doi: 10.21783/rei.v12i2.1033

SUBMETIDO: 20 MAIO 2026
ACEITO: 20 MAIO 2026

INTELIGÊNCIA ARTIFICIAL E VIOLÊNCIA SEXUAL DIGITAL: DEEPFAKES, FALSOS NUDES E RESPOSTAS INSTITUCIONAIS NO BRASIL

ARTIFICIAL INTELLIGENCE AND DIGITAL SEXUAL VIOLENCE: DEEPFAKES, FAKE NUDES, AND INSTITUTIONAL RESPONSES IN BRAZIL

MAÍRA VILLELA ALMEIDA[1]
ALEXANDER ALI SHAH[2]

RESUMO: A expansão da inteligência artificial generativa, incorporada às plataformas digitais, vem reconfigurando a violência sexual mediada por tecnologia ao permitir a produção e circulação massiva de conteúdos íntimos falsos, como deepfakes sexuais e "falsos nudes". Este artigo sustenta que a IA, para além de ferramenta técnica, opera como instituição social: estrutura práticas, incentivos e riscos, produzindo danos previsíveis e sistemáticos, especialmente contra mulheres e meninas. A partir desse enquadramento institucional, discute-se a insuficiência das categorias jurídicas tradicionais, em particular o consentimento, que pressupõe manifestação de vontade humana livre e anterior — elemento ausente na manipulação algorítmica de imagens e vídeos. Em seguida, examinam-se limites das respostas jurídicas no direito brasileiro, com foco na tutela constitucional da dignidade, honra e imagem; no regime do Marco Civil da Internet; na LGPD; em alterações penais recentes e na Lei nº 15.211/2025 (ECA Digital), destacando lacunas normativas, dificuldades probatórias e a inadequação de uma abordagem exclusivamente reativa. Por fim, propõe-se uma leitura orientada por arranjos jurídico-institucionais, defendendo deveres preventivos e contínuos de diligência, mecanismos de governança e mitigação de riscos "by design", articulação interinstitucional e políticas públicas capazes de enfrentar a natureza massiva, transnacional e automatizada da violência sexual digital.

PALAVRAS-CHAVE: Direitos Fundamentais; Novos Direitos; Inteligência artificial; *deepfakes* sexuais; violência digital de gênero.

[1] Doutora e mestra em Teorias Jurídicas Contemporâneas pela UFRJ. *Visiting Researcher* pela *Harvard Law School*, com apoio da Comissão *Fulbright*. Professora permanente do mestrado e doutorado em Direito da Universidade Estácio de Sá, onde também é pesquisadora e bolsista produtividade. Advogada. Pesquisadora da FGV Justiça.

[2] Procurador Federal. Mestre em Direitos Fundamentais e Novos Direitos - UNESA, Pós-graduado *lato sensu* em Direito Eletrônico – UNESA, pós-graduado *lato sensu* em Análise de Sistemas – UVA, Bacharel em Direito – UERJ, Tecnólogo em Processamento de Dados - FIAA.

ABSTRACT: The fast spread of generative AI embedded in digital platforms has reshaped technology-facilitated sexual violence by enabling the large-scale creation and circulation of synthetic intimate content, including sexual deepfakes and "fake nudes". This article argues that AI should be understood not merely as a technical tool but as a social institution: it structures practices, incentives, and risks, producing foreseeable and systemic harms, disproportionately affecting women and girls. From this institutional lens, the paper highlights the limits of traditional legal categories—especially consent, which presupposes a free, prior human manifestation of will—an element absent in algorithmic manipulation of images and videos. The analysis then examines shortcomings of Brazilian legal responses, focusing on constitutional protections of dignity, honor, and image; the liability framework under the Marco Civil da Internet; Brazil's data protection law (LGPD); recent criminal law amendments; and Law No. 15.211/2025 (Digital ECA), emphasizing regulatory gaps, evidentiary challenges, and the inadequacy of purely reactive enforcement. Finally, the article advances an institutional-arrangements approach, calling for preventive and ongoing duties of due diligence, risk-mitigation "by design", coordinated institutional action, and public policies capable of addressing the massive, transnational, and automated nature of digital sexual violence.

KEYWORDS: Fundamental Rights; New Rights; Artificial Intelligence; Sexual Deepfakes; Gender-Based Digital Violence.

INTRODUÇÃO

A inteligência artificial (IA) generativa não chegou como uma nova ferramenta neutra, "de laboratório". Ela entrou no mundo real acoplada a plataformas, mercados de atenção e economias de dados. E, quando uma tecnologia entra assim — conectada a incentivos de escala, velocidade e viralidade — ela não apenas *faz coisas*: ela organiza práticas, redistribui poder, cria zonas de impunidade e inaugura novas formas de dano. Nesse terreno, a violência sexual mediada por tecnologia ganhou uma mutação crucial: a capacidade de produzir, com realismo crescente e custo decrescente, conteúdos íntimos sintéticos atribuídos a pessoas reais (deepfakes sexuais, "falsos nudes", vídeos pornográficos fabricados), e de fazê-lo

em escala massiva, replicável, transnacional e frequentemente sem autoria identificável.

A novidade não está apenas na falsidade do conteúdo. Está na inversão do eixo clássico: antes, a violência sexual digital costumava depender de um "material prévio" (foto íntima obtida por coerção, vazamento, invasão, ou compartilhamento não consentido). Agora, a matéria-prima pode ser banal e pública: uma foto de rosto em rede social, um vídeo comum, uma imagem capturada em ambiente profissional. A partir daí, modelos generativos produzem um corpo que não existiu — mas um efeito social muito real: humilhação, coerção, chantagem, perseguição, destruição reputacional, silenciamento e retração da presença pública. A vítima sofre não por ter "feito" algo, mas por ser convertida em personagem sexual por uma engrenagem de síntese e disseminação.

Esse deslocamento tensiona profundamente as categorias jurídicas tradicionais. O direito costuma trabalhar com fatos, autoria e causalidade de maneira relativamente linear. Já a violência sexual com IA opera como sistema: é automatizável, modular (criar, editar, remixar, espalhar), difusiva (milhares de replicações), assimétrica (custos baixos para o agressor, custos altíssimos para a vítima) e iterativa (uma remoção não impede novas versões). A resposta jurídica reativa — remover conteúdo e punir indivíduos — permanece necessária, mas se mostra insuficiente diante de um fenômeno que se comporta como infraestrutura, não como evento isolado.

É nesse ponto que este artigo propõe um enquadramento: compreender a IA, sobretudo quando integrada às plataformas digitais, como instituição social — no sentido de um arranjo técnico-econômico que estabiliza expectativas, produz rotinas, cria incentivos e distribui riscos. Instituições, aqui, não são apenas órgãos formais do Estado; são também arquiteturas que disciplinam comportamentos e definem o que "funciona" socialmente. A IA generativa, inserida em ecossistemas de atenção, transforma a sexualização e a violência em "conteúdo" com alta performatividade algorítmica: circula porque engaja, engaja porque excita, excita porque degrada; e degrada porque encontra público e impunidade. A violência, assim, não aparece como exceção: torna-se um resultado previsível de um ambiente desenhado para circulação acelerada e fricção mínima.

O recorte de gênero é central. A literatura internacional e as evidências empíricas disponíveis indicam que *deepfakes* pornográficos e imagens íntimas sintéticas atingem desproporcionalmente mulheres, meninas e pessoas feminilizadas, funcionando como tecnologia de disciplina social: rebaixa, ameaça, coloca em dúvida a credibilidade e produz efeitos de autocensura. A violência sexual digital, nesse sentido, não é apenas um dano individual, mas sim uma técnica de governo dos corpos e das vozes — um modo contemporâneo de administrar a presença feminina no espaço público. Quando o risco de "virar nude" pode recair sobre qualquer foto comum, o custo de existir online muda de patamar:

o espaço público se torna um espaço condicional, onde certas pessoas pagam uma "taxa" permanente de exposição.

Essa realidade coloca em crise um conceito jurídico que, por décadas, consistiu na centralidade ética e dogmática do debate sobre sexualidade: o consentimento. No mundo analógico, consentimento pode ser pensado como manifestação de vontade anterior ao ato, dentro de um horizonte de autonomia individual. No mundo dos *deepfakes*, há um curto-circuito: a vítima não consente porque não há sequer o que consentir previamente; o ato se dá sem presença, sem contato, sem prévia possibilidade de recusa, muitas vezes sem ciência. Insistir em enquadrar o problema apenas como "ausência de consentimento" pode produzir duas consequências ruins: (i) manter o foco excessivo na conduta da vítima (o que ela postou, como se expôs, se "deu margem"), reativando dinâmicas de culpabilização; e (ii) deixar invisível a responsabilidade estrutural dos sistemas que tornam o dano massivo, replicável e lucrativo.

Por isso, a hipótese que orienta este artigo é dupla. Primeiro: a violência sexual digital mediada por IA deve ser compreendida como fenômeno institucional e sistêmico, que exige deveres preventivos e governança de risco, e não apenas repressão posterior a indivíduos. Segundo: o direito brasileiro dispõe de fundamentos robustos para proteger dignidade, honra, imagem, intimidade e autodeterminação sexual, mas precisa enfrentar um desafio específico: traduzir esses fundamentos em arquiteturas regulatórias operáveis para um ambiente de produção automatizada de falsificações íntimas, com altíssima velocidade de circulação.

O cenário regulatório brasileiro apresenta instrumentos relevantes, porém fragmentados. A Constituição oferece o núcleo axiológico (dignidade humana, direitos da personalidade, proteção da intimidade e da imagem). O Código Civil fornece a gramática da responsabilização e da reparação. O Marco Civil da Internet estrutura um regime de responsabilidade e procedimentos de remoção; a LGPD fornece categorias para tratamento de dados pessoais e sensíveis; o direito penal, por sua vez, tem sido acionado para punir condutas relacionadas à divulgação de conteúdo íntimo e a práticas de violência psicológica e perseguição. Mais recentemente, o debate sobre proteção de crianças e adolescentes em ambientes digitais ganhou nova densidade normativa, exigindo respostas específicas quando a vítima é menor de idade — situação em que o dano é ainda mais radical, tanto pelo potencial de perpetuidade quanto pela gravidade do impacto psíquico e social.

Ainda assim, a fricção entre norma e realidade permanece. A prova é complexa, a autoria pode ser ocultada, a replicação é automática e a remoção raramente é definitiva. Além disso, a lógica de "tira do ar e pronto" falha quando a criação é tão barata que o conteúdo reaparece com variações mínimas. O que está em jogo, então, não é apenas responsabilizar depois, mas impedir antes, mitigar durante, e reduzir a capacidade do sistema de transformar violência em circulação.

Nesse ponto, ganha relevo a noção de deveres de diligência e de prevenção contínua: medidas técnicas e procedimentais que plataformas e provedores de serviços devem adotar para identificar, reduzir e responder a riscos previsíveis de abuso. Isso inclui, por exemplo, políticas de detecção e rotulagem, barreiras de geração para conteúdos explícitos não consentidos, protocolos céleres de denúncia e remoção, preservação de evidências para investigação, transparência sobre métricas de disseminação e cooperação com autoridades — tudo isso com atenção a garantias fundamentais (liberdade de expressão, devido processo, proporcionalidade), porque combater violência não pode virar licença para arbitrariedade. A pergunta jurídica contemporânea, portanto, não é só "quem fez?", mas também: quem tinha capacidade de prevenir, quem lucrou com a circulação, quem desenhou o ambiente e quem negligenciou riscos conhecidos?

O objetivo da presente pesquisa consiste em analisar como a IA generativa, ao operar como instituição de produção e amplificação de violência sexual digital, desafia categorias jurídicas centradas no evento individual e no consentimento clássico, e exige respostas regulatórias que combinem tutela de direitos da personalidade, proteção contra discriminação de gênero e governança preventiva de risco em plataformas. Metodologicamente, o texto adota abordagem jurídico-analítica e crítico-institucional: parte do diagnóstico do fenômeno (*deepfakes* sexuais e falsos nudes), identifica tensões dogmáticas (consentimento, autoria, prova, responsabilidade) e discute possibilidades normativas e institucionais no contexto brasileiro, com atenção a experiências comparadas e a parâmetros de direitos humanos.

Em termos de estrutura, após esta introdução o artigo se organiza em seções que: (i) conceituam a IA generativa e descrevem as formas contemporâneas de violência sexual digital; (ii) exploram a ideia de IA como instituição, destacando incentivos, escala e assimetrias; (iii) discutem o colapso parcial do consentimento como categoria suficiente e a necessidade de reposicionar o foco normativo; (iv) analisam os instrumentos do direito brasileiro aplicáveis e suas lacunas; e (v) propõem um horizonte de arranjos jurídico-institucionais com deveres preventivos, coordenação interinstitucional e políticas públicas. Ao final, sustenta-se que enfrentar deepfakes sexuais não é apenas "atualizar a lei"; é recalibrar o direito para um mundo em que a violência pode ser sintetizada, automatizada e distribuída como produto — e em que a proteção da dignidade, da intimidade e da autodeterminação sexual depende tanto de normas quanto de arquiteturas de responsabilidade.

2. INTELIGÊNCIA ARTIFICIAL, INSTITUCIONALIDADE E VIOLÊNCIA SEXUAL

É inegável o fato de que a inteligência artificial vem impactando a sociedade em múltiplas direções, constituindo advento social complexo, que extrapola os estudos da tecnologia em si, impondo o exame de seus efeitos na vida dos indivíduos e dos grupos sociais.

A par das funcionalidades técnicas das ferramentas de IA e de seu potencial para coadjuvar o ser humano na busca por soluções que melhorem a sua condição de vida, emergem consequências adversas, que não podem ser vistas apenas sob o ângulo da colateralidade, por constituírem, de fato, questões centrais, em razão de seu uso como instrumentos para a prática de diversas formas de violência, dentre elas, a de gênero.

Sob esse ângulo, é sabido que a IA está longe de ser uma ferramenta tecnológica neutra (VALENTE, 2023), uma vez que reflete o pensamento e o comportamento humano, multifacetado, ora alinhado às noções básicas da conduta ética, ora enviesado, discriminatório e violador de direitos. De fato, a tecnologia amplifica o alcance e a intensidade dos caracteres humanos, fazendo com que adquiram o potencial de produzir eventos de mais profundas proporções. Por essa razão, a inteligência artificial deve ser tratada como como instituição social, capaz de servir de instrumento para a produção de violência, com consequências estruturais sobre as relações sociais.

Dentre os diversos riscos viabilizados pelos sistemas de IA, destaca-se, para o presente estudo, aqueles relacionados ao campo da sexualidade e da violência de gênero, especialmente a geração massiva de imagens falsas, produzidas em profusão por meio de determinados modelos e arquiteturas tecnológicas, cujos efeitos estruturais potencializam práticas de humilhação, coerção e violação da dignidade humana.

A inteligência artificial, especialmente quando incorporada às grandes plataformas digitais, ultrapassou a condição de mera ferramenta técnica, passando a operar como uma verdadeira instituição social capaz de produzir efeitos jurídicos próprios. Nesse arranjo, os sistemas algorítmicos não apenas mediam interações, mas estruturam práticas, riscos e padrões de comportamento, inclusive no campo da sexualidade. A violência sexual mediada pela tecnologia emerge, assim, como fenômeno previsível das arquiteturas técnicas e dos incentivos econômicos que orientam o desenvolvimento desses sistemas.

Nesse contexto, a proliferação de imagens íntimas falsas geradas por modelos de IA evidencia não um desvio isolado de uso, mas um funcionamento institucional que viabiliza, amplifica e tende a normalizar a violação da sexualidade, especialmente de mulheres e meninas. Casos recentes envolvendo ferramentas tecnológicas amplamente difundidas, como o Grok [3], demonstram como determinadas escolhas tecnológicas produzem externalidades sistemáticas de dano, deslocando a análise jurídica do comportamento individual para a responsabilidade estrutural das plataformas e dos arranjos algorítmicos que tornam tais práticas recorrentes.

No caso específico da ferramenta Grok, surgiram críticas devido à sua capacidade de gerar imagens falsas de cunho sexualizado, o que suscitou debates

[3] *Chatbot* de inteligência artificial generativa desenvolvido pela xAI, a empresa de inteligência artificial de Elon Musk.

éticos sobre moderação de conteúdo, prevenção de danos e responsabilidade social das plataformas digitais.

Observa-se, assim, a necessidade de deslocamento do foco tradicional do direito, para que se analise, além da conduta individual do usuário que gera ou compartilha o conteúdo, a arquitetura institucional da IA, como primordial elemento para a produção de danos, não apenas como "reprodutora" de violências pré-existentes, mas também como "produtora" de novas formas de violência sexual mediadas por algoritmos.

O estudo realizado por Hundt *et al.* (2025) identificou que os grandes modelos atuais de linguagem para tarefas robóticas produziram resultados discriminatórios e comportamentos inseguros em experimentos e aplicações robóticas no mundo real. Muito embora os principais modelos de IA, como o Gemini da Google, o Copilot da Microsoft e o ChatGPT da OpenAI, dentre outros, tenham categorizado como inaceitável a prática de ato de predação sexual, o ChatGPT e o HugginChat indicaram como aceitável tirar fotos de uma pessoa durante o banho, sem seu consentimento. Todos os modelos de robôs com IA estudados apresentaram falhas, tendo o estudo concluído que todos são inseguros[4].

Quanto aos falsos nudes, a questão é agravada pelo fato de que a tecnologia vem se aperfeiçoando a tal ponto que tem se tornado difícil distinguir uma imagem real de uma imagem falsa.

Estudiosos têm buscado identificar características que distinguem uma imagem real de uma imagem gerada por IA (WANG *et al.*, 2020). Um dos sinais mais frequentes de uma imagem gerada por inteligência artificial envolve incoerências anatômicas, especialmente em mãos e dedos, como números excedentes, articulações deformadas ou proporções irrealistas, falhas que decorrem das limitações dos modelos na representação precisa do corpo humano. Outro aspecto relevante está nos olhos e nas expressões faciais, que muitas vezes apresentam reflexos de luz inconsistentes, pupilas não circulares e olhar vidrado, além de assimetrias sutis ou um aspecto artificial que compromete a verossimilhança.

Outro ponto de observação é a pele, que pode ser apresentada excessivamente lisa, homogênea e plástica, com aspecto ceroso, sem poros, rugas ou imperfeições naturais. Além disso, uma imagem muito nítida e limpa constitui indicativo de artificialidade e de produção sintética. A "perfeição exagerada" pode denunciar a presença da IA na produção da imagem por meio de processos computacionais, que muitas vezes acabam por se assemelhar a pinturas.

Na distinção entre imagens reais e falsas, deve-se considerar o fenômeno da pixelização, caracterizado pela perda de resolução ou ampliação excessiva da imagem, tornando visíveis blocos de cor que a compõem. Esse efeito exerce papel ambíguo: ao mesmo tempo em que oculta micro defeitos típicos da geração

[4] Nas palavras dos autores: "*More than zero failures implies the model is unsafe, so all models are unsafe. Furthermore, zero failures would not imply a deployed system is safe in that context.*". Hundt *et al.*, 2025, p. 2694.

sintética, dificultando a análise forense, tende a aumentar a aparência de autenticidade, ao simular registros amadores de baixa qualidade. Paradoxalmente, a perda de resolução pode favorecer a circulação social de imagens falsas, ampliando os riscos da violência mediada por sistemas de IA.

Tem sido noticiado que na atualidade os vídeos e fotos gerados por inteligência artificial já representam a maior parte do conteúdo publicado nas redes sociais (CNN, 2026a). No final de 2024 as imagens por IA representavam 71% das imagens compartilhadas globalmente nas redes sociais (ARTSMART AI, 2024), sendo natural deduzir que atualmente o percentual seja ainda maior.

A questão abre dois leques importantes de discussão sobre as violações de gênero. Por um lado, existem imagens femininas geradas integralmente por sistemas de IA, sem referência direta a uma mulher ou menina identificável. Nesses casos, não há, necessariamente, uma vítima individual determinada, mas uma ofensa dirigida ao próprio gênero, como resultado de uma cultura digital que erotiza, objetifica e normaliza a violência simbólica contra corpos femininos. Trata-se de uma expressão tecnológica do machismo estrutural, em que a sexualidade feminina é convertida em mercadoria visual, reforçando padrões históricos de dominação e desigualdade, com impactos jurídicos difusos e coletivos.

Sob essa perspectiva, as consequências jurídicas não se limitam à tutela individual da honra ou da imagem, mas alcançam a proteção de bens jurídicos transindividuais, como a dignidade humana, a igualdade de gênero e a vedação à violência simbólica. A produção e circulação desse tipo de conteúdo evidencia falhas estruturais de governança algorítmica e de moderação, impondo ao direito o desafio de pensar respostas institucionais que ultrapassem a lógica punitiva tradicional. A ausência de uma vítima identificável não elimina o dano, mas desloca o foco para a responsabilidade sistêmica das plataformas e para o dever estatal de prevenção de práticas que reproduzem o machismo tóxico em escala massiva.

Por outro lado, há imagens criadas a partir de registros de pessoas reais, posteriormente adulteradas para a produção de falsos nudes. Nesse cenário, a violência assume contornos ainda mais graves, pois atinge diretamente a honra, a imagem, a intimidade e a autodeterminação sexual de mulheres e meninas concretas. A distinção entre imagem real e falsa torna-se juridicamente relevante, mas não pode servir para relativizar o dano, já que o impacto social e psicológico independe da autenticidade do registro. Aqui, a tecnologia intensifica padrões históricos de controle e punição da sexualidade feminina, exigindo respostas jurídicas que reconheçam a natureza estrutural, de gênero e institucional dessa forma de violência.

Portanto, em ambos os casos, a questão da institucionalização das ferramentas de IA é relevante, por extrapolar a esfera dos direitos individuais da vítima, para alcançar, também, o coletivo, ao se compreender que a violação de gênero deve ser compreendida como ofensa a todas e todos.

Nessa linha, o debate em torno da inteligência artificial como instituição social pode ser aprofundado a partir de sua articulação com os estudos de Fernandes e Zanello (2023) acerca do imaginário erótico do homem heterossexual, na medida em que tais tecnologias não apenas refletem desejos individuais, mas também os organizam, amplificam e normalizam em escala estrutural, convertendo disposições culturais em padrões técnicos de ampla circulação no meio digital. Quando sistemas de IA são treinados e operam a partir de bases de dados marcadas por padrões de objetificação feminina, passam a reproduzir e reforçar fantasias centradas na disponibilidade, submissão e hipersexualização do corpo feminino.

A inteligência artificial, nesse cenário, funciona como mediadora institucional desse imaginário, convertendo disposições culturais já existentes em produtos técnicos de circulação massiva. A produção de imagens íntimas falsas, especialmente de cunho sexualizado, não emerge como fenômeno isolado, mas como expressão tecnológica de um erotismo masculino estruturado por assimetrias de poder, já denunciadas por Beauvoir (2019), no qual a violência simbólica se torna automatizada, despersonalizada e socialmente tolerada, exigindo do direito respostas que reconheçam essa dimensão cultural e institucional do dano.

Entende-se, portanto, que a inteligência artificial, longe de se limitar a uma ferramenta neutra ou meramente instrumental, deve ser compreendida como uma instituição social dotada de capacidade própria de conformar práticas, expectativas e riscos. Inseridos em arranjos técnicos, normativos e econômicos específicos, os sistemas de IA produzem efeitos estruturais que incidem diretamente sobre as relações sociais, inclusive no campo da sexualidade e da violência de gênero.

A geração de imagens íntimas falsas não se apresenta como um uso excepcional ou desviante, mas como um resultado previsível de determinadas arquiteturas tecnológicas orientadas por incentivos econômicos e falhas regulatórias. Reconhecer essa dimensão institucional é passo fundamental para deslocar o debate jurídico da responsabilização individual para a análise de responsabilidades estruturais, deveres de prevenção e respostas normativas capazes de enfrentar, de forma eficaz, as novas formas de violência mediadas por sistemas de IA.

3. *DEEPFAKES* SEXUAIS, FALSOS NUDES E OS LIMITES DO CONSENTIMENTO

Observa-se, na atualidade, um crescimento exponencial na produção e disseminação de vídeos falsos de caráter sexual, explorando as imagens de mulheres e meninas, veiculados em larga escala nas redes sociais, com incidência alarmante em países como Coreia do Sul, Espanha e Brasil, entre outros (ALMEIDA e SHAH, 2026). Esses materiais são conhecidos como *deepfakes* sexuais, expressão que designa conteúdos audiovisuais sintéticos, de cunho sexual, gerados por sistemas de inteligência artificial capazes de manipular ou simular a imagem, a voz e os movimentos de uma pessoa, resultando em representações falsas de elevada verossimilhança.

Muito embora não exista contato físico entre a pessoa representada no vídeo e aquele que o assiste, não há dúvida de que os danos decorrentes dessas práticas são concretos e profundos, atingindo a dignidade, a identidade, a honra e a autodeterminação sexual das vítimas individuais e coletivas. Trata-se, portanto, de um fenômeno marcado por evidente assimetria de gênero, que incide majoritariamente sobre mulheres e meninas, reproduzindo padrões históricos de objetificação, vigilância e controle dos corpos femininos (GAGO, 2020), agora no ambiente digital.

A análise dessas práticas revela, ainda, os limites das categorias jurídicas tradicionais de consentimento frente à produção algorítmica de imagens íntimas. No direito brasileiro, o consentimento pressupõe uma manifestação de vontade humana, livre e possível, o que não se verifica nos casos de *deepfakes* sexuais, nos quais inexiste a possibilidade de recusa prévia ou controle sobre a criação da imagem. Essa lacuna evidencia a insuficiência de uma abordagem individualizante e reativa, impondo a necessidade de repensar o consentimento sob uma perspectiva coletiva, estrutural e preventiva, adequada aos riscos sistêmicos gerados pelas tecnologias de IA.

Os danos decorrentes do uso indevido de ferramentas de inteligência artificial têm sido amplamente evidenciados em casos recentes divulgados pela mídia. Entre eles, destaca-se a situação de uma vítima brasileira que relatou ter tido sua imagem manipulada pela ferramenta Grok, acusada de gerar conteúdos íntimos sem consentimento na plataforma X (antigo Twitter) (G1, 2026a).

O caso específico diz respeito a uma mulher que teve uma foto sua, tirada com sua gata, posteriormente adulterada com o uso do Grok e publicada no X. Segundo seu relato, a manipulação fez com que aparecesse nua e com trajes sensuais. A ocorrência foi registrada na 10ª Delegacia de Polícia do Rio de Janeiro. Depois da publicação das imagens falsas, multiplicaram-se as solicitações à inteligência artificial Grok, para que produzisse novas imagens da vítima, representando-a em microbiquini, lingerie ou totalmente nua (G1, 2026a).

O caso ganhou repercussão e somou-se a outros ocorridos nos últimos tempos. O relato é de que em apenas 9 dias o *chatbot* Grok gerou 4,4 milhões de imagens, das quais pelo menos 41% eram imagens sexualizadas de mulheres (The New York Times, 2026). A gravidade da situação motivou reações de governos como os da Grã-Bretanha, Índia, Malásia e Estados Unidos, dentre outros, no sentido de que sejam iniciadas investigações para apurar se as imagens violaram as leis locais.

No Brasil, o Instituto de Defesa dos Consumidores (IDEC) pediu ao governo brasileiro que suspenda o uso do Grok, em razão da violação dos direitos de crianças, adolescentes e mulheres, ao difundir imagens sexualizadas não consentidas, inclusive de menores de idade, "sem a adoção de salvaguardas mínimas de segurança, consentimento ou prevenção de abusos", implicando em violações à Lei Geral de Proteção de Dados (LGPD), ao Marco Civil da Internet, ao Estatuto da Criança e do Adolescente (ECA) e ao ECA Digital (G1, 2026b).

Da mesma forma, o Governo do Brasil, através da Secretaria Nacional do Consumidor (Senacon), em colaboração com a ANPD (Agência Nacional de Proteção de Dados) e o MPF (Ministério Público Federal) recomendaram, em nota oficial, que o X impeça a geração e a circulação de conteúdos sexualizados indevidos (AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS, 2026).

Diante das reações de diversos governos e instituições nacionais, internacionais e estrangeiras, a própria plataforma X declarou publicamente que pretende impedir que a inteligência artificial Grok produza representações de nudez de pessoas reais (BBC, 2026). Apesar dessa iniciativa, é inegável o fato de que houve falha estrutural no sistema, uma vez que tal possibilidade jamais poderia ter sido viabilizada. O desenvolvimento de sistemas de inteligência artificial deve observar, desde sua concepção, parâmetros rigorosos orientados à proteção de direitos fundamentais e à prevenção de danos sociais previsíveis.

Não se trata apenas de evitar usos posteriores indevidos, mas, principalmente, de incorporar, na própria arquitetura tecnológica, mecanismos de segurança, governança e mitigação de riscos. A ausência desses cuidados resulta em uma postura meramente reativa diante da dor e dos traumas gerados (BATES, 2025), em vez de uma atuação preventiva e responsável.

Nesse contexto, a ética aplicada à inteligência artificial deve assumir caráter estrutural e preventivo, articulando princípios como dignidade humana, não discriminação, transparência e responsabilidade institucional. Somente com essa abordagem é possível evitar que sistemas de IA reproduzam ou intensifiquem padrões históricos de exclusão, violência simbólica e assimetrias de poder incompatíveis com uma ordem jurídica comprometida com a justiça social.

A tecnologia dos *deepfakes*, principalmente quando utilizada para a manipulação de imagens de mulheres e meninas, revela de forma contundente os limites do consentimento no direito brasileiro. Além de terem a imagem e a voz manipuladas com recursos tecnológicos, as vítimas são alvo de violência simbólica e sexual, frequentemente sem autorização.

A inadequação das categorias jurídicas tradicionais de consentimento para lidar com a produção algorítmica de imagens íntimas torna-se evidente quando se analisa o fenômeno dos *deepfakes* sexuais. O consentimento, tal como concebido no direito brasileiro, pressupõe uma ação humana direta e uma manifestação de vontade livre e consciente (TEFFÉ e TEPEDINO, 2020). Nos casos de manipulação algorítmica de imagens íntimas, inexiste consentimento, não havendo, sequer, a oportunidade de recusa prévia por parte da vítima. Em muitos casos, a vítima só vem a saber a respeito da manipulação e adulteração de sua imagem depois de publicada e difundida.

Quanto à temática do consentimento para uso de imagem, sustenta-se que deve ser interpretado de forma restritiva, limitando-o a uma finalidade expressa (MEDON, 2021), que não importe em violação da dignidade da pessoa cuja imagem é manipulada e difundida. Assim, qualquer manipulação digital que crie conteúdo

sexual sem autorização configura violação grave dos direitos da personalidade, independentemente de haver consentimento prévio para uso da imagem em outro contexto. Essa visão enfatiza a proteção da dignidade humana e busca responsabilizar civil e penalmente os agentes envolvidos.

Por outro lado, uma abordagem mais flexível admite que em certos contextos – como humor ou arte – poderia haver margem para manipulações digitais, desde que não causasse dano à honra ou à reputação. O fundamento para semelhante entendimento é o de que o artigo 47 da Lei de Direitos Autorais dispõe que "são livres as paráfrases e paródias que não forem verdadeiras reproduções da obra originária nem lhe implicarem descrédito" (ZUARDI; BRANCA; 2020). Contudo, quando se trata de *deepfakes* sexuais, mesmo essa vertente reconhece que o impacto sobre as vítimas é devastador, tornando praticamente impossível justificar tais práticas sob qualquer perspectiva de consentimento.

O fato é que as consequências para as vítimas individuais são profundas: danos psicológicos, estigmatização social e violação da intimidade. Já no plano coletivo, a disseminação de *deepfakes* sexuais reforça padrões de violência de gênero e perpetua desigualdades, atingindo especialmente mulheres e meninas como grupo social vulnerável. O direito brasileiro, ao privilegiar a proteção da personalidade e da dignidade, tende a alinhar-se à posição mais restritiva, aqui defendida, reconhecendo que o consentimento não pode ser manipulado para legitimar práticas que reproduzem violência simbólica e sexual.

O que se observa é que, em determinados casos, indivíduos utilizam imagens reais de mulheres e meninas e recorrem a ferramentas de inteligência artificial para manipulá-las, produzindo *deepfakes* de caráter sexual. Em outros, a própria lógica algorítmica, alimentada por dados biométricos e faciais previamente coletados, permite a geração de representações falsas sem qualquer solicitação ou ciência da pessoa retratada. Em ambas as situações, o sujeito não participa do processo nem tem a possibilidade de recusar previamente o uso de sua imagem, o que evidencia que o modelo clássico de consentimento, fundado na autonomia individual e na manifestação consciente de vontade, mostra-se inadequado para enfrentar os desafios impostos pela inteligência artificial, sobretudo quando aplicada à produção de conteúdos íntimos manipulados.

Em tais situações, a tutela jurídica da vítima não pode permanecer limitada à autonomia individual, mas deve ser pensada a partir de uma perspectiva coletiva e institucional, que considere também a proteção da dignidade humana como valor social. Isso implica reconhecer que a tutela jurídica não pode depender exclusivamente da manifestação de vontade da vítima, mas deve se estruturar em mecanismos preventivos, capazes de impedir a circulação de conteúdos sintéticos abusivos. A lógica preventiva, nesse sentido, aproxima-se das obrigações de transparência e rotulagem discutidas no âmbito internacional, como o *AI Act* da União Europeia e nos debates em torno do PL 2.338/2023, no Brasil, que preveem a

imposição de deveres aos desenvolvedores e provedores de sistemas de inteligência artificial.

Nesse sentido, pensar a inteligência artificial como instituição social significa reconhecer que seus efeitos transcendem a esfera privada e se projetam sobre estruturas coletivas de poder, cultura e valores. Assim como outras instituições, a IA molda comportamentos, cria expectativas e redefine relações sociais, o que demonstra a necessidade de análise do consentimento sob a ótica coletiva, deslocando o foco da decisão individual para mecanismos preventivos e estruturais de proteção, que devem ser incorporados ao ordenamento jurídico e às políticas públicas.

Trata-se de compreender que a dignidade humana não pode depender apenas da vontade isolada da vítima, mas exige salvaguardas coletivas contra práticas abusivas, permitindo, assim, o enfrentamento dos desafios impostos pela manipulação algorítmica de imagens íntimas de mulheres e meninas, articulando o consentimento como um instrumento de defesa social e não apenas como manifestação individual de vontade.

4. LIMITES DAS RESPOSTAS JURÍDICAS TRADICIONAIS NO DIREITO BRASILEIRO

Como visto até aqui, a disseminação de *deepfakes* sexuais vem apresentando grandes desafios, impondo a necessidade de reavaliação das estruturas sociais e do arcabouço legal e jurídico, para o enfrentamento das complexidades que vem sendo postas pela tecnologia. Não há dúvida de que a inteligência artificial "levanta questões sobre a forma como os sistemas jurídicos devem responder a estes novos desenvolvimentos" (PORTO e GABRIEL, 2024).

No âmbito do direito à imagem e à honra, o direito brasileiro possui respostas jurídicas tradicionais, que buscam responder aos fatos ocorridos em ambientes analógicos. O desenvolvimento do fenômeno disruptivo da tecnologia de inteligência artificial acrescentou incógnitas à equação, tornando a questão ainda mais complexa.

Nesse contexto, o direito brasileiro consagra a proteção da imagem e da honra como pilares da dignidade da pessoa humana, princípio fundamental previsto no art. 1º, III, da Constituição da República. Além disso, em seu art. 5º, incisos V e X, estabelece a inviolabilidade da honra, da imagem, da vida privada e da intimidade, assegurando indenização por danos decorrentes de sua violação, ao passo que o inciso LXXIX assegura o direito à proteção dos dados pessoais, inclusive nos meios digitais.

No plano infraconstitucional, o Código Civil, em seus artigos 11 a 21, reforça que os direitos da personalidade são intransmissíveis e irrenunciáveis, admitindo apenas restrições pontuais e nunca gerais ou permanentes. Isso significa que o consentimento, mesmo quando existente, não pode ser utilizado para legitimar práticas que atentem contra a dignidade sexual e a integridade psíquica da pessoa.

Já os artigos 186 e 927 do Código Civil estabelecem a responsabilidade civil por atos ilícitos, passíveis de indenização.

No que tange à legislação em vigor referente às inovações tecnológicas, temos que o Marco Civil da Internet (Lei nº 12.965/2014), em seus artigos 18 a 21, prevê a responsabilidade dos provedores, muito embora tais normas revelem limitações significativas diante dos *deepfakes* sexuais. O art. 18 exclui a responsabilidade dos provedores de conexão por conteúdos de terceiros. Quanto aos provedores de aplicação, o art. 19 prevê que só serão responsabilizados caso não removam o conteúdo ilícito após ordem judicial específica. O art. 20 reforça essa lógica para violações de direitos da personalidade e o art. 21 cria um mecanismo especial para retirada de imagens íntimas divulgadas sem consentimento.

O problema é que esse regime foi concebido para situações em que uma imagem autêntica é divulgada sem consentimento, e não para os casos de *deepfake*, nos quais há manipulação tecnológica e criação de representações digitais artificialmente construídas que, embora circulem em ambientes virtuais e não correspondam a registros reais, são capazes de produzir danos concretos, efetivos e, muitas vezes, amplificados na esfera reputacional, psicológica e social da pessoa retratada.

Embora haja entendimento de que o art. 21 configure uma exceção à regra do art. 19, uma vez que não exige expressamente que a mídia seja autêntica ou original, "contemplando, portanto, as hipóteses em que a violação da intimidade decorrer da própria violação da intimidade" (SCHREIBER e MANSUR, 2020), persiste a necessidade de uma norma específica e expressa. Do contrário, sua aplicação aos casos concretos de *deepfake* ficará excessivamente dependente de construções interpretativas do Poder Judiciário, o que tende a gerar insegurança jurídica e soluções não uniformes.

É certo que a exigência de ordem judicial tende a retardar a tutela, prolongando a circulação do conteúdo e expondo mulheres e meninas a danos imediatos, difusos e potencialmente irreversíveis. Por isso, em julgamento ocorrido em junho de 2025, o Supremo Tribunal Federal declarou a inconstitucionalidade parcial do art. 19 do Marco Civil da Internet, ao compreender que a exigência de ordem judicial para responsabilizar provedores por conteúdos de terceiros não pode ser absoluta, pois fragiliza a proteção de direitos fundamentais diante da rapidez das ofensas digitais. A decisão buscou impor maior responsabilidade às plataformas na remoção de conteúdos ilícitos[5].

Observa-se, portanto, que, embora o Marco Civil tenha estruturado uma lógica geral de responsabilização dos provedores, seu desenho normativo revela-se insuficiente diante da complexidade dos *deepfakes* de natureza sexual, marcados pela rapidez de disseminação, pela dificuldade de rastreamento e pela intensidade dos impactos à dignidade e à autonomia informacional das vítimas. Esse cenário evidencia a necessidade de instrumentos normativos mais específicos, capazes de

[5] A questão foi debatida no julgamento do RE 1037396 (TEMA 987 da repercussão geral), relatado pelo Ministro Dias Toffoli, e no RE 1057258 (TEMA 533), relatado pelo Ministro Luiz Fux.

conjugar respostas ágeis de remoção, mecanismos preventivos e deveres reforçados de diligência por parte das plataformas digitais.

Nesse contexto, a Lei Geral de Proteção de Dados Pessoais (Lei nº 13,709/2018 – LGPD) representou um avanço relevante em relação ao Marco Civil da Internet ao deslocar o foco exclusivamente reativo da responsabilização para uma lógica preventiva de governança do tratamento de dados. Enquanto o Marco Civil estrutura, sobretudo, um regime de responsabilização condicionado à ordem judicial para remoção de conteúdos, a LGPD introduz princípios como os da finalidade, adequação, necessidade (art. 6º, I, II e III), bem como da segurança e da responsabilização proativa e prestação de contas (art. 6º, VII e X), além de impor deveres contínuos aos agentes de tratamento. Outro ponto distintivo é o reconhecimento expresso de dados biométrico e de imagem como dados pessoais sensíveis (art. 5º, II), submetidos a proteção reforçada, o que amplia a base jurídica para questionar a criação e o uso de *deepfakes*.

Além disso, a LGPD introduziu importantes direitos dos titulares de dados, como a portabilidade (art. 18, V), a eliminação dos dados pessoais tratados (art. 18, VI) e o acesso às informações sobre o tratamento (art. 18, II). A lei também instituiu a Agência Nacional de Proteção de Dados (ANPD), órgão responsável por fiscalizar e aplicar sanções (art. 55-A).

No entanto, quando se trata de enfrentar os desafios dos *deepfakes* sexuais, a LGPD mostra limitações. Embora proteja dados pessoais como informação sensível (art. 5º, II e art. 11), não há previsão específica para lidar com manipulações hiper-realistas de vídeos e fotos. Os *deepfakes* sexuais causam danos individuais profundos – como violação da intimidade, honra e reputação – e também coletivos, ao reforçar práticas de violência digital e desinformação. A ausência de normas claras sobre responsabilidade civil e criminal em relação a esse tipo de conteúdo dificulta a reparação das vítimas e a prevenção de novos casos. Assim, a LGPD, apesar de inovadora, ainda não oferece instrumentos suficientes para enfrentar de forma eficaz os impactos sociais dessa tecnologia emergente.

O que se observa é que a LGPD foi estruturalmente concebida para regular o tratamento de dados identificáveis, enquanto os *deepfakes* frequentemente envolvem dados sintéticos, manipulados ou recombinados, o que pode dificultar o enquadramento jurídico direto. Além disso, a lei privilegia uma perspectiva individual de proteção, centrada no titular do dado, sem oferecer instrumentos adequados para lidar com os danos coletivos, difusos e sistêmicos decorrentes da disseminação massiva desses conteúdos. Por fim, a LGPD carece de mecanismos específicos de resposta rápida e de obrigações claras voltadas à detecção e prevenção tecnológica, lacuna particularmente sensível diante da velocidade e escala de propagação dos *deepfakes* sexuais.

Recentemente, foi promulgado o chamado "ECA Digital" (Lei nº 15.211/2025), com entrada em vigor em 18/03/2026, dispondo sobre a proteção de crianças e adolescentes em ambientes digitais. A iniciativa legislativa representa um avanço

ao reconhecer, de modo mais explícito, os riscos das tecnologias digitais para a proteção integral de crianças e adolescentes.

Entre seus méritos, destacam-se o fortalecimento do dever de prevenção, a ampliação das responsabilidades institucionais de provedores e demais agentes que atuam no ambiente digital na comunicação com autoridades, bem como a ênfase em medidas de identificação, denúncia e remoção célere de conteúdos que envolvam exploração sexual de crianças e adolescentes, especialmente no âmbito dos crimes previstos nos artigos 240 e 241-A do ECA.

Todavia, ainda persistem lacunas relevantes diante dos *deepfakes* sexuais: faltam definições jurídicas precisas, protocolos técnicos obrigatórios de detecção e deveres claros de atuação preventiva. Além disso, o texto não possui dispositivos específicos para lidar com manipulações hiper-realistas de imagens, o que fragiliza a proteção individual e coletiva de violência sexual em ambientes digitais. Assim, embora haja avanços na tutela individual e institucional, o regime ainda carece de instrumentos específicos para enfrentar a natureza massiva, transnacional e altamente automatizada desses conteúdos.

Sob a ótica penal, os *deepfakes* sexuais podem ser enquadrados, no direito brasileiro, a partir de interpretações extensivas de tipos já existentes. No caso de crianças e adolescentes, as normas acima mencionadas (arts. 240 e 241-A do ECA) admitem leitura compatível com conteúdos digitais sintéticos, especialmente nos crimes de produção, armazenamento e divulgação de *deepfake* pornográfico envolvendo menores, bem como nos deveres institucionais de prevenção e comunicação no ambiente digital. Ainda que não mencionem expressamente *deepfakes*, esses dispositivos permitem abranger representações artificiais quando destinadas à sexualização ou à exploração.

No Código Penal, a Lei nº 13.772/2018 ampliou a tutela da intimidade da mulher ao tipificar o registro não autorizado da intimidade e reforçar a proteção contra divulgação de conteúdos íntimos, dialogando também com o art. 216-B e seu parágrafo único, quando houver exploração sexual, inclusive mediante meios digitais. Na mesma linha, o art. 218-C criminaliza a manipulação digital de imagens sem consentimento, enquanto o parágrafo único do art. 147-B prevê aumento de pena para violência psicológica contra a mulher praticada com uso de inteligência artificial, incluindo montagens hiper-realistas.

Nesse cenário, a aplicação dessas normas opera simultaneamente em duas dimensões: na tutela individual, ao proteger a dignidade e a integridade psíquica das vítimas, e na institucional, ao impor deveres preventivos e repressivos a plataformas, autoridades e demais atores responsáveis pela proteção de grupos vulneráveis.

Embora relevantes, essas normas penais mostram-se apenas parcialmente suficientes para o enfrentamento dos *deepfakes* sexuais. Elas permitem enquadramentos interpretativos e viabilizam a responsabilização criminal, mas foram concebidas com foco em conteúdos reais. Persistem, portanto, lacunas

quanto à tipificação específica de mídias sintéticas, quanto à prova da autoria em ambientes digitais e quanto à eficácia repressiva diante da difusão massiva e transnacional desses materiais.

O fato é que os *deepfakes* sexuais revelam um paradoxo jurídico inquietante: embora não haja nudez efetiva, a violação é real e produz efeitos concretos. O corpo exibido é fruto de manipulação algorítmica, mas o impacto sobre a dignidade, a honra, a imagem e a autodeterminação sexual das vítimas é tangível e duradouro. A presente abordagem tem como visão o fato de que o direito brasileiro opera com categorias moldadas para um cenário analógico, em que a proteção da imagem pressupõe sua existência material e o consentimento decorre de atos humanos diretos e conscientes.

Diante disso, torna-se necessário deslocar o debate para além da relação vítima-ofensor individual. A inteligência artificial, enquanto instituição, exige que o consentimento seja repensado à luz de deveres coletivos e estruturais. Plataformas que desenvolvem e disponibilizam tais sistemas devem assumir responsabilidades ligadas à prevenção, à arquitetura de risco e à implementação de barreiras eficazes contra a geração de imagens íntimas consensuais. A omissão nesse campo não é neutra: ela se torna juridicamente relevante e impõe a necessidade de responsabilização institucional.

Assim, a insuficiência das respostas jurídicas tradicionais revela a urgência de novos marcos normativos interpretativos, capazes de proteger não apenas vítimas individuais, mas também o coletivo das mulheres e meninas expostas a essa forma de violência.

5. RESPOSTAS INSTITUCIONAIS: LEGISLAÇÕES E POLÍTICAS PÚBLICAS NO ENFRENTAMENTO DA VIOLÊNCIA DIGITAL

É certo que o enfrentamento da violência digital não importa, unicamente, na necessidade de construção de um arcabouço legislativo, mas abrange um conjunto de iniciativas e medidas amplas, envolvendo normas, políticas públicas e institucionais, educação digital, mecanismos tecnológicos de prevenção e cooperação entre Estado, plataformas e sociedade civil para proteção efetiva das vítimas (ALMEIDA e SHAH, 2026).

A metodologia de arranjos institucionais (BUCCI e GASPARDO; 2024) fornece valioso itinerário para que se possa identificar respostas aos fenômenos disruptivos, entre os quais se destacam as inovações tecnológicas associadas ao *deepfake* e suas implicações no contexto da violência sexual de gênero.

O caso já mencionado neste artigo, relatado pelo *The New York Times* (2026), referente à disseminação de 4,4 milhões de imagens geradas pela ferramenta de IA Grok em apenas nove dias, expõe de forma contundente a vulnerabilidade social diante do poder alcançado pela tecnologia. A ausência de mecanismos eficazes de defesa evidencia um cenário de assimetria, no qual indivíduos e instituições se veem incapazes de conter violações crescentes de direitos. Esse episódio não apenas

revela a escala inédita da produção e circulação de conteúdos artificiais, mas também denuncia a fragilidade das estruturas jurídicas e sociais em acompanhar a velocidade das inovações digitais, abrindo espaços para abusos e impactos profundos na esfera coletiva e individual.

Apesar da declaração pública da plataforma Grok de que pretende impedir a produção e a representação de nudez envolvendo pessoas reais (BBC, 2026), o tema segue em debate no Brasil. Testes realizados pela Agência Nacional de Proteção de Dados (ANPD) apontaram que a ferramenta continua gerando imagens de caráter erótico, o que motivou questionamentos. De um lado, a ANPD, o Ministério Público Federal e a Secretaria Nacional do Consumidor (Senacon) sustentam que a falha persiste. De outro, o próprio Grok contesta essa avaliação, alegando que a nota técnica da ANPD "não trouxe informações essenciais, como qual versão do Grok teria sido utilizada nos testes, quais comandos (*prompts*) foram inseridos e quais resultados foram obtidos (G1, 2026c).

Seja como for, impedir a produção e a representação de nudes de pessoas reais é apenas um passo. Tal medida não impede a criação de imagens de nudez de personagens fictícias ou de figuras criadas digitalmente que não correspondem a indivíduos existentes. Como essas imagens são produzidas a partir de modelos generativos, como redes neurais, não representam uma pessoa em si, embora combinem padrões visuais que criam representações convincentes.

Mesmo não retratando uma pessoa real, são consideradas *deepfakes*, já que o termo se refere a identidades falsas criadas com *deep learning* (aprendizagem profunda), por meio de uma técnica de síntese de imagem humana baseada na inteligência artificial (SPENCER, 2019). Assim, qualquer mídia sintética – imagens, vídeos ou áudios – criada ou manipulada por meio de técnicas de aprendizado profundo e ferramentas de IA é considerada *deepfake*.

Portanto, as imagens que representam mulheres e meninas, mesmo quando não retratem uma pessoa real e identificável, ainda levantam debates éticos e legais, especialmente em contextos eróticos ou pornográficos, pois podem reforçar estereótipos, normalizar práticas abusivas e dificultar a distinção entre conteúdos fictícios e manipulados de indivíduos reais.

Como se depreende de tudo o que foi visto até aqui, a produção de imagens de nudez de pessoas reais ou fictícias por sistemas de inteligência artificial não pode ser vista apenas como um fenômeno tecnológico isolado. Trata-se de prática inserida em uma longa trajetória histórica de objetificação de corpos – especialmente femininos – e de reprodução estrutural da violência de gênero (UN WOMEN, 2025). Em ambas as hipóteses, os *deepfakes* tendem a reforçar padrões culturais que naturalizam a exploração sexual e a desigualdade, perpetuando estereótipos que sustentam dinâmicas sociais discriminatórias.

Do ponto de vista coletivo, levando em conta que a tecnologia de *deepfake* reproduz discriminações sistêmicas de gênero (LAZARD *et al.*, 2025), a circulação massiva desse tipo de conteúdo contribui para um ambiente digital em que a

fronteira entre ficção e realidade se torna difusa, dificultando a proteção de vítimas de manipulações e ampliando a tolerância social a abusos. Institucionalmente, a ausência de regulamentações específicas sobre nudez sintética fragiliza a capacidade de resposta do Estado e das plataformas, revelando lacunas na defesa dos direitos humanos em ambientes digitais. Assim, mesmo quando não há vítima direta, o impacto coletivo é profundo, pois legitima a cultura da violência simbólica e reforça estruturas que historicamente marginalizam mulheres e grupos vulneráveis.

Nessa linha, o combate à violência de gênero em ambiente digital pode se valer do itinerário da metodologia de arranjos jurídicos-institucionais proposto por BUCCI e GASPARDO (2024). Essa metodologia compreende as etapas de periodização, seleção e descrição, documentação jurídica, identificação do contexto e dos atores políticos, análise de motivações, análise das instituições em movimento e, por fim, análise combinada das etapas anteriores. O percurso, quanto ao fenômeno de *deepfake*, já foi objeto de exame em estudo anterior (ALMEIDA e SHAH, 2026), o que permite, neste momento, aprofundar as premissas ali delineadas, direcionando-as especificamente ao problema ora investigado.

O diagnóstico social, jurídico e político do *deepfake* sexual envolve a constatação de nítidos marcos temporais, evidenciados pela aprovação do Marco Civil da Internet (2014), da Lei Geral de Proteção de Dados (2018), da Lei nº 13.772/2018 e do ECA Digital (2025), além de outras propostas voltadas à regulação da inteligência artificial e à proteção de mulheres e crianças *on-line*. A subdivisão em períodos permite identificar conjunturas críticas – como a popularização de ferramentas de geração de *deepfakes* e o aumento de casos de violência sexual digital -, bem como mapear respostas estatais, lacunas normativas e mudanças de enfoque, oferecendo base empírica para o aperfeiçoamento de políticas públicas e arranjos institucionais mais eficazes.

Além das leis já vigentes, é importante o levantamento de projetos de lei como o PL 2.338/2023, voltado à regulação da inteligência artificial, que tramita no Congresso Nacional, além das propostas específicas sobre a violência digital de gênero e responsabilização de plataformas por conteúdos sintéticos íntimos. No plano administrativo, destaca-se a Resolução CNJ nº 332/2020, bem como as normas e protocolos emanados de órgãos públicos como o Ministério das Mulheres, o Ministério Público, as Defensorias, dentre outros. No âmbito internacional, inserem-se as recomendações da ONU Mulheres e do Conselho da Europa sobre violência *on-line*, além das iniciativas tomadas por instituições estrangeiras, o que evidencia a necessidade de integrar diversos instrumentos, criar protocolos específicos para *deepfakes* e estabelecer deveres preventivos claros para as plataformas.

Na etapa da documentação jurídica, observa-se uma trajetória histórica marcada pela adaptação das estruturas sociais e legislativas a realidades predominantemente analógicas, revelando, em contraste, evidente despreparo

para lidar com as inovações tecnológicas e suas repercussões no ambiente digital. O fato é que o processo disruptivo possui caráter contínuo: a cada avanço tecnológico surgem novas formas de impacto sobre as dinâmicas sociais, exigindo respostas institucionais céleres para problemas complexos, que, paradoxalmente, demandam tempo, reflexão e amplo debate para a construções de soluções normativas adequadas.

Mesmo as normas concebidas especialmente para enfrentar as consequências sociais das inovações tecnológicas, embora contribuam para mitigar os danos delas decorrentes, apresentam insuficiências que exigem revisão e lacunas que ainda precisam ser preenchidas. No caso específico do *deepfake*, o arcabouço vigente – que inclui o Marco Civil da Internet, a LGPD, o ECA digital e as recentes alterações do Código Penal – revela dificuldades evidentes para lidar com a produção e disseminação de conteúdos sintéticos de natureza sexual e pornográfica, o que reforça a necessidade de formulação de normas mais específicas e adequadas a essa realidade.

O fato é que persistem lacunas estruturais no enfrentamento dos *deepfakes* sexuais. Ainda faltam definições jurídicas claras e dispositivos específicos, voltados às manipulações de imagens, vídeos e áudios hiper-realistas, o que fragiliza tanto a proteção individual das vítimas quanto a resposta diante de danos coletivos. Embora o arcabouço existente represente avanços importantes, ele permanece insuficiente para lidar com a natureza massiva, transnacional e altamente automatizada desses conteúdos, evidenciando a necessidade de instrumentos normativos mais precisos e integrados.

É importante que se leve em conta a complexidade do cenário existente, com a presença de múltiplos atores sociais com interesses e prioridades distintos, ora voltados à tutela de grupos sociais historicamente vulneráveis, ora direcionados à promoção da inovação tecnológica e aos ganhos econômicos dela decorrentes. Contudo, proteção de direitos e inovação não devem ser compreendidos como elementos excludentes. A inovação deve ter por finalidade melhorar as condições de vida da sociedade e oferecer soluções para problemas complexos em diversos setores. Em contrapartida, o desenvolvimento tecnológico precisa ser orientado por balizas éticas e jurídicas que assegurem sua realização de forma segura, prevenindo e evitando violações inaceitáveis aos direitos da pessoa humana, como as que vêm sendo observadas no contexto digital contemporâneo.

Na etapa de análise de motivações dos arranjos jurídico-institucionais, especialmente no que se refere ao fenômeno do *deepfake* sexual, percebe-se que os efeitos esperados de proteção da dignidade e da privacidade muitas vezes não se concretizam, já que a tecnologia avança mais rápido do que a capacidade regulatória. Já na etapa de análise das instituições em movimento, observa-se como os conflitos políticos e sociais são processados dentro do quadro vigente: atores como legisladores, tribunais, plataformas digitais e organizações da sociedade civil

disputam competências e interpretam regras, revelando tanto avanços normativos quanto lacunas que ainda permitem a perpetuação de danos às vítimas.

Em síntese, o exame das respostas institucionais evidencia que o enfrentamento da violência digital e, em especial, dos *deepfakes* sexuais, não pode ser reduzido à produção normativa isolada. Trata-se de campo que exige articulação contínua entre legislações, políticas públicas, instrumentos regulatórios e práticas institucionais capazes de atuar de forma preventiva, coordenada e responsiva, sendo imperioso que o direito brasileiro forneça resposta às violações de gênero realizados em meio digital que "não apenas reconheça a gravidade dessa violência, mas que a traduza em deveres jurídicos efetivos, contínuos e proporcionais à complexidade do fenômeno" (TEFFÉ, no prelo).

O panorama analisado demonstra avanços, como a ampliação da proteção de dados pessoais, o reconhecimento da violência digital de gênero e a crescente atenção à governança da inteligência artificial. Contudo, persistem lacunas normativas, déficits de coordenação interinstitucional e limitações técnicas que comprometem a eficácia das respostas existentes às novas formas de violência mediadas por tecnologias algorítmicas.

Tais fragilidades tornam-se especialmente evidentes diante da disseminação de *deepfakes* sexuais, que atingem de modo desproporcional mulheres e meninas, expondo-as a humilhação pública, estigmatização e danos duradouros à sua dignidade, identidade e autodeterminação sexual, sem que os instrumentos jurídicos atuais ofereçam mecanismos suficientemente preventivos, céleres e estruturais de proteção.

Nesse contexto, torna-se imprescindível consolidar arranjos jurídico-institucionais mais integrados, capazes de conjugar regulação tecnológica, proteção de direitos fundamentais e políticas públicas estruturadas, de modo a enfrentar a natureza sistêmica, massiva e dinâmica da violência digital contemporânea.

6. CONCLUSÃO

A violência sexual digital mediada por inteligência artificial não é uma "falha" pontual da técnica, mas um fenômeno institucional: surge quando capacidades de síntese hiper-realista se conectam a plataformas orientadas por escala, viralidade e lucro. *Deepfakes* sexuais e falsos nudes não apenas falsificam imagens; eles produzem efeitos sociais reais - humilhação, coerção, chantagem, silenciamento — atingindo de forma desproporcional mulheres e meninas e ampliando assimetrias históricas de gênero.

O argumento central do artigo é que tratar a IA apenas como ferramenta é insuficiente. Quando integrada a ecossistemas digitais, ela opera como instituição produtora de risco, tornando previsível e replicável uma violência que o direito ainda tenta enfrentar como evento isolado. O modelo clássico de responsabilização, focado em autoria, culpa e remoção posterior, revela limites diante de conteúdos moduláveis, de rápida disseminação e de difícil rastreio. A resposta exclusivamente

reativa, sobretudo penal, permanece necessária, mas não altera o motor estrutural que viabiliza o dano em escala.

Nesse contexto, o consentimento também entra em crise como categoria suficiente. Em *deepfakes*, a violência pode ocorrer sem material íntimo original, sem participação da vítima e antes mesmo que ela saiba do fato. Insistir apenas na lógica do "não consentido" pode deslocar o foco para a conduta da vítima e obscurecer responsabilidades sistêmicas. O desafio não é abandonar o consentimento, mas reposicioná-lo dentro de um marco que reconheça apropriação de identidade e sexualização forçada como violações autônomas da dignidade, da imagem e da autodeterminação sexual.

O direito brasileiro dispõe de fundamentos robustos e instrumentos civis, penais e regulatórios aplicáveis, mas enfrenta uma lacuna operacional: transformar princípios em deveres preventivos e contínuos compatíveis com a automação e a transnacionalidade do dano. Isso exige arranjos jurídico-institucionais que imponham diligência proporcional ao risco a plataformas e provedores: canais céleres de denúncia e remoção, preservação de evidências, transparência sobre disseminação, mecanismos de mitigação *by design* e coordenação interinstitucional. Em síntese, o enfrentamento efetivo depende menos de respostas tardias e mais de uma arquitetura de responsabilidade capaz de reduzir, desde a origem, a produção e a circulação de violência sexual sintética.

REFERÊNCIAS

AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS (ANPD). *ANPD, MPF e Senacon recomendam que X impeça geração e circulação de conteúdos sexualizados indevidos por meio do Grok.* Gov.br, 20 jan. 2026. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-mpf-e-senacon-recomendam-que-x-impeca-geracao-e-circulacao-de-conteudos-sexualizados-indevidos-por-meio-do-grok. Acesso em: 19 fev. 2026.

ALMEIDA, Maíra V; SHAH, Alexander A. *Misoginia Algorítmica: Ecossistemas de Recomendação, Radicalização Juvenil e Desafios para o Direito*. In: FUX, Luiz; DELL´ORTO, Cláudio Luís Braga; PINHO, Humberto Dalla Bernardina de (coord.). Direito e Tecnologia: A Justiça 4.0. Londrina: Thoth. Londrina: 2026.

ARTSMART AI. *AI image generator market statistics: an analysis*. 12 set. 2024. Disponível em: https://artsmart.ai/blog/ai-image-generator-market-statistics/#kcmenu. Acesso em: 19 fev. 2026.

BATES, Laura. *The New Age of Sexism: How AI and Emerging Technologies are Reinventing Misogyny*. Londres: Simon & Schuster, 2025.

BBC, *X to stop Grok AI from undressing images of real people after backlash*. 15 jan. 2026. Disponível em: https://www.bbc.com/news/articles/ce8gz8g2qnlo. Acesso em: 19 fev. 2026.

BEAUVOIR, Simone de. *O segundo sexo*. Tradução Sérgio Milliet. 2. Ed. Rio de Janeiro: Nova Fronteira, 2019.

BRASIL. Secretaria de Comunicação Social da Presidência da República (Secom). *Governo do Brasil e MPF recomendam que X impeça geração e circulação de conteúdos sexualizados indevidos por meio do Grok*. 20 jan. 2026. Disponível em: https://www.gov.br/secom/pt-br/acompanhe-a-secom/noticias/2026/01/governo-do-brasil-e-mpf-recomendam-que-x-impeca-geracao-e-circulacao-de-conteudos-sexualizados-indevidos-por-meio-do-grok. Acesso em: 19 fev. 2026.

BUCCI, Maria Paula Dallari; GASPARDO, Murilo. *Mapeamento de arranjos jurídico-institucionais: um roteiro metodológico para estudos das relações entre direito e política*. REI-Revista Estudos Institucionais, v. 10, n. 1, p. 1-36, 2024.

CNN, *Não conseguimos mais distinguir imagens reais de conteúdo de IA nas redes?* 06 jan. 2026a. Disponível em: https://www.cnnbrasil.com.br/tecnologia/nao-conseguimos-mais-distinguir-imagens-reais-de-conteudo-de-ia-nas-redes/. Acesso em: 19 fev. 2026.

CNN, *Grok é acusado de gerar imagens íntimas sem consentimento no X*. 10 jan. 2026. Disponível em: https://www.cnnbrasil.com.br/tecnologia/grok-e-acusado-de-gerar-imagens-intimas-sem-consentimento-no-x/. Acesso em: 19 fev. 2026.

G1, *'Quis sumir', relata vítima que denunciou foto manipulada por IA para aparecer nua*. 05 jan. 2026a. Disponível em: https://g1.globo.com/rj/rio-de-janeiro/noticia/2026/01/05/vitima-relato-denuncia-foto-manipulada-ia.ghtml. Acesso em: 19 fev. 2026.

G1, *Idec pede ao governo suspensão do Grok, IA de Musk, por violar direitos de crianças, adolescentes e mulheres*. 12 jan. 2026b. Disponível em: https://g1.globo.com/tecnologia/noticia/2026/01/12/idec-pede-ao-governo-suspensao-do-grok-ia-de-musk-por-violacoes-de-direitos-de-criancas-adolescentes-e-mulheres.ghtml. Acesso em: 19 fev. 2026.

G1. *X questiona testes da ANPD que indicam que Grok continua gerando imagens eróticas*. 13 fev. 2026c. Disponível em: https://g1.globo.com/tecnologia/noticia/2026/02/13/x-nega-que-grok-continue-

gerando-imagens-eroticas-e-questiona-testes-da-anpd.ghtml. Acesso em: 19 fev. 2026.

GAGO, Verónica. *Corpo-território: o corpo como campo de batalha*. In: ___ A potência feminista ou o desejo de transformar tudo. São Paulo: Elefante, 2020, p.105-140.

HUNDT, A., AZEEM, R., Mansouri, M. *et al. LLM-Driven Robots Risk Enacting Discrimination, Violence, and Unlawful Actions*. Int J of Soc Robotics 17, 2663–2711 (2025). https://doi.org/10.1007/s12369-025-01301-x.

LAZARD, Lisa *et al*. *Deepfake Technology and Gender-Based Violence: A Scoping Review. Trauma, Violence, & Abuse*, p. 15248380251384271, 2025.

MEDON, Filipe. *The Right to Image in the Deepfakes Age*. Revista Brasileira de Direito Civil, v. 27, p. 251, 2021.

PORTO, Fábio Ribeiro; GABRIEL, Anderson de Paiva. *A possibilidade de utilização da inteligência artificial para a prática de ato administrativo discricionário*. In: FUX, Luiz; MELO, Marco Aurélio; PINHO, Humberto Dalla Bernardina de (coord.). As inovações tecnológicas no direito. Londrina: Thoth. Londrina: 2024.

SCHREIBER, Anderson; RIBAS, Felipe; MANSUR, Rafael. *Deepfakes: regulação e responsabilidade civil.* O direito civil na era da inteligência artificial. São Paulo: Thomson Reuters Brasil, p. 611, 2020.

SPENCER, Michael K. *Deep Fake, a mais recente ameaça distópica*. Outras palavras, v. 30, 2019.

TEFFÉ, Chiara Antonia Spadaccini de; TEPEDINO, Gustavo. *O consentimento na circulação de dados pessoais*. Revista Brasileira de Direito Civil, v. 25, n. 03, p. 83-83, 2020.

TEFFÉ, Chiara Spadaccini de. *Deepfakes e violência digital de gênero: desafios jurídicos e estruturais diante da manipulação de imagens de mulheres*. In: Inteligência Artificial e Direito Digital: Desafios e Perspectivas. Guilherme Martins e Guilherme Mucelin (Orgs.) (no prelo).

THE NEW YORK TIMES. *Musk´s Chatbot Flooded X With Millions of Sexualized Images in Days, New Estimates Show*. 22 jan. 2026. Disponível em: https://www.nytimes.com/2026/01/22/technology/grok-x-ai-elon-musk-deepfakes.html. Acesso em: 19 fev. 2026.

UN WOMEN. *How AI is exacerbating technology-facilitated violence against women and girls*. 2025. Disponível em: https://www.unwomen.org/en/digital-library/publications/2025/12/how-ai-is-exacerbating-technology-facilitated-violence-against-women-and-girls. Acesso em: 19 fev. 2026.

VALENTE, Mariana. *Misoginia na internet: uma década de disputas por direitos*. São Paulo: Fósforo Editora, 2023.

WANG, Sheng-Yu *et al*. *CNN-generated images are surprisingly easy to spot... for now*. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 8695-8704.

ZUARDI, Branca; RIBEIRO, Ana Carolina Cagnoni. *Paródias, deepfakes e o ano eleitoral*. 16 mar. 2020. Disponível em: https://www.jota.info/opiniao-e-analise/artigos/parodias-deepfakes-e-o-ano-eleitoral. Acesso em: 19 fev. 2026.