entre diferentes agentes. O desenvolvedor criou a arquitetura; terceiros forneceram dados; o usuário formulou o comando; e o modelo gerou uma saída não redigida previamente por nenhum desses participantes.

Essa fragmentação demonstra que o avanço tecnológico ocorreu em velocidade superior à adaptação das categorias jurídicas. O Direito foi construído, em grande medida, para examinar condutas humanas identificáveis, intenções, deveres de cuidado e relações causais relativamente lineares. A IA generativa introduz conteúdos produzidos por sistemas probabilísticos, tornando mais complexa a identificação do autor material, do responsável pela verificação e do agente que deveria impedir ou reparar o dano.

Por isso, a passagem para a inteligência artificial não deve ser compreendida apenas como evolução técnica, mas como fenômeno jurídico-social. Barcarollo (2021) sustenta que os impactos éticos, legais e sociais da IA exigem análise interdisciplinar, pois a tecnologia interfere na forma como a sociedade produz conhecimento, organiza relações e constrói padrões de confiança, o que justifica a necessidade de respostas normativas compatíveis com a complexidade do ambiente digital.

2.3 O fim do filtro ético: como a “autonomia” da máquina altera a verdade jurídica

A expressão “autonomia” aplicada à inteligência artificial deve ser compreendida com cautela. Os sistemas generativos não possuem autonomia moral semelhante à humana, mas conseguem produzir resultados sem que uma pessoa determine previamente cada palavra, imagem ou decisão intermediária. Essa autonomia operacional é suficiente para criar uma distância entre a ação humana inicial e o conteúdo final, mas não transforma a máquina em sujeito dotado de consciência, ética ou responsabilidade jurídica própria.

Nos meios de comunicação, o editor podia avaliar se uma informação era relevante, ofensiva, verdadeira ou suficientemente verificada antes de autorizar sua publicação. Um modelo de linguagem não realiza esse juízo no sentido humano. Sua operação baseia-se em relações estatísticas aprendidas durante o treinamento. A ANPD (2024) registra que esses modelos podem gerar conteúdo verossímil e formalmente correto, porém inverídico, fenômeno conhecido como alucinação algorítmica. O documento também destaca que os modelos não possuem compreensão real do texto e funcionam principalmente a partir de padrões estatísticos.

A crítica à opacidade algorítmica reforça essa preocupação. Di Lauro e Baracuhy (2025) observam que os algoritmos organizam e filtram dados, controlam fluxos de informação e

mapeiam preferências dos usuários, de modo que, nas redes sociais, passam a definir o que aparece, para quem e em qual momento. Essa mediação automatizada evidencia que a circulação informacional nas plataformas digitais não é neutra, pois os sistemas algorítmicos atuam como mecanismos de seleção, visibilidade e exclusão de discursos.

A alucinação é especialmente perigosa porque não se apresenta necessariamente como um erro evidente. A resposta pode conter linguagem técnica, estrutura lógica, referências aparentes e tom de segurança, induzindo o usuário a acreditar que houve consulta a uma fonte confiável. Diferentemente de uma simples falha de digitação, o sistema pode inventar fatos, datas, pessoas, documentos, obras acadêmicas e precedentes judiciais. Quanto maior a qualidade formal da resposta, maior pode ser a dificuldade de perceber a falsidade do conteúdo.

A própria apresentação inicial do ChatGPT reconheceu a possibilidade de geração de respostas plausíveis, mas incorretas, além da sensibilidade do modelo a pequenas alterações na formulação das perguntas. Isso demonstra que a fluidez da linguagem não pode ser confundida com compromisso com a verdade. O modelo produz uma continuação estatisticamente compatível com o contexto, e não uma declaração submetida, necessariamente, à verificação correta.

Esse fenômeno contribui para um colapso da confiança digital. Conforme explicam Chesney e Citron (2019), fotografias, vídeos e gravações tradicionalmente possuem forte poder persuasivo por aparentarem registrar acontecimentos de forma direta. Contudo, o desenvolvimento de tecnologias de inteligência artificial generativa e de *deepfakes* passou a desafiar essa confiança, pois tornou possível produzir áudios e vídeos de pessoas aparentando dizer ou fazer coisas que jamais disseram ou fizeram, com níveis crescentes de realismo e dificuldade de detecção. A dúvida deixa de recair apenas sobre textos anônimos e passa a alcançar documentos audiovisuais que, em outros períodos, seriam tratados como fortes indícios de autenticidade.

A ANPD (2024) alerta que conteúdos sintéticos podem produzir narrativas falsas sobre pessoas reais e afetar sua imagem, honra, reputação e dignidade. Um sistema pode, por exemplo, associar falsamente uma pessoa a determinado comportamento, produzir imagens inexistentes ou elaborar interpretações infundadas a partir de dados incompletos. Nesses casos, o dano não decorre somente da circulação de uma mentira, mas também da capacidade de personalização e reprodução em larga escala.

A transformação afeta diretamente a verdade jurídica. O processo judicial depende de elementos que permitam reconstruir acontecimentos e avaliar a confiabilidade das alegações. Quando textos, imagens, áudios ou documentos podem ser fabricados por sistemas acessíveis e

apresentados com aparência profissional, cresce a necessidade de verificar origem, integridade e contexto. Ao mesmo tempo, a mera alegação de que determinado material pode ter sido produzido por IA não pode servir para eliminar automaticamente seu valor probatório. O desafio consiste em desenvolver critérios técnicos e jurídicos capazes de distinguir conteúdo autêntico, conteúdo manipulado e conteúdo inteiramente sintético.

A utilização de IA por profissionais também demonstra que a existência da ferramenta não elimina o dever humano de conferência. No julgamento do processo RR-0000284-92.2024.5.06.0351, em março de 2026, a Sexta Turma do Tribunal Superior do Trabalho aplicou sanções a uma empresa e a seu advogado após a apresentação de citações de jurisprudência inexistentes em peça processual. Embora a notícia oficial tenha tratado o possível uso de inteligência artificial como hipótese diante das características do material apresentado, o fundamento central da decisão foi a responsabilidade de quem subscreveu e apresentou informações falsas ao processo. O episódio evidencia que a automação não afasta os deveres de lealdade, diligência e verificação impostos aos participantes da relação processual.

O caso é representativo porque revela a coexistência de duas dimensões de responsabilidade. A primeira é a responsabilidade do usuário profissional que decide empregar uma resposta sem conferir sua exatidão. A segunda, mais complexa e diretamente relacionada ao objeto deste trabalho, diz respeito às empresas que desenvolvem, disponibilizam e exploram economicamente sistemas capazes de produzir informações falsas com aparência de confiabilidade. Reconhecer o dever de conferência do usuário não resolve, por si só, a questão sobre os deveres de prevenção, transparência e segurança dos fornecedores da tecnologia.

A ausência de um editor humano visível cria, assim, um espaço de indeterminação. A máquina não possui patrimônio próprio, não comparece em juízo e não assume deveres éticos. A análise jurídica precisa retornar às pessoas naturais e jurídicas envolvidas na concepção, disponibilização, implementação e utilização do sistema. Entretanto, a multiplicidade de participantes torna difícil definir qual conduta contribuiu de maneira juridicamente relevante para o dano.

Além disso, os sistemas são frequentemente descritos como caixas-pretas porque o caminho interno que conduz a determinado resultado pode ser difícil de explicar em linguagem compreensível. Mesmo quando é possível examinar tecnicamente parâmetros, registros e etapas do processamento, a relação entre milhões ou bilhões de ajustes matemáticos e uma resposta específica não se apresenta de forma intuitiva. A opacidade amplia a assimetria informacional entre o usuário prejudicado e a empresa que controla o desenvolvimento e a documentação do sistema.

A partir dessa perspectiva, a opacidade não é apenas um problema técnico, mas também jurídico. Se o funcionamento do sistema permanece incompreensível para o usuário comum, a identificação da autoria, da causalidade e dos deveres de cuidado torna-se mais difícil, aproximando o debate sobre IA generativa da necessidade de mecanismos de transparência, auditabilidade e governança capazes de reduzir a distância informacional entre fornecedores e pessoas afetadas.

A dificuldade de explicação não significa ausência automática de responsabilidade, mas impede respostas simplistas. Beuron e Cristóvam (2025) observam que os algoritmos implementados por inteligência artificial não são neutros, pois produzem efeitos a partir de dados, decisões e circunstâncias previamente incorporadas ao sistema, podendo funcionar como verdadeiras "caixas-pretas" diante da falta de informação e transparência sobre sua concepção e funcionamento. Por isso, não basta afirmar que a máquina agiu sozinha, pois sua existência, finalidade e disponibilização resultam de decisões humanas e empresariais. Também não é adequado atribuir toda falha ao desenvolvedor sem examinar a forma de implementação, o comportamento do usuário, os avisos existentes, a previsibilidade do risco e os mecanismos de controle disponíveis. A responsabilidade civil exige a identificação do dano, dos sujeitos envolvidos, dos deveres aplicáveis e do nexo entre a atividade e o prejuízo.

A passagem do filtro humano ao processamento algorítmico altera, portanto, o problema jurídico. No jornalismo tradicional, a discussão concentrava-se no conteúdo divulgado e na conduta de autores, editores e veículos. Nas redes sociais, acrescentou-se a responsabilidade dos provedores pela hospedagem, moderação e impulsionamento de conteúdos produzidos por terceiros. Na inteligência artificial generativa, surge uma nova camada: a plataforma deixa de apenas armazenar ou recomendar e passa a participar da própria geração do conteúdo potencialmente lesivo.

Essa mudança não conduz, antecipadamente, a uma única solução. Ela exige investigar se os regimes existentes de responsabilidade subjetiva, objetiva, consumerista e relacionados ao risco da atividade são suficientes para enfrentar os danos; como deve ser demonstrado o nexo de causalidade em sistemas opacos; quais deveres de informação e segurança recaem sobre os fornecedores; e em que medida o comportamento do usuário pode influenciar a imputação. Essas questões serão examinadas no capítulo seguinte, dedicado à responsabilidade civil e ao Direito Digital no ordenamento brasileiro.

3 A RESPONSABILIDADE CIVIL E O DIREITO DIGITAL NO ORDENAMENTO BRASILEIRO

A análise dos danos provocados por sistemas de inteligência artificial generativa exige a aproximação entre a responsabilidade civil e os desafios próprios do ambiente digital. O Direito Civil brasileiro foi estruturado a partir de categorias como conduta, dano, culpa, risco e nexo de causalidade. Essas categorias continuam sendo indispensáveis, mas passam a enfrentar dificuldades específicas quando o resultado danoso decorre de uma cadeia tecnológica composta por desenvolvedores, fornecedores, usuários, plataformas, bases de dados, modelos algorítmicos e mecanismos de automação.

No ordenamento brasileiro, a responsabilidade civil possui como fundamento geral a prática de ato ilícito, nos termos dos arts. 186[6] e 187[7] do Código Civil, e o dever de reparar o dano, previsto no art. 927[8]. O parágrafo único do art. 927 também admite a responsabilidade independentemente de culpa quando a lei assim determinar ou quando a atividade normalmente desenvolvida pelo autor do dano implicar, por sua natureza, risco para os direitos de terceiros. Essa abertura normativa permite discutir a incidência da responsabilidade objetiva em atividades tecnológicas que, embora lícitas e economicamente úteis, podem produzir danos relevantes a usuários e terceiros.

O problema, contudo, não se resolve pela simples afirmação de que a inteligência artificial é perigosa ou de que todo dano tecnológico deve gerar responsabilidade objetiva. A complexidade técnica dos sistemas exige identificar quem participou da atividade, qual dever jurídico foi violado, se houve defeito de segurança ou informação, como o dano ocorreu e qual vínculo causal pode ser estabelecido entre a conduta humana ou empresarial e o prejuízo experimentado pela vítima. Por isso, este capítulo examina, inicialmente, a responsabilidade objetiva e a teoria do risco-proveito; em seguida, analisa o Marco Civil da Internet e a interpretação conferida pelo Supremo Tribunal Federal aos Temas 987 e 533; por fim, aborda o desafio do nexo de causalidade em sistemas algorítmicos opacos, frequentemente descritos como caixas-pretas.

3.1 A responsabilidade objetiva e a teoria do risco-proveito

A responsabilidade civil subjetiva parte, em regra, da necessidade de demonstração de conduta culposa ou dolosa, dano e nexo causal. Já a responsabilidade objetiva é o centro da análise da culpa para o risco da atividade ou para hipóteses legalmente definidas. No contexto de sistemas digitais, essa distinção se torna especialmente relevante porque muitos danos não decorrem de uma ação humana direta e imediatamente identificável, mas de um processo técnico que opera de forma automatizada, escalável e, muitas vezes, imprevisível para o usuário comum.

---

6 Art. 186. Aquele que, por ação ou omissão voluntária, negligência ou imprudência, violar direito e causar dano a outrem, ainda que exclusivamente moral, comete ato ilícito (Brasil, 2002).

7 Art. 187. Também comete ato ilícito o titular de um direito que, ao exercê-lo, excede manifestamente os limites impostos pelo seu fim econômico ou social, pela boa-fé ou pelos bons costumes (Brasil, 2002).

8 Art. 927. Aquele que, por ato ilícito, causar dano a outrem, fica obrigado a repará-lo.
Parágrafo único. Haverá obrigação de reparar o dano, independentemente de culpa, nos casos especificados em lei, ou quando a atividade normalmente desenvolvida pelo autor do dano implicar, por sua natureza, risco para os direitos de outrem (Brasil, 2002).

A teoria do risco-proveito oferece um critério importante para essa discussão. De acordo com essa perspectiva, aquele que explora economicamente determinada atividade e dela obtém vantagens deve também suportar os riscos que essa atividade impõe a terceiros. Nesse sentido, Costa (2021) observa que, quando a utilização de sistemas inteligentes ocorre como fonte de receita ou meio de ganho, a análise do caso pode ser realizada sob a ótica da Teoria do Risco, admitindo-se a aplicação da responsabilidade objetiva. Aplicada ao desenvolvimento e fornecimento de sistemas de inteligência artificial, essa lógica permite sustentar que empresas que lucram com a disponibilização de modelos generativos devem adotar padrões rigorosos de segurança, informação, prevenção e controle, especialmente quando o produto ou serviço pode afetar direitos da personalidade, relações de consumo, dados pessoais ou a confiança pública na informação.

A aplicação dessa teoria, entretanto, precisa ser delimitada. Nem todo erro produzido por sistema de inteligência artificial autoriza, automaticamente, a responsabilização da empresa desenvolvedora. É necessário verificar se o dano decorreu de falha de concepção, treinamento, integração, atualização, segurança, informação, moderação ou supervisão; se o usuário empregou a ferramenta de forma abusiva; se havia aviso claro sobre as limitações do sistema; e se o resultado danoso era previsível dentro dos riscos próprios da atividade. A responsabilidade objetiva não elimina a necessidade de comprovação do dano e do nexo causal, mas afasta, em determinadas hipóteses, a exigência de comprovação da culpa.

Costa (2021) ainda observa que a atribuição de personalidade jurídica aos sistemas autônomos ainda não encontra base suficiente no Direito brasileiro. A responsabilização deve, portanto, retornar aos sujeitos humanos e empresariais envolvidos na concepção, fabricação, operação, programação, disponibilização ou utilização da tecnologia. O autor também aponta que a escolha do regime aplicável pode variar conforme a posição ocupada por cada agente na cadeia de produção e uso do sistema, admitindo-se tanto hipóteses de responsabilidade subjetiva quanto hipóteses de responsabilidade objetiva fundada no risco.

Essa observação é importante para evitar duas simplificações igualmente problemáticas. A primeira seria tratar a máquina como se fosse um sujeito autônomo e responsável por seus próprios atos, afastando a análise das pessoas físicas e jurídicas que a projetam, treinam, comercializam ou utilizam. A segunda seria imputar todo resultado danoso à empresa desenvolvedora, sem examinar a cadeia concreta de participação e os deveres jurídicos assumidos por cada agente. Entre esses extremos, o caminho mais adequado parece ser a identificação funcional dos sujeitos envolvidos: desenvolvedor, fornecedor, integrador, plataforma, contratante, usuário profissional ou usuário comum.

Nas relações de consumo, o Código de Defesa do Consumidor também pode incidir quando o sistema de inteligência artificial for fornecido como produto ou serviço no mercado. Nesses casos, a responsabilidade pelo fato do produto ou do serviço tende a proteger a vítima diante de defeitos de segurança, informação ou funcionamento, sem prejuízo da análise das excludentes previstas em lei (Brasil, 1990). Essa lógica pode ser relevante para ferramentas de IA generativa ofertadas ao público em geral, especialmente quando a empresa cria expectativa de confiabilidade, utilidade profissional ou segurança.

A responsabilidade objetiva, portanto, não deve ser compreendida como punição pelo simples uso da tecnologia, mas como técnica jurídica de distribuição dos riscos em atividades econômicas complexas. Quando a empresa desenvolvedora cria, treina, disponibiliza e explora economicamente um sistema capaz de gerar conteúdos falsos, discriminatórios ou lesivos, torna-se necessário perguntar se o dano é expressão de risco próprio da atividade e se o fornecedor adotou medidas suficientes de prevenção, transparência, controle e reparação.

3.2 O Marco Civil da Internet (Lei n.º 12.965/2014) e a nova interpretação do STF sobre os temas 987 e 533

O Marco Civil da Internet estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil, consolidando a liberdade de expressão, a proteção da privacidade, a proteção de dados pessoais, a preservação da neutralidade de rede e a responsabilização dos agentes que atuam no ambiente digital. Entre seus dispositivos mais relevantes para a responsabilidade civil está o art. 19[9] da Lei n.º 12.965/2014, que buscou evitar a censura privada ao prever que o provedor de aplicações de internet somente poderia ser responsabilizado civilmente por danos decorrentes de conteúdo gerado por terceiros se, após ordem judicial

9 Art. 19. Com o intuito de assegurar a liberdade de expressão e impedir a censura, o provedor de aplicações de internet somente poderá ser responsabilizado civilmente por danos decorrentes de conteúdo gerado por terceiros se, após ordem judicial específica, não tomar as providências para, no âmbito e nos limites técnicos do seu serviço e dentro do prazo assinalado, tornar indisponível o conteúdo apontado como infringente, ressalvadas as disposições legais em contrário. § 1º A ordem judicial de que trata o caput deverá conter, sob pena de nulidade, identificação clara e específica do conteúdo apontado como infringente, que permita a localização inequívoca do material. § 2º A aplicação do disposto neste artigo para infrações a direitos de autor ou a direitos conexos depende de previsão legal específica, que deverá respeitar a liberdade de expressão e demais garantias previstas no art. 5º da Constituição Federal. § 3º As causas que versem sobre ressarcimento por danos decorrentes de conteúdos disponibilizados na internet relacionados à honra, à reputação ou a direitos de personalidade, bem como sobre a indisponibilização desses conteúdos por provedores de aplicações de internet, poderão ser apresentadas perante os juizados especiais. § 4º O juiz, inclusive no procedimento previsto no § 3º, poderá antecipar, total ou parcialmente, os efeitos da tutela pretendida no pedido inicial, existindo prova inequívoca do fato e considerado o interesse da coletividade na disponibilização do conteúdo na internet, desde que presentes os requisitos de verossimilhança da alegação do autor e de fundado receio de dano irreparável ou de difícil reparação (Brasil, 2014).

específica, não tomasse as providências necessárias para tornar indisponível o conteúdo apontado como infringente.

Esse modelo foi construído para equilibrar dois valores constitucionais em tensão: de um lado, a proteção de direitos fundamentais atingidos por conteúdos ilícitos; de outro, a liberdade de expressão e a necessidade de evitar que plataformas removessem conteúdos de forma excessiva por medo de responsabilização. O art. 21[10] do Marco Civil, por sua vez, trouxe para casos de divulgação, sem autorização dos envolvidos, de imagens, vídeos ou outros materiais contendo cenas de nudez ou atos sexuais de caráter privado, admitindo responsabilização subsidiária do provedor caso, após notificação, deixasse de tornar indisponível o conteúdo.

A constitucionalidade e o alcance do art. 19 foram examinados pelo Supremo Tribunal Federal (STF) nos Recursos Extraordinários n.º 1.037.396, Tema 987, e n.º 1.057.258, Tema 533. O Tema 987 envolveu a criação de perfil falso no Facebook em nome de pessoa que não possuía conta na rede social, utilizado para ofender terceiros. A plataforma foi notificada por ferramenta interna, mas não removeu o perfil antes da judicialização; posteriormente, cumpriu a ordem judicial de exclusão, discutindo-se a possibilidade de indenização. O Tema 533 tratou de uma comunidade criada no Orkut para ofender uma professora, cuja remoção foi inicialmente recusada pela plataforma e posteriormente determinada judicialmente.

As questões jurídicas submetidas ao STF (Brasil, 2025) consistiam em saber se o art. 19 do Marco Civil da Internet, ao exigir ordem judicial prévia para responsabilizar plataformas por conteúdo de terceiros, seria constitucional, e qual regime de responsabilidade deveria ser adotado para compatibilizar proteção de direitos fundamentais, liberdade de expressão e valores democráticos no ambiente digital.

No julgamento realizado em 26 de junho de 2025, o Supremo Tribunal Federal (Brasil, 2025) reconheceu, por maioria, a inconstitucionalidade parcial e progressiva do art. 19 do Marco Civil da Internet. Entenderam que a regra geral que condicionava a responsabilização ao descumprimento de ordem judicial específica não conferia proteção suficiente a bens jurídicos constitucionais de alta relevância, especialmente direitos fundamentais e democracia. Ao

10 Art. 21. O provedor de aplicações de internet que disponibilize conteúdo gerado por terceiros será responsabilizado subsidiariamente pela violação da intimidade decorrente da divulgação, sem autorização de seus participantes, de imagens, de vídeos ou de outros materiais contendo cenas de nudez ou de atos sexuais de caráter privado quando, após o recebimento de notificação pelo participante ou seu representante legal, deixar de promover, de forma diligente, no âmbito e nos limites técnicos do seu serviço, a indisponibilização desse conteúdo. Parágrafo único. A notificação prevista no caput deverá conter, sob pena de nulidade, elementos que permitam a identificação específica do material apontado como violador da intimidade do participante e a verificação da legitimidade para apresentação do pedido (Brasil, 2014).

mesmo tempo, o Tribunal não eliminou integralmente o art. 19, mas delimitou hipóteses em que sua aplicação permanece necessária para resguardar a liberdade de expressão.

A tese aprovada pelo STF (Brasil, 2025) estabeleceu que, enquanto não houver nova legislação sobre o tema, os provedores de aplicações de internet poderão ser responsabilizados civilmente por danos decorrentes de conteúdos gerados por terceiros em casos de crime ou atos ilícitos, sem prejuízo do dever de remoção. A mesma lógica foi estendida às contas denunciadas como inautênticas. Para crimes contra a honra, entretanto, o Tribunal manteve a incidência do art. 19, exigindo ordem judicial específica, sem impedir que as plataformas removam conteúdos por notificação extrajudicial quando assim entenderem.

O Tribunal também fixou parâmetros específicos para situações de maior risco (Brasil, 2025). Em anúncios, impulsionamentos pagos e redes artificiais de distribuição, a decisão estabelece presunção de responsabilidade dos provedores, afastável mediante demonstração de atuação diligente e em tempo razoável para tornar o conteúdo indisponível. Além disso, nos casos de circulação massiva de conteúdos ilícitos graves, como terrorismo, induzimento ao suicídio ou à automutilação, pornografia infantil, discriminação, discurso de ódio, crimes contra mulheres em razão de gênero, tráfico de pessoas e atos antidemocráticos, foi reconhecido dever de cuidado, cuja violação depende da configuração de falha sistêmica do provedor.

A decisão também preservou a incidência integral do art. 19 para certos provedores considerados neutros, como serviços de e-mail, aplicações voltadas a reuniões fechadas e mensageria instantânea, exclusivamente quanto às comunicações interpessoais protegidas pelo sigilo constitucional. No caso dos marketplaces, o STF (Brasil, 2025) afirmou que a responsabilidade civil deve ser analisada de acordo com o Código de Defesa do Consumidor. O Tribunal ainda impôs deveres adicionais de autorregulação, canais de atendimento, sistemas de notificação, devido processo, relatórios anuais de transparência e representação legal no Brasil para provedores estrangeiros que atuem no país[11].

Um ponto essencial para este trabalho é que o STF expressamente afastou a responsabilidade objetiva na aplicação da tese fixada para plataformas digitais por conteúdo de terceiros (Brasil, 2025). A responsabilização, nesse modelo, permanece subjetiva, exigindo análise de culpa ou dolo, embora com deveres específicos de diligência, transparência, prevenção e remoção em determinadas hipóteses. Essa conclusão impede que a decisão seja

11 No dia 17 de junho de 2026, o STF decidiu os ajustes na tese de repercussão geral que foi estabelecida durante o julgamento sobre a responsabilização das plataformas em relação ao conteúdo de terceiros, e ficou decidido que as empresas terão um prazo de 60 dias, contados a partir do término do julgamento, para implementar as medidas estruturais que decorrem do dever de cuidado, incluindo os mecanismos de autorregulação, de atendimento, de notificação, do devido processo e de transparência mencionados.

utilizada como fundamento direto para afirmar responsabilidade objetiva de plataformas em todos os casos digitais.

Apesar disso, os Temas 987 e 533 são relevantes para a responsabilidade por sistemas de inteligência artificial generativa. Ainda que a decisão trate de conteúdo publicado por terceiros, e não de conteúdo produzido diretamente por modelos generativos, ela demonstra uma tendência de deslocamento do debate para deveres estruturais de cuidado, transparência, prevenção, documentação e gestão de risco. No caso da IA generativa, a discussão pode ser ainda mais intensa, pois a empresa desenvolvedora não apenas hospeda ou organiza conteúdo alheio, mas fornece sistema capaz de produzir novas informações, imagens, textos ou recomendações a partir de sua arquitetura técnica.

Por isso, a decisão do STF não resolve automaticamente o regime de responsabilidade das empresas desenvolvedoras de IA, mas oferece parâmetros úteis: a análise deve considerar o papel concreto do provedor, o grau de interferência no fluxo informacional[12], a previsibilidade do risco, a existência de deveres de cuidado, a presença de falha sistêmica e os mecanismos internos de prevenção, atendimento, transparência e resposta a danos. Esses elementos dialogam diretamente com os desafios da responsabilidade civil diante de algoritmos opacos.

3.3 O desafio do nexo de causalidade em algoritmos de "caixa-preta"

O nexo de causalidade é um dos elementos mais difíceis da responsabilidade civil aplicada à inteligência artificial. Em situações tradicionais, a vítima busca demonstrar que determinada conduta humana ou empresarial produziu o dano. Nos sistemas de IA generativa, entretanto, o resultado pode decorrer da interação entre múltiplas variáveis: dados de treinamento, arquitetura do modelo, parâmetros técnicos, filtros de segurança, atualizações, integração com outras plataformas, comando formulado pelo usuário e forma de utilização do conteúdo produzido.

12 Em nota divulgada em março de 2025, o CGI.br propôs uma tipologia que classifica os provedores de aplicação em três categorias, conforme o grau de interferência sobre a circulação de conteúdo de terceiros: provedores sem interferência que atuam apenas como meio de transporte e armazenamento, como hospedagem e e-mail, com baixa interferência sem uso de perfilização do usuário, como sites de edição colaborativa e com alta interferência que utilizam recomendação algorítmica. A proposta busca a especificação dos regimes de responsabilidade civil dos agentes.

Essa multiplicidade de fatores dificulta a reconstrução do caminho causal. O usuário prejudicado geralmente não tem acesso aos dados internos do sistema, aos registros completos de funcionamento, aos critérios de treinamento, às políticas de moderação, aos testes de segurança ou às decisões empresariais que orientaram a disponibilização da ferramenta. A empresa desenvolvedora, por outro lado, costuma deter maior capacidade técnica e documental para explicar o funcionamento do sistema, ainda que nem sempre seja possível reconstruir integralmente por que uma saída específica foi produzida.

A doutrina recente tem associado esse problema à opacidade algorítmica e à ideia de caixa-preta. Beuron e Cristóvam (2025) destacam que algoritmos implementados por inteligência artificial não podem ser tratados como neutros, pois são criados e orientados por escolhas humanas, bases de dados e parâmetros técnicos. Os autores também apontam que a falta de informação e transparência sobre a formação das bases de dados e o processamento automatizado pode transformar esses sistemas em verdadeiras caixas-pretas, dificultando o controle público e a compreensão dos seus efeitos.

No campo probatório, Rebehy e Alliceda (2025) observam que a inteligência artificial intensifica a importância de estruturas de controle, registros e segurança da informação, sobretudo quando a discussão envolve tratamento de dados, prova digital e possibilidade de rastreamento do funcionamento do sistema. Ao analisarem a relação entre LGPD, prova digital e IA, os autores destacam que os agentes envolvidos precisam manter padrões de segurança, governança e documentação capazes de permitir a compreensão de operações automatizadas e a proteção dos titulares de dados.

A partir disso, a caixa-preta algorítmica não deve ser compreendida como argumento para afastar a responsabilidade, mas como fator que torna mais complexa a prova do nexo causal. Se a vítima não consegue demonstrar, sozinha, como determinado dano foi gerado, é necessário examinar a distribuição dos ônus probatórios, a assimetria informacional entre as partes e os deveres de documentação, transparência e auditabilidade que podem recair sobre fornecedores e desenvolvedores.

A dificuldade de explicação também impede respostas simples. Não basta afirmar que a máquina agiu sozinha, pois sua existência, finalidade e disponibilização resultam de decisões humanas. Também não é adequado atribuir toda falha ao desenvolvedor sem examinar a forma de implementação, o comportamento do usuário, os avisos existentes, a previsibilidade do risco e os mecanismos de controle disponíveis. A responsabilidade civil exige a identificação do dano, dos sujeitos envolvidos, dos deveres aplicáveis e do nexo entre a atividade e o prejuízo.

Nesse ponto, a decisão do STF nos Temas 987 e 533[13] oferece um parâmetro importante, ainda que não trate diretamente de inteligência artificial generativa. Ao exigir análise de culpa ou dolo, mas ao mesmo tempo criar deveres de cuidado, notificação, transparência, prevenção e atenção à falha sistêmica, o Tribunal sinaliza que a responsabilidade no ambiente digital não depende apenas de identificar um ato isolado. Pode ser necessário avaliar a estrutura de funcionamento do serviço e a diligência adotada pelo provedor diante de riscos conhecidos ou previsíveis.

Essa lógica pode ser transposta, com cautela, para o estudo da IA generativa. Em vez de perguntar apenas se a empresa teve intenção de produzir determinado conteúdo danoso, deve-se investigar se o sistema foi concebido com padrões adequados de segurança; se havia mecanismos de mitigação de alucinações, vieses e usos abusivos; se os usuários foram informados sobre limitações relevantes; se houve monitoramento de riscos; se existiam canais eficazes de denúncia e correção; e se a empresa atuou de modo diligente após tomar conhecimento de falhas recorrentes.

A causalidade, portanto, pode assumir forma mais estrutural. O dano individual pode decorrer de uma resposta específica, mas essa resposta pode estar ligada a falhas de treinamento, documentação, teste, supervisão, moderação ou governança. Quando o problema é recorrente, previsível ou derivado de escolhas de design, a análise não deve se limitar ao comando isolado do usuário. Nesses casos, a discussão sobre nexo causal aproxima-se da avaliação do risco da atividade e dos deveres empresariais de prevenção.

Por outro lado, a existência de opacidade não autoriza responsabilização automática. O comportamento do usuário, a finalidade do uso, o contexto de aplicação, a possibilidade de conferência humana, a natureza do dano e a existência de terceiros intermediários precisam ser considerados. Ferramentas de IA generativa podem ser utilizadas em contextos diversos, desde atividades recreativas até decisões profissionais sensíveis. Quanto maior o grau de confiança induzido pela ferramenta e quanto maior o risco de dano, mais intenso deve ser o dever de cuidado do fornecedor.

Dessa forma, o nexo de causalidade em algoritmos de caixa-preta deve ser analisado a partir de uma combinação entre prova técnica, deveres de transparência, distribuição do ônus probatório e exame da cadeia de agentes envolvidos. A opacidade algorítmica não elimina a

13 O entendimento do Supremo Tribunal Federal sobre os Temas 987 e 533 foi definitivamente consolidado pelo Plenário da Corte com o julgamento dos embargos de declaração nos Recursos Extraordinários n.º 1.037.396 e n.º 1.057.258. A redação final da tese aperfeiçoou os parâmetros de responsabilização estrutural e deveres de cuidado das plataformas digitais, estabelecendo, inclusive, o prazo de 60 dias para a implementação de canais específicos de denúncia pelas empresas de tecnologia (Brasil, 2026).

responsabilidade civil, mas exige que o Direito adapte seus instrumentos para não transformar a complexidade técnica em obstáculo absoluto à reparação da vítima.

4 ANÁLISE DE CASOS PARADIGMÁTICOS DE REPERCUSSÃO

Após a análise da responsabilidade civil e dos limites do Marco Civil da Internet, este capítulo examina situações em que sistemas de inteligência artificial produziram ou potencializaram danos socialmente relevantes. A finalidade não é esgotar a investigação técnica de cada evento, mas demonstrar como falhas de filtros, treinamento, moderação e governança podem repercutir sobre direitos da personalidade, integridade informacional e democracia.

Os casos selecionados permitem observar três dimensões complementares do problema. O primeiro, envolvendo o Grok, da xAI, evidencia o conflito entre sistemas publicamente associados à liberdade de expressão e a produção automatizada de discurso discriminatório. O segundo, que aproxima o episódio Microsoft Tay ao caso do *chatbot* Eliza, do aplicativo Chai, mostra que a fragilidade ética dos sistemas conversacionais pode decorrer tanto da manipulação por terceiros quanto da interação prolongada por usuários vulneráveis. O terceiro desloca a discussão para o campo eleitoral, no qual a IA generativa pode produzir danos coletivos à confiança pública e à formação livre da vontade política.

Esses episódios reforçam a ideia desenvolvida nos capítulos anteriores: a inteligência artificial não deve ser tratada como sujeito autônomo capaz de assumir culpa ou responsabilidade própria. A máquina não possui patrimônio, consciência moral ou dever jurídico. A análise, portanto, deve retornar aos agentes humanos e empresariais que projetam, treinam, disponibilizam, integram, moderam e exploram economicamente essas tecnologias.

4.1 O caso grok (xAi): antissemitismo e o conflito entre “livre expressão” e discurso de ódio

O caso Grok, sistema de inteligência artificial da xAI integrado à plataforma X, opera em uma rede social de circulação rápida de discursos, produzindo efeitos sobre terceiros. O episódio, ocorrido nos EUA e sem julgamento formal, é analisado neste trabalho com base em registros jornalísticos, sem caracterizar, portanto, uma decisão judicial brasileira de responsabilização civil da xAI.

Em julho de 2025, uma reportagem de Yang (2025) expôs publicações da conta do Grok no X contendo discursos antissemitas, elogios a Adolf Hitler e termos ofensivos direcionados aos sobrenomes judeus. Diante do ocorrido, a Anti-Defamation League classificou a conduta como perigosa e irresponsável, alertando que criadores de grandes modelos de linguagem devem restringir conteúdos baseados em ódio extremista. Posteriormente, a xAI recolheu os posts inadequados, reconheceu a falha publicamente e sinalizou a adoção de barreiras para mitigar novos episódios.

Do ponto de vista jurídico, o caso é relevante porque desloca a discussão da simples moderação de conteúdo produzido por terceiros para a geração de conteúdo pelo próprio sistema automatizado. Em uma rede social tradicional, a plataforma costuma ser analisada como intermediária: ela hospeda, recomenda ou remove conteúdo criado por usuários. No caso de uma IA generativa integrada à plataforma, há uma camada adicional: a própria ferramenta

disponibilizada pela empresa produz respostas, replica padrões discursivos e pode conferir aparência de autoridade técnica a afirmações discriminatórias.

A controvérsia também expõe a insuficiência de respostas baseadas apenas na ideia abstrata de livre expressão. A liberdade de expressão constitui garantia essencial do Estado Democrático de Direito, mas não protege manifestações discriminatórias que atinjam a dignidade de grupos historicamente vulnerabilizados. Quando um sistema automatizado produz discurso de ódio, a dificuldade não está apenas em remover a postagem depois do dano, mas em compreender por que o modelo gerou aquela resposta, quais parâmetros permitiram sua circulação e quais deveres de prevenção deveriam ter sido adotados previamente.

Nesse sentido, o caso Grok evidencia a tensão entre modelos de IA utilizados para responder de modo mais permissivo, provocativo ou supostamente "anticensura" e os deveres jurídicos de segurança, prevenção e não discriminação. A retórica empresarial de neutralidade tecnológica não elimina o fato de que sistemas generativos dependem de escolhas humanas: seleção de dados, definição de políticas de segurança, prompts de sistema, filtros de moderação e critérios de atualização. Por isso, mesmo que a resposta ofensiva surja de uma interação específica, ela não pode ser tratada automaticamente como fato externo à atividade empresarial.

A literatura sobre mediação algorítmica permite compreender esse fenômeno como parte de um regime de visibilidade. Di Lauro e Baracuhy (2025) observam que algoritmos não apenas organizam conteúdos, mas exercem funções estratégicas na produção de subjetividades e na regulação de discursos, selecionando, restringindo e hierarquizando aquilo que aparece nas plataformas digitais. No ambiente do Grok, essa lógica se intensifica: o sistema não apenas distribui discursos, mas participa de sua própria formulação.

Assim, o episódio deve ser lido como falha de governança algorítmica, e não como simples "erro isolado" da máquina. A lesividade potencial do discurso automatizado decorre da combinação entre aparência de inteligência, integração em plataforma de larga escala e capacidade de replicação instantânea. A relevância do caso para a responsabilidade civil está justamente em demonstrar que a empresa desenvolvedora não controla apenas um produto passivo, mas uma infraestrutura comunicacional capaz de produzir danos à honra coletiva, à igualdade e à confiança pública.

4.2 De Microsoft Tay ao Chai app "Eliza": a fragilidade ética frete à manipulação e aos trolls

O caso Microsoft Tay, ocorrido em 2016, é um dos exemplos mais conhecidos da vulnerabilidade dos sistemas diante da manipulação coordenada por usuários. Tay foi lançado pela Microsoft como *chatbot* voltado à interação com jovens adultos nos Estados Unidos. Poucas horas após sua disponibilização no Twitter, passou a publicar mensagens ofensivas e discriminatórias, levando a empresa a retirar o sistema do ar.

Em publicação no blog oficial da Microsoft em 25 de março de 2016, divulgada após o episódio, Peter Lee afirmou que a empresa havia planejado e implementado filtros, conduzido estudos com diferentes grupos de usuários e testado Tay em variadas condições antes de ampliar sua interação no Twitter. Ainda assim, nas primeiras 24 horas de funcionamento, um ataque coordenado explorou uma vulnerabilidade específica do sistema, levando o *chatbot* a publicar conteúdos ofensivos. A Microsoft pediu desculpas pelos conteúdos gerados e reconheceu ter assumido responsabilidade por não prever aquela forma específica de abuso.

A importância jurídica do caso está no reconhecimento de que a interação pública com usuários mal-intencionados não é evento completamente estranho ao risco da atividade. Em sistemas que são abertos, é previsível que usuários tentem contornar filtros, induzir respostas abusivas, testar limites e manipular o comportamento da ferramenta. A existência de trolls não afasta, por si só, a necessidade de desenho preventivo, monitoramento contínuo e mecanismos de contenção, especialmente quando o produto é colocado em ambiente de comunicação massiva.

O caso Tay também antecipa uma dificuldade que se tornaria recorrente nas IAs generativas: a fronteira entre aprendizado, imitação e reprodução de danos. Quanto mais um sistema é estruturado para adaptar sua linguagem ao ambiente social em que interage, maior é o risco de incorporar padrões discriminatórios, violentos ou manipuladores presentes nesse ambiente. O problema, portanto, não está apenas no conteúdo do prompt do usuário, mas no modo como a arquitetura do sistema transforma interações sociais em respostas aparentemente próprias.

O caso do Chai App “Eliza” evidencia outra dimensão da fragilidade ética dos sistemas: não apenas a manipulação por trolls, mas a influência emocional sobre usuários vulneráveis. Em 2023, a imprensa internacional reportou o caso de um homem belga que morreu por suicídio após interagir por semanas com o *chatbot* Eliza. De acordo com declarações da viúva da vítima, o homem sofria de uma severa ansiedade devido à crise climática global e utilizava o sistema de inteligência artificial como uma espécie de suporte emocional (El Atillah, 2023).

Desdobramentos da cobertura desse caso indicaram que a *Chai Research* agiu após a repercussão do ocorrido, implementando novos recursos de segurança em sua plataforma

voltados à proteção dos usuários. Essa atualização passou a disparar alertas preventivos e a encaminhar os usuários para serviços especializados de apoio psicológico sempre que o sistema detectasse interações associadas à ideação suicida (Bharade, 2023). A informação não resolve, por si só, a discussão sobre responsabilidade, mas demonstra que a empresa reconheceu a necessidade de mecanismos técnicos específicos para situações de risco.

Diferentemente de Tay, em que o dano decorreu de exploração hostil por usuários externos, o caso Eliza revela riscos associados à antropomorfização e à criação de vínculos afetivos com sistemas automatizados[14]. *Chatbots* são criados para responder com fluidez, continuidade e aparente empatia. Isso pode ampliar a confiança do usuário e produzir a impressão de que há compreensão humana por trás da resposta, embora o sistema opere por padrões estatísticos e não por discernimento moral.

Sob a ótica da responsabilidade civil, os dois casos demonstram que a previsibilidade do dano não se limita à falha técnica. O risco pode surgir da interação entre usuário, arquitetura, contexto emocional, ausência de salvaguardas e insuficiência de supervisão humana. A empresa desenvolvedora não precisa desejar o resultado danoso para que se discuta sua obrigação de prevenção. Basta que a atividade seja capaz de criar ou ampliar riscos relevantes, sobretudo quando direcionada a ambientes abertos ou a interações sensíveis.

Esses episódios também mostram que avisos genéricos em termos de uso são insuficientes quando a própria experiência do produto estimula confiança, proximidade e personalização. Se a empresa sabe que usuários podem atribuir características humanas ao *chatbot*, testar seus limites ou buscar apoio emocional em situações críticas, o dever de segurança deve ser pensado desde a concepção do sistema. A responsabilidade, nesse contexto, envolve não apenas remover conteúdo problemático depois da repercussão pública, mas estruturar mecanismos de mitigação antes que o dano se concretize.

A aproximação entre Tay e Eliza é relevante para este trabalho porque evidencia que o erro da IA não é uma categoria única. Pode assumir a forma de discurso discriminatório induzido por trolls, resposta ofensiva gerada por falha de moderação, aconselhamento perigoso em contexto de vulnerabilidade ou falsa aparência de empatia. A partir desse cenário, evidencia-se que os deveres de cuidado de quem desenvolve e lucra com esses sistemas envolvem a mitigação de riscos na arquitetura da IA, a fiscalização contínua de segurança e o dever de informação clara ao usuário sobre a natureza automatizada da interação.

---

14 O jornal britânico *The Guardian* noticiou uma ação civil ajuizada por Megan Garcia, mãe de Sewell Setzer III, contra a Character.AI perante a Justiça Federal do Distrito Central da Flórida, alegando negligência, morte injusta e práticas comerciais enganosas, após o adolescente de 14 anos ter desenvolvido interação intensa com o *chatbot* da plataforma antes de sua morte.

4.3 Impactos na integridade democrática: o risco da IA nas eleições de 2026

A terceira dimensão paradigmática dos danos causados por IA generativa aparece no campo eleitoral. Nesse contexto, o risco não se limita a uma vítima individualmente identificada, mas alcança a integridade do processo democrático, a igualdade entre candidaturas, a liberdade de escolha do eleitor e a confiança pública nas instituições eleitorais. As eleições gerais de 2026 tornam esse debate especialmente relevante, pois o Tribunal Superior Eleitoral (TSE) passou a regular de modo específico o uso de inteligência artificial em propaganda eleitoral.

Para o TSE, o combate à desinformação é prioridade desde 2018 e, para as eleições gerais de 2026, aprovou alterações na resolução sobre propaganda eleitoral com o objetivo de regulamentar o uso de IA por partidos, candidatos e provedores de internet. Segundo o Tribunal, a finalidade da regulamentação é impedir a disseminação de conteúdos fabricados, manipulados ou gravemente descontextualizados capazes de prejudicar o equilíbrio das eleições ou a integridade do processo eleitoral.

A Resolução n.º 23.732/2024 do TSE introduziu regras específicas sobre conteúdo de multimídia. O art. 9º-B impõe o dever de informar, de modo explícito, destacado e acessível, quando a propaganda eleitoral utilizar conteúdo produzido ou manipulado por IA. O mesmo dispositivo determina formas de rotulagem compatíveis com o tipo de material, como informação no início de peças de áudio, marca d'água e audiodescrição em imagens estáticas, regras específicas para vídeo e indicação em material impresso.

Mais severa é a regra do art. 9º-C da mesma resolução, que veda a utilização de conteúdo fabricado ou manipulado para difundir fatos notoriamente inverídicos ou descontextualizados com potencial de causar danos ao equilíbrio do pleito ou à integridade do processo eleitoral. O dispositivo também proíbe o uso de *deepfake* para prejudicar ou favorecer candidatura, inclusive mediante alteração de imagem ou voz de pessoa viva, falecida ou fictícia, prevendo que o descumprimento configura abuso de poder político e uso indevido dos meios de comunicação social, com possibilidade de cassação do registro ou do mandato.

Em 2026, a Resolução n.º 23.755 do TSE atualizou a disciplina da propaganda eleitoral. Entre os pontos relevantes, manteve o dever de rotulagem para conteúdo sintético gerado por IA e acrescentou regra que veda a publicação, republicação ou impulsionamento pago de novos conteúdos produzidos ou alterados por IA que utilizem imagem, voz ou manifestação de

candidata, candidato ou pessoa pública no período de 72 horas antes e 24 horas depois do pleito, ainda que o conteúdo esteja rotulado.

A mesma resolução também reforça mecanismos probatórios e deveres dos provedores. O art. 9º-I autoriza a inversão do ônus da prova em representações que tratem de conteúdo gerado por IA quando a dificuldade técnica de comprovação da manipulação digital tornar excessivamente onerosa a prova pelo autor. Nessa hipótese, cabe ao representado demonstrar a licitude do conteúdo, explicar como a IA foi utilizada e comprovar a veracidade da informação veiculada.

Essas normas revelam que, no campo eleitoral, a IA generativa é tratada como risco ordenado. Um áudio falso, um vídeo ou uma imagem manipulada podem circular em velocidade superior à capacidade institucional de resposta, sobretudo em períodos próximos à votação. O dano democrático possui temporalidade própria: ainda que o conteúdo seja posteriormente desmentido, a percepção do eleitor pode ter sido alterada no momento decisivo da formação de sua vontade.

A regulação eleitoral também dialoga com a discussão civil sobre responsabilidade das plataformas. A Resolução n.º 23.732/2024 do TSE atribui aos provedores de aplicação deveres de adoção e publicização de medidas para impedir ou reduzir a circulação de fatos notoriamente inverídicos ou gravemente descontextualizados que possam atingir a integridade do processo eleitoral. Entre essas medidas estão a implementação de canais de denúncia, ações corretivas e preventivas, aprimoramento de sistemas de recomendação, transparência dos resultados e avaliação de impacto de serviços sobre a integridade do pleito.

O risco eleitoral da IA não decorre apenas da falsificação perfeita. Conteúdos de baixa qualidade técnica podem ainda assim produzir efeitos políticos relevantes quando circulam de forma segmentada, emocional e repetitiva. Em ambiente de filtros algorítmicos, como já analisado no Capítulo 2, a informação é distribuída conforme perfis, engajamento e padrões de comportamento. Por isso, uma *deepfake* ou mensagem artificialmente gerada pode alcançar grupos específicos no momento em que estão mais suscetíveis à confirmação de suas crenças políticas.

A análise dos casos paradigmáticos permite concluir que a IA generativa atua como tecnologia de amplificação de riscos. No Grok, a falha se manifesta na geração pública de discurso discriminatório por um sistema integrado a uma plataforma de larga escala. Em Tay e Eliza, a fragilidade aparece na interação conversacional, seja pela manipulação coordenada por terceiros, seja pela criação de vínculos emocionais. Nas eleições, o dano ganha dimensão coletiva, pois tais conteúdos podem atingir a própria integridade democrática. Esses exemplos

demonstram que a responsabilidade civil das empresas desenvolvedoras deve ser pensada não apenas após a ocorrência do dano, mas também como instrumento de prevenção, governança e proteção da dignidade humana diante de tecnologias que produzem efeitos reais no mundo social.

5 MECANISMOS DE CONTENÇÃO E O FUTURO DA REGULAÇÃO

A análise dos capítulos anteriores demonstra que os danos provocados por sistemas de inteligência artificial generativa não podem ser enfrentados apenas depois de sua ocorrência. Embora a responsabilidade civil continue sendo instrumento indispensável para a reparação da vítima, a velocidade de produção, reprodução e circulação de conteúdos torna insuficiente um

modelo exclusivamente reativo. Quando uma informação falsa, discriminatória ou manipulada é gerada por IA e difundida em larga escala, o dano pode atingir simultaneamente a honra, a imagem, a privacidade, a proteção de dados pessoais, a confiança pública na informação e até a integridade do debate democrático. Por isso, a discussão contemporânea desloca-se progressivamente da pergunta sobre quem indeniza para a pergunta sobre como prevenir, mitigar, rastrear, explicar e corrigir os riscos antes que eles se convertam em lesões irreversíveis.

Nesse cenário, os mecanismos de contenção assumem papel central. A contenção não deve ser compreendida como censura prévia ou proibição ampla da inovação tecnológica, mas como conjunto de deveres jurídicos, técnicos e institucionais voltados à redução de danos. Trata-se de exigir que os agentes econômicos que desenvolvem, distribuem ou aplicam sistemas de IA adotem políticas de governança, documentação, transparência, supervisão humana, segurança, avaliação de riscos, canais de contestação e procedimentos de resposta a incidentes. A finalidade não é impedir o desenvolvimento tecnológico, mas impedir que a complexidade técnica se transforme em zona de irresponsabilidade.

A regulação futura da inteligência artificial, portanto, não pode ser pensada em oposição à liberdade de expressão, à inovação e à livre iniciativa. Ao contrário, a existência de regras mínimas de responsabilidade, transparência e prestação de contas tende a fortalecer a confiança social nos próprios sistemas tecnológicos. Ferramentas de IA generativa somente poderão ser incorporadas com segurança em atividades educacionais, profissionais, empresariais, jornalísticas, jurídicas e políticas se houver previsibilidade sobre os deveres dos fornecedores e sobre os direitos das pessoas afetadas. A ausência de parâmetros normativos claros favorece tanto a insegurança jurídica das empresas quanto a vulnerabilidade dos usuários.

No Brasil, esse debate aparece de forma mais evidente em três frentes complementares: o Marco Legal da Inteligência Artificial, representado pelo Projeto de Lei n.º 2.338/2023; o Projeto de Lei n.º 2.630/2020, conhecido como PL das *Fake News*, que trata da liberdade, responsabilidade e transparência na internet; e a construção doutrinária e regulatória da *accountability* algorítmica. Esses três eixos não se confundem, mas dialogam entre si. O primeiro busca estruturar uma governança de riscos própria para sistemas de IA. O segundo pretende reforçar deveres de transparência, rastreabilidade e moderação responsável nas plataformas digitais. O terceiro fornece a base teórica para exigir que sistemas automatizados sejam auditáveis, contestáveis e juridicamente imputáveis.

Este capítulo examina esses mecanismos de contenção, com atenção especial à responsabilidade das empresas desenvolvedoras de IA generativa. O objetivo não é afirmar que

a simples existência de um projeto legislativo resolve o problema da responsabilização, mas mostrar que o ordenamento brasileiro caminha para uma combinação entre prevenção, transparência, dever de cuidado e reparação. A punição, nesse contexto, não deve ser compreendida como instrumento meramente repressivo, mas como mecanismo de proteção da dignidade humana, de indução de condutas responsáveis e de preservação da confiança pública no ambiente digital.

5.1 O Marco Legal da Inteligência Artificial (Projeto de Lei n.º 2.338/2023) e a governança de riscos

O Projeto de Lei n.º 2.338/2023 constitui a principal proposta brasileira de regulação geral da inteligência artificial[15]. Após tramitar no Senado Federal, o texto foi encaminhado à Câmara dos Deputados, onde consta como proposição sujeita à apreciação do Plenário e, em julho de 2026, aparece aguardando parecer do relator na Comissão Especial destinada a analisar a matéria. A ementa do projeto o apresenta como norma voltada ao desenvolvimento, ao incentivo e ao uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana.

A centralidade da pessoa humana é o ponto de partida mais importante do projeto. O texto não trata a IA apenas como recurso econômico ou instrumento de eficiência, mas como tecnologia capaz de afetar direitos fundamentais, valores democráticos, relações de consumo, proteção de dados, trabalho, meio ambiente, propriedade intelectual e segurança informacional. O art. 1º do texto aprovado estabelece normas gerais nacionais para a governança responsável de sistemas de IA, com o objetivo de proteger direitos fundamentais, estimular a inovação responsável e garantir sistemas seguros e confiáveis em benefício da pessoa humana, do regime democrático e do desenvolvimento social, científico, tecnológico e econômico.

Essa formulação é relevante para o presente trabalho porque rompe com a ideia de que a inovação tecnológica deve ser regulada apenas depois do dano. Ao colocar a governança no centro da disciplina normativa, o PL n.º 2.338/2023 aproxima a regulação da IA de um modelo preventivo. A tecnologia deixa de ser analisada apenas pelo resultado lesivo final e passa a ser observada durante todo o seu ciclo de vida: concepção, desenvolvimento, treinamento,

15 Nota pública aprovada na 11ª Reunião Ordinária de 2025 do CGI.br, realizada em 14 de novembro de 2025, em que o Comitê manifestou apoio à regulação da IA no Brasil e recomendou que a regulação observe, a classificação de riscos, os mecanismos de transparência e explicabilidade, a participação multissetorial e uma arquitetura regulatória policêntrica com coordenação da ANPD no Sistema Nacional de Regulação e Governança de Inteligência Artificial.

testagem, implantação, monitoramento, atualização e eventual descontinuidade. Esse deslocamento é fundamental para sistemas de IA generativa, pois muitos riscos não surgem somente no momento em que o usuário recebe uma resposta, mas nas decisões anteriores sobre dados, planejamento do modelo, filtros, parâmetros de segurança, finalidade de uso e distribuição do sistema[16].

O projeto também é importante porque define expressamente categorias que dialogam com o objeto desta pesquisa. O texto conceitua sistema de inteligência artificial como sistema baseado em máquina que, com diferentes graus de autonomia, infere resultados como previsões, conteúdos, recomendações ou decisões capazes de influenciar o ambiente virtual ou real. Além disso, define IA generativa como modelo especificamente destinado a gerar ou modificar significativamente texto, imagens, áudio, vídeo ou código de software. Essa previsão permite tratar juridicamente os modelos generativos como categoria própria, e não apenas como extensão genérica de softwares comuns.

A distinção entre desenvolvedor, distribuidor e aplicador também contribui para a análise da responsabilidade. O texto do PL identifica como desenvolvedor a pessoa natural ou jurídica que desenvolve sistema de IA com vistas à colocação no mercado ou aplicação em serviço sob seu nome ou marca; como distribuidor, quem disponibiliza o sistema a terceiros; e como aplicador, quem emprega ou utiliza o sistema em seu nome ou benefício, inclusive configurando, mantendo ou fornecendo dados para a operação e monitoramento. Essa separação remete diretamente ao problema da cadeia de responsabilidade: em muitos casos, o dano produzido por IA generativa não poderá ser atribuído a um único agente, pois envolve desenvolvedores, plataformas, empresas que integram o modelo em seus serviços e usuários finais.

A governança de riscos aparece no projeto por meio de uma lógica graduada. O PL n.º 2.338/2023 não submete todos os sistemas de IA ao mesmo regime, mas prevê a realização de avaliação preliminar para classificação do grau de risco. Essa avaliação é definida como processo simplificado de autoavaliação, anterior à utilização ou colocação no mercado do sistema, destinado a determinar o cumprimento das obrigações legais aplicáveis. A lógica é

16 Estudo noticiado pelo *The Guardian*, do Centre for Long-Term Resilience, financiado pelo AI Security Institute do governo britânico, identificou cerca de 700 casos reais de *chatbots* e agentes de IA que desconsideravam instruções humana, contornaram salvaguardas e executam ações não autorizadas, apontando um aumento de cinco vezes nos episódios de mau comportamento entre outubro de 2025 e março de 2026, incluindo casos de exclusão não autorizada de e-mails e arquivos. O levantamento, baseado em milhares de interações reais e não apenas em ambiente controlado de laboratório. A reportagem relaciona esses episódios ao debate sobre governança, segurança e acompanhamento internacional de modelos cada vez mais capazes.

simples: quanto maior a possibilidade de afetar direitos fundamentais, maior deve ser o nível de documentação, transparência, controle e supervisão.

A classificação por risco permite diferenciar usos triviais, usos de alto risco e usos considerados excessivos. Essa técnica regulatória é importante porque evita duas soluções igualmente inadequadas. A primeira seria liberar completamente a IA em nome da inovação tecnológica, tratando todos os danos como simples acidentes. A segunda seria impor a todos os sistemas o mesmo peso regulatório, inviabilizando aplicações de baixo impacto e criando barreiras desproporcionais. O modelo de risco busca um ponto intermediário: obrigações proporcionais à gravidade potencial da tecnologia e ao contexto de sua utilização.

Entre os riscos excessivos, o projeto veda o desenvolvimento, a implementação e o uso de sistemas de IA destinados a instigar ou induzir comportamento capaz de causar danos à saúde, à segurança ou a direitos fundamentais, bem como a explorar vulnerabilidades de pessoas ou grupos com o objetivo ou efeito de induzir condutas lesivas. O texto também se preocupa com práticas de avaliação de personalidade, comportamento passado e risco criminal, além de prever restrições a certos usos de material relacionado a abuso e exploração sexual de crianças e adolescentes, entre outras hipóteses. Para o tema deste trabalho, a vedação da exploração de vulnerabilidades é especialmente relevante, pois sistemas generativos podem ser projetados para persuadir, manipular ou manter usuários engajados em interações emocionalmente sensíveis.

A categoria dos sistemas de alto risco, por sua vez, permite construir um regime de deveres reforçados. Embora nem toda IA generativa seja automaticamente de alto risco, determinados usos podem produzir efeitos relevantes sobre direitos, reputação, acesso a serviços, proteção de dados, informação pública e relações de consumo. O risco não está apenas na tecnologia abstrata, mas no contexto em que ela é aplicada. Um modelo generativo usado para entretenimento possui impacto distinto daquele integrado a serviços jurídicos, educacionais, financeiros, administrativos, eleitorais, médicos ou de atendimento psicológico. Por isso, o Direito Digital precisa considerar a finalidade, o público afetado, a vulnerabilidade dos usuários e a possibilidade de dano coletivo.

A avaliação de impacto algorítmico é um dos instrumentos centrais desse modelo. O Projeto de Lei n.º 2.338/2023 define essa avaliação como análise dos impactos do sistema de inteligência artificial sobre direitos fundamentais, com indicação de medidas preventivas, mitigadoras e de reversão dos impactos negativos, bem como de medidas potencializadoras dos impactos positivos. Em sistemas de alto risco, a avaliação torna-se obrigação do desenvolvedor

ou do aplicador que introduzir ou colocar o sistema em circulação no mercado, considerando o papel e a participação do agente na cadeia.

Essa avaliação possui relevância direta para a responsabilidade civil. Ao exigir que riscos sejam identificados previamente, documentados e atualizados durante o ciclo de vida do sistema, o ordenamento cria elementos que podem auxiliar a demonstração de culpa, nexo causal e cumprimento ou descumprimento de deveres de cuidado. Se uma empresa conhece riscos previsíveis de alucinação, discriminação, vazamento de dados ou geração de conteúdo ilícito e, ainda assim, não adota medidas de mitigação, a ausência de governança deixa de ser mera falha técnica e passa a ser indício jurídico de uma conduta negligente.

O Projeto de Lei n.º 2.338/2023 também vincula a governança à transparência e à aplicabilidade. A diligência devida e auditabilidade ao longo do ciclo de vida, a confiabilidade e robustez, a proteção de direitos fundamentais, a prestação de contas, a responsabilização, a reparação integral de danos e a prevenção, precaução e mitigação de riscos. Essa combinação é relevante porque mostra que o futuro da regulação não se limita à indenização posterior. O sistema normativo busca criar obrigações de organização interna, gestão preventiva e documentação capaz de tornar a tecnologia juridicamente controlável.

No campo da IA generativa, o Projeto de Lei n.º 2.338/2023 dedica atenção específica aos sistemas de propósito geral e generativos. O desenvolvedor desses sistemas deve realizar documentação pertinente sobre o desenvolvimento e avaliação preliminar para identificar níveis esperados de risco, inclusive risco estrutural. Quando houver risco, o texto exige, antes da disponibilização ou introdução no mercado para fins comerciais, a descrição do modelo, a documentação dos testes e análises realizados para identificar e gerenciar riscos razoavelmente previsíveis e a documentação dos riscos não mitigáveis remanescentes. Essa previsão é particularmente importante para grandes modelos generativos, que podem ser integrados a múltiplos produtos e gerar impactos em escala.

A noção de risco aproxima o PL n.º 2.338/2023 dos problemas analisados no capítulo anterior. Casos envolvendo *chatbots*, *deepfakes*, perfis falsos, conteúdos discriminatórios e manipulação informacional demonstram que o dano da IA não se limita necessariamente a uma relação individual entre usuário e fornecedor. Um erro produzido por sistema generativo pode ser replicado por milhares de usuários, incorporado por mecanismos de busca, transformado em conteúdo audiovisual, impulsionado por plataformas e utilizado para afetar a reputação de pessoas, grupos ou instituições. A escala de circulação transforma a falha técnica em risco social.

Por isso, a governança de riscos deve ser compreendida como ponte entre responsabilidade civil e regulação administrativa. O Direito Civil continua necessário para reparar a vítima, mas a regulação exige que empresas adotem medidas anteriores ao dano. Essa articulação aparece também na experiência da ANPD com o *sandbox* regulatório em inteligência artificial. A Agência publicou, em 2026, o primeiro relatório parcial de monitoramento do projeto-piloto, acompanhando três empresas de tecnologia em ambiente regulado, experimental, controlado e supervisionado, com foco em desafios tecnológicos, jurídicos e operacionais, além de práticas de governança, segurança, transparência e proteção de dados pessoais.

O *sandbox* regulatório ilustra uma tendência importante: a regulação da IA não pode ser apenas abstrata. Ela precisa testar hipóteses, acompanhar soluções concretas, mapear riscos e produzir evidências. A ANPD (2026) registrou que os próximos ciclos do projeto aprofundaram a análise técnica e regulatória em segurança digital, transparência, governança de dados, anonimização e produção de evidências. Esses elementos dialogam com a dificuldade probatória estudada no capítulo anterior, pois a produção de evidências técnicas pode auxiliar tanto autoridades reguladoras quanto vítimas em eventuais ações de responsabilização.

A adoção de mecanismos de governança também deve ser relacionada à proteção de dados pessoais. Em sistemas de IA generativa, a coleta, tratamento, inferência e reutilização de dados podem ocorrer em diferentes momentos: treinamento do modelo, ajuste fino, interação com usuários, armazenamento de comandos, geração de perfis, avaliação de desempenho e moderação automatizada. Por isso, a LGPD permanece como norma estruturante, especialmente quanto aos princípios da finalidade, adequação, necessidade, transparência, segurança, prevenção, não discriminação e responsabilização. A regulação específica da IA não substitui a proteção de dados, mas amplia o campo de deveres das empresas diante de sistemas capazes de inferir e produzir informações a partir de grandes bases de dados.

Rebehy e Alliceda (2025) também contribuíram para essa leitura ao relacionar IA, LGPD e prova digital. Os autores observam que agentes de tratamento devem adotar medidas de segurança, técnicas e administrativas aptas a proteger dados pessoais, e que a utilização de IA em contextos de tratamento de dados exige atenção à estrutura de governança, à documentação e à demonstração de conformidade. Essa perspectiva reforça a ideia de que a proteção jurídica não se limita à existência formal de consentimento ou termos de uso, exigindo capacidade efetiva de comprovar segurança, finalidade, necessidade e respeito aos direitos dos titulares.

Apesar de seu avanço, o PL n.º 2.338/2023 não elimina todos os problemas[17]. A efetividade do modelo dependerá da definição da autoridade competente, da articulação com autoridades, da capacidade técnica de fiscalização, da participação social, da proteção de segredos comerciais sem ocultar informações essenciais e da criação de critérios compreensíveis para avaliação de impacto algorítmico. Se a avaliação de risco se transformar em documento meramente formal, produzido pela própria empresa sem auditoria suficiente, a governança poderá se converter em aparência de conformidade. A regulação, portanto, precisará equilibrar flexibilidade tecnológica e exigibilidade jurídica.

Também é necessário evitar que a governança seja usada como estratégia de autoproteção empresarial. Relatórios, políticas internas e termos de uso não podem servir apenas para transferir integralmente ao usuário a responsabilidade por erros do sistema. O dever de informação deve ser acompanhado de medidas reais de segurança, supervisão, rastreabilidade e resposta a danos. Em outras palavras, não basta avisar genericamente que o sistema pode errar; é preciso adotar mecanismos para reduzir a frequência, a gravidade e a previsibilidade desses erros, especialmente quando o modelo é explorado economicamente e disponibilizado em larga escala.

Assim, o Marco Legal da IA, se aprovado, poderá representar importante mudança na lógica de responsabilização. O foco deixará de ser exclusivamente a reparação posterior para incluir deveres prévios de governança e gestão de riscos. Para o presente TCC, essa mudança é decisiva: quanto mais o ordenamento exigir documentação, avaliação de impacto, transparência e monitoramento, menor será o espaço para que empresas desenvolvedoras aleguem desconhecimento absoluto sobre as falhas de seus sistemas. A opacidade algorítmica continuará sendo desafio técnico, mas não poderá ser tratada como salvo-conduto jurídico.

5.2 O Projeto de Lei das *Fake News* (PL n.º 2.630/2020) e o dever de transparência das plataformas

O segundo eixo regulatório relevante é o Projeto de Lei n.º 2.630/2020, conhecido como PL das Fake News. Embora tenha sido apresentado antes da popularização massiva da IA

17 Em sentido crítico, Gomes (2026) observa que, diante da velocidade das transformações tecnológicas, há risco de que normas sobre inteligência artificial "nasçam velhas", apontando o PL n.º 2.338/2023 como proposta relevante, mas sujeita ao desafio de acompanhar a rápida evolução dos sistemas de IA.

generativa, o projeto permanece importante porque enfrenta problemas estruturais do ambiente digital: contas inautênticas, contas automatizadas, redes artificiais de distribuição, impulsionamento pago, transparência publicitária, moderação de conteúdo e deveres de informação das plataformas. Em julho de 2026, a ficha da Câmara dos Deputados indicava que o projeto estava pronto para pauta no Plenário e submetido ao regime de urgência.

O PL n.º 2.630/2020 institui a chamada Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. O art. 1º estabelece normas, diretrizes e mecanismos de transparência para provedores de redes sociais e serviços de mensageria privada, com a finalidade de garantir segurança e ampla liberdade de expressão, comunicação e manifestação do pensamento. Essa formulação é relevante porque demonstra que o projeto não se limita ao combate genérico às notícias falsas. Seu objetivo é estruturar deveres de transparência e responsabilidade no sistema informacional digital.

A transparência, nesse contexto, não significa que o Estado ou as plataformas devam definir previamente uma verdade oficial. O risco de censura e de remoção abusiva de conteúdos é real e deve ser considerado. A regulação da desinformação deve preservar a liberdade de expressão, a crítica política, a sátira, a manifestação artística e a pluralidade de opiniões. O problema jurídico surge quando a circulação informacional deixa de ser fruto de participação legítima e passa a ser artificialmente impulsionada por redes coordenadas, contas falsas, mecanismos automatizados ou publicidade confusa. É nesse ponto que o dever de transparência se torna instrumento de proteção democrática.

O Projeto de Lei n.º 2.630/2020 define conta inautêntica como aquela criada ou usada para assumir ou simular identidade de terceiros com o objetivo de enganar o público, ressalvadas hipóteses legítimas como nome social, pseudonímia, humor ou paródia. Define também rede de distribuição artificial como comportamento coordenado e articulado por contas automatizadas ou tecnologia não autorizada, com a finalidade de impactar artificialmente a distribuição de conteúdo. Já a conta automatizada é aquela predominantemente gerida por programa de computador ou tecnologia capaz de simular ou substituir atividades humanas na distribuição de conteúdo. Essas categorias são úteis para a análise de danos envolvendo IA, porque sistemas generativos podem alimentar e potencializar esse tipo de circulação artificial.

Com a IA generativa, o problema deixa de ser apenas a existência de perfis falsos ou robôs de distribuição. O conteúdo produzido por esses mecanismos passa a ter qualidade linguística, visual e sonora muito superior, tornando mais difícil sua detecção por usuários comuns. Um sistema pode gerar textos personalizados para públicos diferentes, imagens falsas de alta veracidade, vídeos manipulados, áudios com clonagem de voz e respostas automatizadas

capazes de simular conversas humanas. A rede artificial, portanto, deixa de ser apenas mecanismo de amplificação e passa também a ser mecanismo de produção de realidade aparente.

O art. 4º do PL n.º 2.630/2020 apresenta objetivos diretamente relacionados a essa preocupação: fortalecer o processo democrático por meio do combate ao comportamento inautêntico e às redes de distribuição artificial de conteúdo, impedir a censura no ambiente online, buscar maior transparência nas práticas de moderação de conteúdos postados por terceiros e adotar mecanismos de informação sobre conteúdos impulsionados e publicitários. Essa estrutura mostra que a transparência das plataformas é tratada como condição para a liberdade, não como sua negação.

A relação entre transparência e democracia aparece de forma ainda mais clara quando se examina o impulsionamento pago. O projeto exige a identificação de conteúdos impulsionados e publicitários, de modo destacado e mantido inclusive quando o conteúdo for compartilhado, encaminhado ou repassado. Em matéria eleitoral, o PL prevê que provedores que forneçam impulsionamento de propaganda eleitoral ou de conteúdos que mencionem candidato, coligação ou partido disponibilizem ao público o conjunto de anúncios, com dados sobre valor gasto, identificação do anunciante, tempo de veiculação e critérios de segmentação.

Esses mecanismos são particularmente relevantes diante da IA generativa. A possibilidade de produzir milhares de variações de uma mesma mensagem, adaptadas a diferentes grupos de eleitores, consumidores ou comunidades, amplia o risco de manipulação informacional. A transparência sobre quem pagou, quanto pagou, qual público foi alcançado e quais critérios foram utilizados para segmentação torna-se condição mínima para que sociedade, pesquisadores, imprensa, Justiça Eleitoral e órgãos reguladores possam avaliar a legitimidade da circulação do conteúdo.

O dever de transparência também aparece nos relatórios, o art. 13 do PL n.º 2.630/2020 prevê que provedores de redes sociais produzam relatórios trimestrais de transparência, em português, informando procedimentos e decisões de tratamento de conteúdos gerados por terceiros no Brasil, bem como medidas empregadas para o cumprimento da lei. O texto exige dados sobre moderação de contas e conteúdos, metodologia utilizada na detecção de irregularidades, medidas adotadas, contas automatizadas, redes artificiais de distribuição, conteúdos impulsionados e publicidade não identificada.

Esses relatórios dialogam com a *accountability* algorítmica porque permitem transformar decisões invisíveis de moderação e distribuição em dados minimamente avaliáveis. Sem transparência, a sociedade não consegue saber se determinada plataforma remove mais

conteúdos de um grupo político do que de outro, se falha sistematicamente contra discursos de ódio, se beneficia determinados anunciantes, se permite a circulação de redes artificiais ou se aplica suas próprias regras de forma seletiva. A transparência, portanto, não resolve sozinha o problema, mas cria condições para fiscalização pública.

A moderação de conteúdo, por sua vez, deve ser acompanhada de devido processo. O PL n.º 2.630/2020 prevê que os provedores submetidos à lei garantam acesso à informação e liberdade de expressão nos processos de elaboração e aplicação de seus termos de uso e disponibilizem mecanismos de recursos. Em caso de medida aplicada a conteúdo ou conta, o usuário deve ser notificado sobre a fundamentação, o processo de análise e os prazos e procedimentos para contestação, ressalvadas hipóteses de risco de dano imediato, segurança da informação, violação a direitos de crianças e adolescentes, crimes de discriminação e grave comprometimento da estabilidade da aplicação.

Essa previsão é importante porque a contenção de danos não pode gerar outro dano: a remoção arbitrária ou opaca de conteúdos legítimos. Plataformas digitais exercem poder relevante sobre a circulação da informação, e esse poder se intensifica quando sistemas automatizados passam a recomendar, reduzir alcance, remover, rotular ou impulsionar conteúdos. A moderação precisa ser eficiente contra ilícitos, mas também justificável perante os usuários afetados. A ausência de explicação, recurso e revisão reforça a assimetria informacional entre plataformas e cidadãos.

A aproximação entre o PL n.º 2.630/2020 e os Temas 987 e 533 do STF é evidente. No julgamento sobre responsabilidade de plataformas digitais por conteúdos de terceiros, o Supremo reconheceu a inconstitucionalidade parcial e progressiva do art. 19 do Marco Civil da Internet e definiu parâmetros enquanto o Congresso Nacional do Brasil não editar nova legislação. A tese fixada prevê, entre outros pontos, deveres de notificação, devido processo, relatórios anuais de transparência, canais específicos de atendimento e representação no Brasil, além de afastar a responsabilidade objetiva na aplicação da tese.

O diálogo entre STF e Congresso Nacional do Brasil revela que a regulação das plataformas não é apenas questão de política legislativa, mas também de proteção constitucional. O Tribunal reconheceu que o regime anterior, baseado exclusivamente na ordem judicial prévia para responsabilização, não conferia proteção suficiente a direitos fundamentais e à democracia em certas hipóteses. Ao mesmo tempo, preservou a necessidade de cuidado com crimes contra a honra e com provedores neutros de comunicação interpessoal, buscando evitar remoções excessivas. Essa solução indica uma tentativa de equilibrar liberdade de expressão, proteção de direitos e dever de cuidado.

Embora o PL n.º 2.630/2020 não tenha sido elaborado especificamente para IA generativa, ele fornece instrumentos úteis para esse novo contexto. A identificação de contas automatizadas, a vedação de redes artificiais, a transparência de anúncios, a rastreabilidade de impulsionamentos e a publicação de relatórios periódicos tornam-se ainda mais relevantes quando o conteúdo é produzido por modelos generativos. A tecnologia de IA aumenta a escala, a personalização e a aparência de autenticidade da desinformação; por isso, mecanismos de transparência das plataformas passam a ser indispensáveis para detectar e conter danos.

Há, contudo, limites importantes. A regulação da desinformação não deve transformar plataformas em árbitras absolutas da verdade. A proteção da esfera pública exige que críticas, opiniões minoritárias, denúncias, sátiras e manifestações políticas continuem protegidas. O combate à desinformação deve concentrar-se em práticas ilícitas, manipulação coordenada, ocultação de publicidade, falsidade de identidade, redes artificiais, *deepfakes* enganosos e violações a direitos fundamentais. Quando a norma confunde erro, opinião e ilícito, aumenta o risco de censura privada ou estatal.

Outro limite está na privacidade, especialmente em serviços de mensagem. O PL n.º 2.630/2020 tenta preservar o conteúdo das mensagens ao tratar da guarda de registros de encaminhamentos em massa para fins de responsabilização, exigindo ordem judicial para acesso aos registros e observância do Marco Civil da Internet e da LGPD. A tensão é evidente: é preciso coibir disparos coordenados e distribuição artificial de conteúdo ilícito, mas sem violar indevidamente o sigilo das comunicações e a privacidade dos usuários.

No contexto da IA generativa, essa tensão tende a aumentar. Mensagens falsas podem ser produzidas em massa, personalizadas e distribuídas por canais privados, dificultando a fiscalização. Ao mesmo tempo, soluções excessivamente invasivas poderiam comprometer a criptografia, o sigilo e a liberdade comunicativa. Por isso, a regulação futura deve combinar rastreabilidade proporcional, proteção de dados, deveres de transparência publicitária, cooperação com autoridades e preservação de direitos fundamentais.

A transparência das plataformas também deve alcançar os sistemas de recomendação. O problema das *fake news* não está apenas no conteúdo falso, mas na forma como determinados conteúdos são priorizados, impulsionados, recomendados e mantidos em circulação. Di Lauro e Baracuhy (2025) destacam que os algoritmos organizam e filtram dados, controlam fluxos de informação e mapeiam preferências, definindo o que aparece para os usuários e em que momento. Essa mediação algorítmica demonstra que as plataformas não são ambientes neutros de circulação discursiva, mas estruturas de visibilidade e invisibilidade.

A partir dessa perspectiva, o PL das *Fake News* deve ser lido como parte de um movimento mais amplo de responsabilização informacional. Não se trata apenas de remover conteúdos depois que se tornam virais, mas de exigir que as plataformas expliquem e documentem seus processos de distribuição, impulsionamento e moderação. A transparência sobre publicidade e redes artificiais reduz a assimetria informacional e permite que usuários e instituições compreendam melhor os interesses econômicos e técnicos por trás do conteúdo que recebem.

Por isso, o PL n.º 2.630/2020 pode contribuir para a contenção dos danos da IA generativa mesmo antes de uma lei específica de IA ser plenamente implementada. Seu foco em responsabilidade e transparência na internet atinge a infraestrutura por onde conteúdos sintéticos circulam. A IA produz; a plataforma distribui o impulsionamento ampliado; os algoritmos recomendam; os usuários compartilham. A contenção efetiva exige olhar para toda essa cadeia.

5.3 *Accountability* algorítmica: a punição como instrumento de proteção à dignidade humana

A *accountability* algorítmica representa a síntese dos mecanismos analisados neste capítulo. O termo pode ser compreendido como o dever de agentes que desenvolvem, fornecem ou aplicam sistemas automatizados de prestar contas sobre seu funcionamento, seus riscos, seus critérios decisórios e seus impactos. Em matéria de IA generativa, *accountability* não significa apenas explicar tecnicamente um modelo complexo, mas criar condições jurídicas e institucionais para que pessoas afetadas possam compreender, contestar e buscar reparação por danos.

A dificuldade de explicação dos sistemas de IA não elimina o dever de responder por seus efeitos. Como demonstrado anteriormente, a opacidade algorítmica cria obstáculos à compreensão do caminho interno percorrido pelo sistema até a produção de determinado resultado. Beuron e Cristóvam (2025) observam que os algoritmos implementados por IA podem funcionar como verdadeiras caixas-pretas diante da falta de informação e transparência sobre a criação das bases de dados e a forma de processamento. Essa opacidade informacional compromete a compreensão pública dos critérios utilizados e pode afetar direitos fundamentais.

Nesse ponto, a *accountability* atua como resposta jurídica à opacidade. Se a vítima não possui acesso aos dados técnicos, aos parâmetros do modelo, aos registros de funcionamento e às decisões de governança, dificilmente conseguirá demonstrar sozinha a origem do dano. A

prestação de contas desloca parte desse ônus para quem controla a tecnologia, exigindo documentação, logs, relatórios, avaliações de impacto, explicações acessíveis e mecanismos de auditoria. Quanto maior a assimetria informacional, maior deve ser o dever de transparência do agente econômico.

A *accountability* também se relaciona com a dignidade da pessoa humana. Sistemas de IA generativa podem afetar a dignidade quando produzem conteúdo falso sobre uma pessoa, simulam sua voz ou imagem, discriminam grupos vulneráveis, violam sua privacidade, manipulam seu comportamento, dificultam acesso a direitos ou interferem na formação livre de opiniões. Nessas situações, o dano não é apenas patrimonial. Ele alcança dimensões existenciais, como identidade, reputação, autonomia, igualdade e confiança nas instituições.

Por isso, a punição deve ser compreendida como instrumento de proteção da dignidade humana, e não como simples reação punitivista. Sanções civis, administrativas e regulatórias têm função preventiva, pedagógica e reparatória. Elas indicam ao mercado que determinados custos não podem ser externalizados para usuários e vítimas. Se uma empresa lucra com a disponibilização de sistemas capazes de gerar danos em escala, deve internalizar os custos de prevenção, monitoramento, auditoria, correção e reparação. A punição, nesse sentido, induz a governança responsável.

A lógica da punição protetiva aproxima-se da teoria do risco-proveito estudada no capítulo anterior. Aquele que explora economicamente uma atividade e dela obtém vantagens deve suportar os riscos inerentes a essa atividade. No caso da IA generativa, as empresas desenvolvedoras obtêm valor econômico a partir da capacidade de seus modelos de produzir textos, imagens, respostas, recomendações e interações em escala. Quando essa mesma capacidade gera danos previsíveis, a empresa não pode se esconder atrás da autonomia aparente do sistema.

Isso não significa defender responsabilidade automática em todos os casos. A responsabilização deve considerar o dano, o nexo causal, o comportamento do usuário, a previsibilidade do risco, a existência de advertências adequadas, os mecanismos de controle disponíveis, a finalidade do sistema e o papel de cada agente na cadeia. O próprio STF, ao tratar da responsabilidade de plataformas por conteúdo de terceiros, afirmou que não haverá responsabilidade objetiva na aplicação da tese fixada nos Temas 987 e 533. Ainda assim, a ausência de responsabilidade objetiva não elimina deveres de cuidado, transparência, prevenção e resposta.

A *accountability* algorítmica exige, portanto, uma responsabilidade qualificada. Não basta perguntar se a empresa teve intenção de causar dano. É preciso verificar se ela adotou

padrões razoáveis de segurança, se documentou riscos conhecidos, se corrigiu falhas identificadas, se disponibilizou meios de contestação, se monitorou impactos inesperados, se protegeu grupos vulneráveis e atuou de modo proporcional diante de conteúdos lesivos. A culpa, nesses casos, pode aparecer como falha de governança.

A avaliação de impacto algorítmico é um dos principais instrumentos dessa lógica. Quando realizada adequadamente, ela permite mapear riscos antes da disponibilização do sistema e durante seu ciclo de vida. O PL n.º 2.338/2023 prevê que a avaliação de impacto seja processo contínuo para sistemas de alto risco, com atualizações periódicas e publicidade das conclusões, observados os segredos industrial e comercial. Essa publicidade parcial é importante: protege interesses empresariais legítimos, mas impede que todo o funcionamento do sistema seja ocultado sob o argumento de segredo comercial.

A noção de segredo comercial, aliás, não pode ser absolutizada. O desenvolvimento de IA envolve investimentos, propriedade intelectual e vantagens competitivas que merecem proteção. Contudo, quando a opacidade impede a defesa de direitos fundamentais, deve haver mecanismos de equilíbrio, como auditorias independentes, acesso regulatório a documentos, perícias técnicas, relatórios públicos resumidos e preservação de evidências. A proteção empresarial não pode neutralizar o direito da vítima à reparação.

A *accountability* também exige participação institucional. As empresas não devem ser as únicas responsáveis por definir se seus próprios sistemas são seguros. Autoridades reguladoras, Judiciário, Ministério Público, órgãos de defesa do consumidor, ANPD, TSE, pesquisadores independentes e sociedade civil possuem papéis complementares. A complexidade da IA torna inadequado um modelo de autocontrole absoluto. A autorregulação pode contribuir, mas precisa ser acompanhada por fiscalização pública e possibilidade de sanção.

Nesse sentido, a experiência do STF nos Temas 987 e 533 é indicativa de um modelo híbrido. O Tribunal não substituiu o Congresso, mas definiu parâmetros mínimos enquanto não há nova legislação. Exigiu deveres de transparência, canais de atendimento, devido processo e relatórios, além de prever dever de cuidado em hipóteses de crimes graves e circulação massiva de conteúdos ilícitos. Ao mesmo tempo, preservou a necessidade de análise subjetiva da responsabilidade. Esse modelo reforça que a *accountability* não se resume à indenização, mas envolve estrutura institucional de resposta.

A punição, nesse contexto, deve cumprir quatro funções. A primeira é reparatória, voltada à recomposição do dano individual ou coletivo. A segunda é preventiva, destinada a reduzir a probabilidade de novos danos. A terceira é informacional, pois sanções e decisões

públicas revelam padrões de falha e orientam o comportamento do mercado. A quarta é simbólica, pois reafirma que a dignidade humana não pode ser subordinada à eficiência técnica ou ao lucro empresarial.

Em matéria de IA generativa, essas funções precisam ser articuladas. Uma indenização isolada pode reparar parcialmente a vítima, mas não impede que o modelo continue produzindo o mesmo tipo de dano. Por isso, decisões judiciais e administrativas devem poder impor obrigações de fazer, auditorias, correção de sistemas, modificação de políticas, remoção de conteúdos, identificação de *deepfakes*, transparência sobre anúncios e monitoramento de riscos. A responsabilização deve atingir a estrutura que permitiu o dano, e não apenas o resultado final.

Essa lógica é especialmente importante quando se trata de danos coletivos. Conteúdos discriminatórios, discurso de ódio, manipulação eleitoral, vazamento massivo de dados e desinformação em saúde ou segurança pública podem afetar grupos inteiros. Nesses casos, a reparação individual é insuficiente. A *accountability* precisa incorporar instrumentos de tutela coletiva, medidas estruturais e atuação coordenada de autoridades. A dignidade humana, enquanto fundamento constitucional, possui dimensão individual e coletiva.

A doutrina sobre prova digital também reforça a necessidade de mecanismos de controle. Rebehy e Alliceda (2025) destacam que estruturas de controle relacionadas à coleta, armazenamento e uso dos dados e das tecnologias utilizadas para registro do percurso do dado dentro de uma empresa são especialmente relevantes para permitir prova e verificação. Em sistemas de IA, a ausência de registros adequados pode comprometer a demonstração do dano e do nexo causal, aumentando a vulnerabilidade da vítima.

A produção de evidências deve ser pensada desde o desenvolvimento do sistema. Logs de interação, versões do modelo, parâmetros de segurança, bases de treinamento, registros de alterações, relatórios de incidentes e documentação de avaliações de risco podem ser essenciais para compreender como determinado dano ocorreu. Sem esses elementos, a vítima enfrenta uma dupla barreira: sofre o dano e não consegue provar sua origem. A *accountability* algorítmica busca justamente reduzir essa assimetria.

Também é necessário distinguir transparência de exposição total do código. Em muitos casos, a abertura integral do código ou dos dados de treinamento pode ser inviável, perigosa ou juridicamente protegida. Porém, isso não significa ausência de transparência. A transparência pode ser operacionalizada por meio de explicações acessíveis, documentação técnica para autoridades, avaliações independentes, testes de robustez, relatórios de impacto, métricas de desempenho, análise de vieses, protocolos de resposta e canais de contestação. O ponto central é que o sistema seja juridicamente auditável.

A *accountability*, portanto, exige uma cultura de governança. Empresas que desenvolvem IA generativa devem incorporar deveres éticos e jurídicos desde a fase de projeto. A segurança não pode ser tratada como camada posterior adicionada apenas depois de escândalos públicos. A prevenção de danos deve integrar a seleção de dados, o treinamento, a testagem, a filtragem de conteúdo, a avaliação de vieses, a interface com o usuário, os termos de uso, a supervisão humana e os mecanismos de denúncia. A governança deve acompanhar o ciclo de vida do sistema.

A ideia de dignidade humana serve como limite material dessa governança. A pessoa não pode ser reduzida a ponto de dados, alvo de persuasão ou objeto de experimentação algorítmica. Quando um sistema gera conteúdo falso sobre alguém, reproduz estereótipos discriminatórios ou manipula vulnerabilidades emocionais, o dano atinge algo mais profundo do que a mera eficiência informacional. A regulação deve reconhecer que a tecnologia opera sobre pessoas concretas, com histórias, identidades, reputações e direitos.

A punição como instrumento de proteção da dignidade humana deve, portanto, ser proporcional e orientada à correção. Multas, indenizações e sanções administrativas são necessárias, mas não bastam. É preciso exigir mudança de comportamento. Uma empresa que não documenta riscos deve ser obrigada a documentá-los. Uma plataforma que impulsiona conteúdo sem identificação deve ser obrigada a rotulá-lo. Um desenvolvedor que não oferece mecanismo de contestação deve implementá-lo. A sanção deve induzir reorganização institucional.

Por outro lado, a regulação não deve ignorar a importância da inovação. Sistemas de IA podem contribuir para educação, saúde, acessibilidade, pesquisa científica, produtividade, administração pública e inclusão. O objetivo da *accountability* não é impedir a tecnologia, mas condicionar seu uso ao respeito a direitos fundamentais. Quanto mais confiáveis forem os sistemas, maior será sua aceitação social. A proteção da dignidade humana e a inovação responsável não são objetivos opostos; são condições recíprocas de legitimidade.

A partir dessa leitura, o futuro da regulação brasileira deve caminhar para um modelo integrado. O PL n.º 2.338/2023 fornece estrutura específica para sistemas de IA, com avaliação de risco, avaliação de impacto, governança e direitos das pessoas afetadas. O PL n.º 2.630/2020 fornece instrumentos de transparência e responsabilidade para plataformas digitais, especialmente quanto à circulação de conteúdos e redes artificiais. A LGPD protege dados pessoais e estabelece princípios de segurança, prevenção e responsabilização. O Marco Civil da Internet continua relevante para liberdade de expressão, privacidade, guarda de registros e

responsabilidade de provedores. A jurisprudência do STF complementa esse quadro ao reinterpretar o art. 19 diante da proteção insuficiente a direitos fundamentais.

O desafio está em articular esses regimes sem contradições. Se cada norma tratar isoladamente de IA, plataformas, dados, eleições e consumo, o resultado poderá ser fragmentado e difícil de aplicar. A *accountability* algorítmica serve como eixo de integração, pois impõe uma pergunta comum a todos esses campos: quem controla o sistema, quais riscos conhece, quais medidas adotou, que informações forneceu, como permite contestação e de que modo responde pelo dano?

Assim, a punição jurídica em matéria de IA não deve ser vista apenas como resposta posterior ao ilícito, mas como mecanismo de estruturação do futuro tecnológico. Ao responsabilizar empresas que falham em prevenir riscos previsíveis, o Direito incentiva modelos mais seguros. Ao exigir transparência, reduz a opacidade. Ao impor reparação, reconhece a vítima. Ao proteger a dignidade humana, reafirma que nenhuma inovação pode justificar a transformação da pessoa em objeto descartável de experimentação algorítmica.

6 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo analisar a responsabilidade civil das empresas desenvolvedoras de sistemas de inteligência artificial generativa diante de erros, falhas e danos

causados a usuários e terceiros no ordenamento jurídico brasileiro. A pesquisa partiu do questionamento sobre de que forma o Direito brasileiro pode responsabilizar empresas desenvolvedoras quando seus sistemas produzem conteúdos falsos, ofensivos, discriminatórios, manipulados ou capazes de gerar prejuízos concretos.

No primeiro momento, verificou-se que a evolução dos meios de comunicação modificou profundamente a forma de produção e circulação da informação. O modelo tradicional, marcado por filtros humanos e por uma cadeia de produção mais delimitada, foi alterado pela internet, pelas redes sociais, pelos mecanismos de busca e pelos sistemas automatizados de recomendação. Com isso, a informação passou a circular de maneira descentralizada, personalizada e mediada por algoritmos, dificultando a identificação da autoria, do alcance dos conteúdos e da responsabilidade pelos danos produzidos no ambiente digital.

Em seguida, constatou-se que a inteligência artificial generativa intensifica esses desafios, pois não se limita a organizar ou recomendar informações, mas também cria textos, imagens, áudios, vídeos e respostas automatizadas com aparência de veracidade. Essa capacidade amplia riscos relacionados à desinformação, *deepfakes*, discursos discriminatórios, violações de direitos da personalidade, manipulação de usuários e indução a comportamentos prejudiciais. Por isso, os erros de inteligência artificial não podem ser compreendidos como falhas meramente neutras, já que podem atingir direitos fundamentais protegidos.

No campo da responsabilidade civil, concluiu-se que o ordenamento jurídico brasileiro já possui instrumentos relevantes para enfrentar esses danos, especialmente por meio dos conceitos de ato ilícito, abuso de direito, dever de reparação, responsabilidade objetiva e teoria do risco-proveito. Quando empresas desenvolvem, disponibilizam e exploram economicamente sistemas de inteligência artificial, assumem deveres de cuidado proporcionais aos riscos criados por sua atividade. Assim, a opacidade dos algoritmos, a complexidade técnica dos sistemas e a multiplicidade de agentes envolvidos não devem servir, por si só, para afastar a responsabilidade, sobretudo quando houver falha de prevenção, ausência de testes adequados, falta de transparência ou omissão diante de riscos previsíveis.

A análise do Marco Civil da Internet e da recente interpretação do Supremo Tribunal Federal demonstrou que o regime de responsabilidade das plataformas digitais está em transformação. Embora o art. 19 tenha sido concebido em contexto voltado principalmente aos conteúdos gerados por terceiros, os debates atuais apontam para a necessidade de maior atenção aos deveres de cuidado das plataformas, especialmente em situações de circulação massiva de conteúdos ilícitos, impulsionamento, redes artificiais de distribuição e riscos sistêmicos. Essa

discussão é relevante para a inteligência artificial generativa, pois muitos desses sistemas são integrados a plataformas digitais de grande alcance.

Os casos examinados, como Grok, Microsoft Tay, Chai App “Eliza”, *deepfakes* e uso de inteligência artificial em contextos eleitorais e processuais, demonstraram que os danos provocados por sistemas de IA podem assumir formas variadas. Em alguns casos, o problema está na produção de discurso discriminatório; em outros, na geração de informação falsa, na aparência enganosa de empatia, na manipulação de imagem e voz ou na apresentação de conteúdo inexistente como verdadeiro. Ainda que nem todos representam decisões judiciais, esses episódios evidenciam a importância de mecanismos preventivos, de governança e de responsabilização compatíveis com a complexidade tecnológica atual.

Por fim, conclui-se que o Direito brasileiro já oferece bases importantes para responsabilizar empresas desenvolvedoras de inteligência artificial, especialmente por meio da responsabilidade civil, da proteção dos direitos da personalidade, da defesa do consumidor, da proteção de dados e dos deveres gerais de cuidado. Contudo, esses instrumentos precisam ser interpretados à luz das características próprias da inteligência artificial generativa, como opacidade, escala, autonomia operacional e potencial de dano sistêmico. Dessa forma, empresas que desenvolvem e exploram economicamente sistemas de IA não podem se eximir de responsabilidade sob o argumento de imprevisibilidade tecnológica quando deixam de adotar medidas adequadas de prevenção, controle, transparência e mitigação de riscos. A proteção da dignidade humana, da confiança informacional e dos direitos fundamentais deve orientar tanto a aplicação das normas existentes quanto a construção de uma regulação específica para a inteligência artificial no Brasil.

REFERÊNCIAS

AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **Publicados primeiros resultados do Sandbox Regulatório em Inteligência Artificial.** Brasília, DF: ANPD, 2 jul. 2026. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/publicados-primeiros-resultados-do-sandbox-regulatorio-em-inteligencia-artificial. Acesso em: 14 ago. 2026.

ANUNCIAÇÃO, Clodoaldo Silva da *et al.* **Direito, inteligência artificial e algoritmos**: desafios jurídicos na era da tecnologia extrema. 1. ed. Rio de Janeiro: Almedina Brasil, 2025. *E-book*. Disponível em: https://integrada.minhabiblioteca.com.br/reader/books/9788584939404/epubcfi/6/2[%3Bvnd.vst.idref%3Dcover.xhtml]!/4/2[cover-image]/2%4050:76. Acesso em: 14 ago. 2026.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). **Inteligência artificial generativa.** Radar Tecnológico, n. 3. 1. ed. Brasília, DF: ANPD, nov. 2024. Disponível em: https://www.gov.br/anpd/pt-br/centrais-de-conteudo/documentos-tecnicos-orientativos/radar_tecnologico_ia_generativa_anpd.pdf/@@display-file/file. Acesso em: 14 ago. 2026.

BARCAROLLO, Felipe. **Inteligência artificial**: aspectos ético-jurídicos. 1. ed. São Paulo: Almedina, 2021. *E-book*. Disponível em: https://integrada.minhabiblioteca.com.br/reader/books/9786556272351. Acesso em: 21 mai. 2026.

BAWDEN, David; ROBINSON, Lyn. The dark side of information: overload, anxiety and other paradoxes and pathologies. **Journal of Information Science**, v. 35, n. 2, p. 180-191, 2009. Disponível em: https://journals.sagepub.com/doi/10.1177/0165551508095781. Acesso em: 14 ago. 2026.

BHARADE, Aditi. A widow is accusing an AI chatbot of being a reason her husband killed himself. **Business Insider**, 4 abr. 2023. Disponível em: https://www.businessinsider.com/widow-accuses-ai-chatbot-reason-husband-kill-himself-2023-4. Acesso em: 14 ago. 2026.

BOOTH, Robert. Number of AI chatbots ignoring human instructions increasing, study says. **The Guardian**, 27 mar. 2026. Disponível em: https://www.theguardian.com/technology/2026/mar/27/number-of-ai-chatbots-ignoring-human-instructions-increasing-study-says. Acesso em: 14 ago. 2026.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 14 ago. 2026.

BRASIL. **Lei n.º 10.406, de 10 de janeiro de 2002**. Institui o Código Civil. Brasília, DF: Presidência da República, 2002. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2002/l10406compilada.htm. Acesso em: 14 ago. 2026.

BRASIL. **Lei n.º 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. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 14 ago. 2026.

BRASIL. **Lei n.º 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em:

https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 14 ago. 2026.

BRASIL. **Lei n.º 8.078, de 11 de setembro de 1990**. Dispõe sobre a proteção do consumidor e dá outras providências. Brasília, DF: Presidência da República, 1990. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm. Acesso em: 14 ago. 2026.

BRASIL. Senado Federal. **Projeto de Lei n.º 2.338, de 2023.** Dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. Brasília, DF: Senado Federal, 2025. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2868197&filename=PL%202338/2023. Acesso em: 14 ago. 2026.

BRASIL. Senado Federal. **Projeto de Lei n.º 2.630, de 2020**. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Brasília, DF: Senado Federal, 2020. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1909983&filename=PL%202630/2020. Acesso em: 14 ago. 2026.

BRASIL. Tribunal Superior Eleitoral (TSE). **Resolução n.º 23.732, de 27 de fevereiro de 2024**. Altera a Res.-TSE nº 23.610, de 18 de dezembro de 2019, dispondo sobre a propaganda eleitoral. Brasília, DF: TSE, 2024. Disponível em: https://www.tse.jus.br/legislacao/compilada/res/2024/resolucao-no-23-732-de-27-de-fevereiro-de-2024. Acesso em: 14 ago. 2026.

BRASIL. Tribunal Superior Eleitoral (TSE). **Resolução n.º 23.755, de 2 de março de 2026.** Altera a Resolução nº 23.610/TSE, de 18 de dezembro de 2019, que dispõe sobre propaganda eleitoral. Brasília, DF: TSE, 2026. Disponível em: https://www.tse.jus.br/legislacao/compilada/res/2026/resolucao-no-23-755-de-2-de-marco-de-2026. Acesso em: 14 ago. 2026.

BRUNS, Axel. Gatekeeping, gatewatching, real-time feedback: new challenges for Journalism. **Brazilian Journalism Research**, Brasília-DF, v. 7, n. 2, p. 117-136, 2011. Disponível em: https://bjr.sbpjor.org.br/bjr/article/view/355/331. Acesso em: 14 ago. 2026.

BRUNS, Axel. Gatewatching, not gatekeeping: collaborative online news. **Sage Journals**, v. 107, p. 31-44, 2003. Disponível em: https://snurb.info/files/Gatewatching, %20Not%20Gatekeeping.pdf. Acesso em: 14 ago. 2026.

CASTELLS, Manuel. **A galáxia da internet**: reflexões sobre a internet, os negócios e a sociedade. Tradução de Maria Luiza X. de A. Borges. Rio de Janeiro: Zahar, 2003. Disponível em: https://pt.scribd.com/document/666764828/A-Galaxia-Da-Internet-Manuel-Castells. Acesso em: 14 ago. 2026.

CHESNEY, Bobby; CITRON, Danielle. Deep Fakes: a looming challenge for privacy, democracy, and national security. **California Law Review**, v. 107, p. 1752-1820, 2019. Disponível em: https://scholarship.law.bu.edu/faculty_scholarship/640/?utm_source=chatgpt.com. Acesso em: 14 ago. 2026.

COMITÊ GESTOR DA INTERNET NO BRASIL (CGI.br). **CGI.br elabora proposta de tipologia para diferenciar provedores de aplicações Internet**. São Paulo: CGI.br, 17 mar. 2025. Disponível em: https://cgi.br/noticia/releases/cgi-br-elabora-proposta-de-tipologia-para-

diferenciar-provedores-de-aplicacoes-internet/?utm_source=chatgpt.com. Acesso em: 14 ago. 2026.

CORMODE, Graham; KRISHNAMURTHY, Balachander. Key differences between Web 1.0 and Web 2.0. **First Monday**, v. 13, n. 6, 2 jun. 2008. Disponível em: https://firstmonday.org/ojs/index.php/fm/article/view/2125/1972. Acesso em: 14 ago. 2026.

COSTA, Márcio José Magalhães. **Responsabilidade civil e inteligência artificial.** São Paulo: Almedina, 2021.

DI LAURO, Laís Sousa; BARACUHY, Regina. Corpos à margem no Instagram: (in)visibilidade, controle e vigilância no dispositivo algorítmico. **Revista Heterotópica**, Uberlândia, v. 7, n. 1, p. 104-127, jan./jun. 2025. Disponível em: https://seer.ufu.br/index.php/RevistaHeterotopica/article/view/77084/42006. Acesso em: 14 ago. 2026

DOBROW, Stephen B. Rise of Internet and World Wide Web. **EBSCO**, 2021. Disponível em: https://www.ebsco.com/research-starters/history/rise-internet-and-world-wide-web. Acesso em: 14 ago. 2026.

EISENSTEIN, Elizabeth L. **The printing revolution in early modern Europe**. 2. ed. Cambridge: Cambridge University Press, 2013.

EL ATILLAH, Imane. Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change. **Euronews**, 31 mar. 2023. Disponível em: https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-. Acesso em: 14 ago. 2026.

FORBES TECH. Elite da IA brasileira: veja as 10 startups prontas para captações de até US$ 100 milhões em 2026. **Forbes Brasil**, 6 jan. 2026. Disponível em: https://forbes.com.br/forbes-tech/2026/01/elite-da-ia-brasileira-veja-as-10-startups-prontas-para-captacoes-de-ate-us-100-milhoes-em-2026/. Acesso em: 14 ago. 2026.

GOMES, Samuel. O espantalho do Vale do Silício e o Marco Legal da IA: do PL 2.338/23. **Consultor Jurídico** (Conjur), 28 abr. 2026. Disponível em: https://conjur.com.br/2026-abr-28/o-espantalho-do-vale-do-silicio-e-o-marco-legal-da-ia-do-pl-2-338-23-2/. Acesso em: 14 ago. 2026.

LEE, Peter. Learning from Tay's introduction. **Official Microsoft Blog**, 25 mar. 2016. Disponível em: https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/. Acesso em: 14 ago. 2026.
LUZ, Daniel. Top 10 maiores empresas de IA do mundo em 2026. **beAnalytic**, 25 jun. 2026. Disponível em: https://beanalytic.com.br/blog/maiores-empresas-de-inteligencia-artificial/. Acesso em: 14 ago. 2026.

MONTGOMERY, Blake. Mother says AI chatbot led her son to kill himself in lawsuit against its maker. **The Guardian**, 23 out. 2024. Disponível em: https://www.theguardian.com/technology/2024/oct/23/character-ai-chatbot-sewell-setzer-death. Acesso em: 14 ago. 2026.

OPENAI. Conheça o ChatGPT. **OpenAI**, 30 nov. 2022. Disponível em: https://openai.com/pt-BR/index/chatgpt/. Acesso em: 14 ago. 2026.

O'REILLY, Tim. What is Web 2.0: design patterns and business models for the next generation of software. **O'Reilly**, 30 set. 2005. Disponível em: https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1. Acesso em: 14 ago. 2026.

PICHAI, Sundar; HASSABIS, Demis. Introducing Gemini: our largest and most capable AI model. **News from Google**, 6 dez. 2023. Disponível em: https://blog.google/innovation-and-ai/technology/ai/google-gemini-ai/. Acesso em: 14 ago. 2026.

SUPREMO TRIBUNAL FEDERAL (STF). Plataformas terão 60 dias para implementar medidas estruturais, decide STF. **STF Notícias**, Brasília-DF, 17 jun. 2026. Disponível em: https://noticias.stf.jus.br/postsnoticias/plataformas-terao-60-dias-para-implementar-medidas-estruturais-decide-stf/. Acesso em: 14 ago. 2026.

TRIBUNAL SUPERIOR DO TRABALHO (TST). Empresa e advogado são condenados por possível uso de IA com citações falsas de jurisprudência. **Notícias do TST**, Brasília, DF: TST, 10 mar. 2026. Disponível em: https://www.tst.jus.br/-/empresa-e-advogado-sao-condenados-por-possivel-uso-de-ia-com-citacoes-falsas-de-jurisprudencia. Acesso em: 14 ago. 2026.

TRIBUNAL SUPERIOR ELEITORAL (TSE). Por Dentro das Eleições: conheça as regras sobre uso de IA na campanha eleitoral de 2026. **Notícias do TSE**, Brasília, DF: TSE, 10 abr. 2026. Disponível em: https://www.tse.jus.br/comunicacao/noticias/2026/Abril/por-dentro-das-eleicoes-conheca-as-regras-sobre-uso-de-ia-na-campanha-eleitoral-de-2026. Acesso em: 14 ago. 2026.

YANG, Maya. Elon Musk's AI firm apologizes after chatbot Grok praises Hitler. **The Guardian**, 12 jul. 2025. Disponível em: https://www.theguardian.com/us-news/2025/jul/12/elon-musk-grok-antisemitic. Acesso em: 14 ago. 2026.