REVISTA DIREITO FAE

DEEPFAKE COMO DESAFIO JURÍDICO E SOCIAL ANÁLISE JURÍDICO COMPARADA EM PERSPECTIVA INTERNACIONAL

DEEPFAKE AS A LEGAL AND SOCIAL CHALLENGE A COMPARATIVE INTERNATIONAL ANALYSIS

***Ana Carolina Fukuha*[1]**
***Maria Eduarda Fürst*[2]**
***João Victor Archegas*[3]**

RESUMO

O avanço das tecnologias de inteligência artificial, especialmente aquelas voltadas à geração de conteúdos audiovisuais sintéticos como as *deepfakes*, tem imposto novos desafios ao Direito e à sociedade. Ao manipular de forma hiper-realista imagens, vozes e vídeos, tais ferramentas suscitam riscos relevantes à privacidade, à honra, à integridade informacional e à própria estabilidade democrática. O presente trabalho tem como objetivo analisar o fenômeno das *deepfakes* sob uma perspectiva jurídico-comparada, identificando impactos sociais e riscos jurídicos, bem como avaliando as iniciativas normativas em curso no Brasil, na União Europeia, nos Estados Unidos e na China. A pesquisa, de natureza qualitativa, fundamenta-se em revisão bibliográfica e documental de fontes acadêmicas, legislativas e institucionais. A análise contempla não apenas os riscos associados às *deepfakes*, mas também a forma como sua regulação pode afetar direitos fundamentais, em especial a privacidade, a proteção de dados, a liberdade de expressão e a tutela da imagem e da personalidade. Os resultados evidenciam que, embora haja esforços significativos em diferentes jurisdições — como o *AI Act* europeu, as normas chinesas de síntese profunda e os projetos legislativos brasileiros — ainda

[1] Bacharel em Direito FAE – Centro Universitároio. Contato: anafukuha3@gmail.com.

[2] Bacharel em Direito FAE – Centro Universitároio. Contato: miafurst.contato@gmail.com.

[3] Professor de Direito na FAE e Coordenador no ITS Rio. Mestre e Doutorando em Direito pela UFPR e Master of Laws pela Harvard Law School, onde foi Gammon Fellow de excelência acadêmica. Contato: joao.archegas@fae.edu.

persistem lacunas regulatórias quanto à transparência, à responsabilização de provedores e à proteção da identidade digital. Conclui-se que a regulação das *deepfakes* exige um equilíbrio entre inovação e proteção de direitos fundamentais, com ênfase em medidas preventivas, educação digital e governança cooperativa entre Estados, plataformas e sociedade civil.

Palavras-chave: Análise comparativa. *deepfake*. Direitos fundamentais. Inteligência artificial. Regulação jurídica.

ABSTRACT

The advancement of artificial intelligence technologies aimed at generating synthetic audiovisual content, such as deepfakes, has posed new challenges to the legal system and society. By hyper-realistically manipulating images, voices, and videos, these tools give rise to significant risks to privacy, honor, informational integrity, and even democratic stability. This study aims to analyze the deepfake phenomenon from a comparative legal perspective, identifying social impacts and legal risks, as well as assessing ongoing regulatory initiatives in Brazil, the European Union, the United States, and China. The present study is grounded in a meticulous review of bibliographic and documentary sources, encompassing academic, legislative, and institutional materials. The research addresses not only the risks associated with deepfakes, but also the ways in which their regulation may affect fundamental rights, particularly privacy, data protection, freedom of expression, and the safeguarding of image and personality rights. The findings show that, although significant efforts exist across different jurisdictions — such as the European Union's *AI Act*, China's deep synthesis regulations, and Brazilian legislative proposals — regulatory gaps remain regarding transparency, provider accountability, and the protection of digital identity. The study concludes that deepfake regulation requires a balance between innovation and the protection of fundamental rights, with emphasis on preventive measures, digital education, and cooperative governance among States, platforms, and civil society.

Keywords: Artificial intelligence. Comparative analysis. *deepfake*. Digital regulation. Fundamental rights.

1 INTRODUÇÃO

A ascensão das tecnologias baseadas em inteligência artificial (IA) tem provocado transformações profundas nas formas de produção, circulação e consumo de conteúdo no ambiente digital. Entre essas inovações, destaca-se a *deepfake*, tecnologia que permite a criação de áudios, vídeos ou imagens hiper-realistas por

meio de técnicas como *machine learning* e redes neurais profundas — especialmente as chamadas *Generative Adversarial Networks* (GANs). Esses sistemas são capazes de simular com extrema fidelidade expressões faciais, entonações vocais e movimentos corporais, produzindo conteúdos muitas vezes indistinguíveis da realidade perceptível a olho nu (CHADHA *et al*., 2021).

Na chamada sociedade em rede, a informação tornou-se não apenas um recurso estratégico, mas também um instrumento de poder, dominação e disputa simbólica — como observou Castells (2015) ao apontar que os fluxos comunicacionais, mediados por tecnologias digitais, moldam realidades sociais, políticas e econômicas. Nesse contexto, as *deepfake*s emergem como artefatos de manipulação da verdade, capazes de desestabilizar narrativas coletivas, corroer a confiança pública e intensificar a desinformação. Como afirmam Chesney e Citron (2019), trata-se de uma tecnologia que pode subverter os pilares epistêmicos da democracia ao comprometer a confiabilidade do que vemos e ouvimos, erodindo a "realidade consensual" que sustenta o discurso público.

Ainda que possuam aplicações legítimas — como preservação histórica, reconstituição de vozes extintas, entretenimento e inovação educacional —, as *deepfake*s têm sido instrumentalizadas de forma maliciosa, desafiando a integridade da imagem pessoal, a soberania digital e a própria governança informacional. Gambín *et al*. (2024) observam que a sofisticação dos algoritmos atuais permite não apenas a manipulação de traços físicos, mas também a fabricação de situações inteiras, com potencial de impacto militar, político ou econômico. A problemática se intensifica na era da pós-verdade, em que emoções e crenças pessoais frequentemente se sobrepõem aos fatos objetivos. Siqueira e Andrade (2024) apontam que a manipulação audiovisual alimenta a polarização política e desestabiliza processos democráticos, ao passo que desafia os sistemas jurídicos tradicionais, ainda pouco preparados para lidar com a fluidez e a viralidade desses conteúdos.

A gravidade do fenômeno torna-se evidente em episódios recentes no Brasil: entre novembro de 2023 e março de 2024, mais de 60 adolescentes — meninas entre 13 e 16 anos — foram vítimas da criação e disseminação de conteúdos pornográficos sintéticos produzidos por colegas de escola. O caso, amplamente divulgado pela mídia, revelou lacunas normativas na proteção da imagem digital e na prevenção da violência virtual (G1, 2023). Situações como essa evidenciam a urgência de respostas institucionais capazes de enfrentar as múltiplas camadas do problema — técnica, jurídica, ética e educacional.

Diante desse cenário, a presente pesquisa propõe-se a investigar o fenômeno das *deepfake*s sob duas frentes analíticas complementares: (i) os impactos sociais, jurídicos e políticos decorrentes da disseminação de conteúdos audiovisuais manipulados; e (ii) as iniciativas regulatórias implementadas no Brasil e em ordenamentos estrangeiros, como a União Europeia, os Estados Unidos e a China. A partir de uma perspectiva comparada, o objetivo geral é analisar como diferentes sistemas jurídicos têm buscado responder aos desafios normativos impostos pela IA generativa, especialmente quanto à sua aplicação na adulteração de conteúdos audiovisuais.

De forma mais específica, busca-se identificar os fundamentos técnicos e os usos aplicáveis das tecnologias de *deepfake*, a fim de examinar os impactos sociais e os riscos jurídicos decorrentes do uso indevido dessa tecnologia, bem como mapear os desafios normativos enfrentados pelo direito brasileiro e pelo direito internacional diante desse fenômeno. Por fim, pretende-se comparar as legislações da União Europeia, dos Estados Unidos e da China no tocante à regulamentação das *deepfake*s, avaliando as perspectivas para uma governança digital eficiente, pautada em educação, ética e responsabilidade informacional.

A relevância deste estudo reside em sua atualidade e pertinência: as tecnologias de IA generativa remodelam os fluxos comunicacionais globais e desafiam os limites entre o real e o fabricado. O debate sobre a regulação das *deepfake*s

transcende o campo jurídico e alcança dimensões fundamentais como privacidade, liberdade de expressão, segurança informacional, confiabilidade institucional e autodeterminação informacional (KUGLER e PACE, 2021).

A base teórica que sustenta esta pesquisa foi construída com rigor metodológico, a partir da seleção criteriosa de artigos científicos, documentos normativos, resoluções eleitorais, projetos de lei e marcos legais internacionais. Essa pluralidade — ancorada em áreas como Direito, Comunicação, Ciência Política e Tecnologia — permite uma abordagem verdadeiramente interdisciplinar, indispensável para compreender um fenômeno de natureza híbrida, situado entre o ciberespaço e o campo jurídico. A estrutura deste artigo foi organizada a partir de uma revisão de literatura distribuída em cinco eixos temáticos interdependentes: fundamentos técnicos da *deepfake*; usos lícitos e potenciais aplicáveis; impactos sociais e riscos jurídicos; desafios normativos e limitações do direito vigente; e modelos regulatórios internacionais em comparação. Cada um desses tópicos contribui para a construção de um panorama abrangente e crítico sobre o fenômeno, estabelecendo o alicerce teórico para uma análise comparada entre os marcos normativos do Brasil, da União Europeia, dos Estados Unidos e da China. Essa abordagem busca examinar a evolução legislativa, a abrangência das normas, os mecanismos de *enforcement* e as estratégias de prevenção adotadas em cada contexto, com o objetivo de identificar boas práticas e eventuais lacunas. Ao final, pretende-se oferecer subsídios concretos para o aprimoramento da regulação brasileira sobre o uso de inteligência artificial na criação e circulação de conteúdos audiovisuais manipulados.

2 FUNDAMENTOS TÉCNICOS DA *DEEPFAKE*

A *deepfake* representa um dos avanços mais notáveis no campo da inteligência artificial (IA), particularmente nas subáreas do aprendizado de máquina

(machine learning) e do aprendizado profundo (deep learning). O termo surge da junção entre "*deep learning*" e "*fake*", evidenciando sua

essência: trata-se da produção de conteúdo audiovisual sintético — como vídeos, áudios e imagens — gerado ou alterado por modelos computacionais sofisticados, com a intenção de simular a realidade de maneira verossímil e, muitas vezes, imperceptível aos sentidos humanos.

Do ponto de vista técnico, o fundamento central das *deepfake*s reside nas Redes Geradoras Adversariais (Generative Adversarial Networks – GANs), inicialmente propostas por Ian Goodfellow em 2014. Estas redes funcionam com base em dois sistemas neurais em constante confronto: o gerador, responsável por criar imagens ou sons sintéticos, e o discriminador, encarregado de avaliar se o conteúdo gerado é real ou falso. A interação contínua entre ambos resulta na progressiva melhoria do conteúdo produzido, que tende a se tornar indistinguível do material autêntico (GAMBÍN *et al*., 2024).

Outras arquiteturas complementares também desempenham papel crucial na composição de *deepfakes* de alta qualidade. Modelos baseados em autoencoders, redes convolucionais profundas (*Convolutional Neural Networks – CNNs*), algoritmos de decisão e técnicas de face *reenactment* são amplamente utilizados na criação de vídeos em que há substituição facial, sincronização labial e imitação vocal (AMERINI *et al.,* 2025; PATEL *et al.,* 2023; KAUR *et al*., 2024). A combinação desses métodos com bancos de dados extensos — geralmente compostos por milhares de imagens ou trechos de voz

— viabiliza a criação de manipulações extremamente convincentes.

Embora no início essas tecnologias estivessem restritas a centros de pesquisa e grandes corporações, atualmente sua acessibilidade se ampliou drasticamente. A popularização de softwares de código aberto e o barateamento dos recursos computacionais permitiram que usuários comuns, sem formação técnica especializada, passassem a produzir *deepfakes* com relativa facilidade. Aplicativos

como Zao, Reface, FaceSwap, Descript, entre outros, disponibilizam ferramentas intuitivas capazes de gerar vídeos manipulados em poucos minutos (WESTERLUND, 2019; SIQUEIRA; ANDRADE, 2024).

Esse cenário é impulsionado por uma cultura digital baseada na hipervisibilidade e na viralização de conteúdo. Como observam Twomey *et al*. (2025), a disseminação de *deepfakes* não é apenas um produto técnico, masum reflexo de uma dinâmica comunicacional marcada pela instantaneidade, pela lógica algorítmica e pela busca de engajamento. Tal contexto agrava os riscos de uso indevido, sobretudo na manipulação de dados biométricos como voz e imagem, que compõem o núcleo dos direitos da personalidade. Paralelamente ao avanço dos métodos de geração, desenvolvem-se técnicas de detecção e análise forense digital, com o objetivo de identificar sinais de falsificação. Tais métodos incluem a detecção de inconsistências de iluminação, imperfeições de piscar de olhos, distorções faciais ou discrepâncias no espectro sonoro (RANA *et al*., 2022).

Ainda assim, como destaca Westerlund (2019), o progresso das técnicas de síntese supera frequentemente o das técnicas de detecção, criando um cenário de constante corrida tecnológica entre criadores e investigadores. Trata-se de uma missão particularmente desafiadora, visto que, ao identificarem marcadores capazes de indicar conteúdos distorcidos ou gerados por inteligência artificial, os pesquisadores acabam por fornecer subsídios que podem ser utilizados pelos próprios desenvolvedores desses sistemas para o aperfeiçoamento de seus modelos. Estes, por sua vez, incorporam tais informações, tornando os marcadores anteriormente eficazes obsoletos em curto espaço de tempo. Essa dinâmica revela a complexidade inerente à regulação e ao controle das *deepfakes*, exigindo soluções técnicas e normativas em constante aperfeiçoamento.

Esse ciclo de retroalimentação revela um aspecto fundamental do fenômeno: os fundamentos técnicos da *deepfake* não são neutros. Chesney e Citron (2019) ressaltam que a sofisticação da tecnologia, aliada à facilidade de uso e à ausência de

barreiras normativas eficazes, amplia o risco de disseminação maliciosa e de manipulação da realidade de forma deliberada. Por essa razão, o entendimento técnico da tecnologia deve ser acompanhado de uma reflexão ética e normativa[4].Em suma, esta tecnologia representa uma convergência de inovação técnica, potencial de impacto social e complexidade regulatória. Seu domínio requer não apenas conhecimento sobre algoritmos e arquitetura de redes neurais, mas também uma abordagem interdisciplinar, capaz de compreender os efeitos sociopolíticos da manipulação da realidade. Conforme destacam Kugler e Pace (2021), o Direito se vê diante de um dilema estrutural: como proteger os direitos fundamentais sem sufocar o potencial transformador da IA?

2.1 *DEEPFAKE* E DIREITOS FUNDAMENTAIS: USOS LÍCITOS E POTENCIAIS

Embora a tecnologia *deepfake* esteja frequentemente associada à desinformação e ao uso malicioso, é imperioso reconhecer seu potencial de inovação e contribuição social quando aplicada de forma ética, legal e supervisionada. Tais aplicações lícitas já vêm sendo desenvolvidas em campos como cultura, arte, educação, marketing e comunicação, demonstrando que os mesmos algoritmos capazes de fabricar mentiras podem também ser instrumentos de criatividade, inclusão e preservação da memória. Nesse contexto, não se pode perder de vista que o uso legítimo da tecnologia deve sempre ser ponderado à luz dos direitos fundamentais, como a proteção da privacidade e dos dados pessoais, o direito à

[4] CHESNEY, Robert; CITRON, Danielle Keats. Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, v. 107, n. 1, p. 1753-1819, 2019. p. 1753. Os autores alertam que: *"Harmful lies are nothing new. But the ability to distort reality has taken an exponential leap forward with 'deep fake' technology. This capability makes it possible to create audio and video of real people saying and doing things they never said or did. Machine learning techniques are escalating the technology's sophistication, making deep fakes ever more realistic and increasingly resistant to detection. Deep-fake technology has characteristics that enable rapid and widespread diffusion, putting it into the hands of both sophisticated and unsophisticated actors."*

imagem e à personalidade, bem como a liberdade de expressão e suas limitações. Trata-se, portanto, de um campo que exige constante equilíbrio entre inovação tecnológica e garantias constitucionais.

Ressalta-se, contudo, a existência de um debate terminológico relevante acerca dessa temática. O termo " *deepfake*" decorre da junção entre "*deep learning*" e "*fake*", tendo sido inicialmente associado à criação de vídeos manipulados com propósitos antiéticos e maliciosos, especialmente no contexto de pornografia não consensual (AJDER, 2023)[5]. Esse histórico de origem contribuiu para que o vocábulo adquirisse uma conotação predominantemente negativa, frequentemente vinculado à manipulação da verdade e a práticas lesivas. Em razão disso, alguns especialistas sustentam que o termo " *deepfake*" deveria ser reservado apenas às aplicações prejudiciais da tecnologia, recomendando o uso de expressões alternativas — como "inteligência artificial", "mídia sintética gerada por IA" ou "síntese audiovisual baseada em IA" — para designar seus usos lícitos e socialmente construtivos (AJDER, 2023; MIT Sloan, 2023).

Uma provocação pertinente é se, diante de um uso ético e positivo, ainda seria apropriado classificar a tecnologia como *deepfake* ou se não se trataria, simplesmente, de uma aplicação benéfica da inteligência artificial. Além disso, observa-se que diferentes jurisdições adotam entendimentos diversos sobre a definição e o uso do termo, o que evidencia a complexidade terminológica envolvida. Nos Estados Unidos, por exemplo, legislações estaduais apresentam definições distintas de *deepfake*, dificultando a uniformização jurídica (GW Law, 2024; Tech Policy Press, 2023). Na Europa, por sua vez, o Artificial Intelligence Act propõe uma definição mais abrangente, considerando como *deepfake* qualquer conteúdo sintético

[5] "*O termo deepfake surgiu em 2017 quando um usuário do Reddit usou o apelido 'deepfakes' para postar vídeos pornográficos alterados digitalmente com imagens de celebridades. A tecnologia foi aplicada usando como base inúmeras imagens e vídeos de celebridades para aprender a imitar as expressões faciais e sobrepor em um vídeo o rosto de uma celebridade no rosto de atrizes de filmes pornô*" (MOLINA; BERENGUEL, 2022, p. 2).

que simula a realidade, independentemente da intenção de seu criador (ACIG Journal, 2023). Diante desse cenário plural e ainda em formação, opta-se, no presente trabalho, por empregar o termo " *deepfake*" em sentido amplo, abrangendo tanto suas manifestações negativas quanto aquelas positivas, com vistas a oferecer uma análise jurídica mais completa sobre os desafios regulatórios da tecnologia.

2.2 *DEEPFAKE*S NA CULTURA, ARTE E ENTRETENIMENTO

Um dos usos mais emblemáticos e legítimos das *deepfake*s ocorreu em 2019 no Museu Dalí, localizado na Flórida (EUA), que utilizou essa tecnologia para recriar digitalmente o artista Salvador Dalí. A proposta possibilitou ao público interagir com a figura do pintor, em uma experiência imersiva que aliava memória histórica e inovação tecnológica (WESTERLUND, 2019). A iniciativa foi amplamente elogiada por seu caráter educativo, ao aproximar visitantes da história da arte por meio de recursos interativos.

De forma semelhante, a indústria cinematográfica tem recorrido às *deepfake*s para rejuvenescer atores ou recriar personagens falecidos. É o caso do filme Star Wars: Rogue One (2016), no qual o ator Peter Cushing foi digitalmente inserido por meio de algoritmos de IA (RAMOS, 2022). Um exemplo notório da aplicação de técnicas de clonagem de voz baseadas em inteligência artificial é a aparição do personagem Luke Skywalker rejuvenescido na série The Mandalorian. Para esse feito, a voz de um Luke mais jovem foi sintetizada a partir da análise de entrevistas e transmissões de rádio antigas de Mark Hamill, permitindo a aplicação conjunta de recursos de *deepfake* e clonagem vocal com alto grau de realismo (RESPEECHER, 2022)[6]. A aplicação de *deepfake*s para recriar performances vocais também tem sido

[6] Segundo a empresa Respeecher, "a clonagem de voz traz oportunidades sem precedentes para novos talentos e criadores de conteúdo, permitindo que eles equilibrem melhor sua carga de trabalho entre atores altamente requisitados e aqueles que não enfrentam essa demanda intensa. O Respeecher transformou a maneira como as produções lidam com ADR e dublagem, economizando

explorada em experiências performáticas, como simulações de artistas interpretando obras em estilos distintos ou idiomas diversos, com finalidade educativa ou memorial (SIQUEIRA; ANDRADE, 2024)[7].

2.3 POTENCIAL CIENTÍFICO E EDUCACIONAL

O campo educacional figura entre os mais promissores para a aplicação ética e inovadora das *deepfake*s, especialmente no que diz respeito à inclusão, personalização da aprendizagem e preservação histórica. A tecnologia pode ser empregada na criação de vídeos educativos adaptados para públicos com diferentes perfis cognitivos ou com deficiência auditiva e visual, por meio de recursos como sincronização labial automatizada, legendas dinâmicas e interpretação em linguagem de sinais (TWOMEY *et al*., 2025).

Adicionalmente, o uso da inteligência artificial pode contribuir para revitalizar o ensino de ciências humanas e sociais, a partir da criação de vídeos interativos nos quais figuras históricas, como cientistas, escritores e líderes políticos, apresentam suas ideias em linguagem acessível. Essa estratégia facilita o engajamento dos alunos e a compreensão de conceitos complexos (PATEL *et al*., 2023).

Um exemplo interessante de aplicação memorial é o projeto do *Illinois Holocaust Museum and Education Center*, que, em 2018, desenvolveu um módulo de entrevistas em holograma, no qual os visitantes podiam interagir com sobreviventes do Holocausto. Embora, à época, a tecnologia utilizada não se baseasse em

tempo para os atores e dinheiro para as equipes de produção". (RESPEECHER. Respeecher synthesized younger Luke Skywalker's voice for Disney's Mandalorian, 2022. Disponível em: https://www.respeecher.com/case-studies/respeecher-synthesized-younger-luke-skywalkers-voi ce-disneys-mandalorian. Acesso em: 7 maio 2025).

[7] Esse tipo de uso tem suscitado inúmeros debates no campo do direito autoral, especialmente quanto à necessidade de autorização prévia para o uso da voz e da imagem de artistas falecidos ou vivos. Discute-se se a reprodução sintética de uma performance constitui violação de direitos patrimoniais e morais do autor ou intérprete, ou se se enquadra como nova obra derivada.

*deepfake*s, sua integração posterior com ferramentas de IA permitiria ampliar significativamente a interatividade e o realismo da experiência. Conforme apontam Magno e Magela Pieroni (2024), com a incorporação de *deepfake*s, “essa experiência ganharia novos contornos hoje em dia”, permitindo respostas mais fluídas, sincronização labial realista e maior imersão narrativa.

Karnouskos (2020) acrescenta que, em ambientes de realidade aumentada e virtual, as *deepfake*s têm contribuído para o desenvolvimento de simulações médicas, treinamentos corporativos e jogos imersivos, permitindo que avatares realistas representem pacientes, professores ou instrutores em experiências controladas. Tais recursos são especialmente valiosos em cenários de aprendizado profissional, como o treinamento de médicos e pilotos, nos quais a simulação precisa da realidade é essencial para a formação técnica segura. Um exemplo marcante no que tange ao setor de saúde, foi quando a tecnologia foi utilizada na campanha “Malaria Must Die”, em que o ex-jogador David Beckham teve sua voz manipulada para transmitir mensagens em nove idiomas diferentes (WESTERLUND, 2019).

Além disso, como destacam Siqueira e Andrade (2024), o uso de vídeos sintéticos com personalização de linguagem e estilo de comunicação pode facilitar o acesso ao conteúdo para estudantes com dislexia, deficiência auditiva ou transtornos de aprendizagem, promovendo equidade educacional. As *deepfake*s, nesse contexto, devem ser compreendidas como ferramentas que potencializam a democratização do conhecimento, desde que aplicadas com responsabilidade, consentimento e supervisão adequada.

2.4 COMUNICAÇÃO E PUBLICIDADE

A neutralidade técnica da tecnologia *deepfake* permite sua aplicação para fins construtivos e comerciais em diversos setores, como saúde, entretenimento, turismo e, notadamente, marketing (WESTERLUND, 2019). A capacidade de personalizar

conteúdo audiovisual com rostos, vozes e expressões corporais adaptáveis ao público-alvo tem despertado o interesse de empresas que desejam aumentar o engajamento de suas campanhas e aproximar a comunicação institucional das preferências do consumidor.

Nesse sentido, Caporusso (2020) apresenta uma aplicação que utiliza algoritmos de *deepfake* para construir um "gêmeo digital interativo" — um modelo acurado de uma pessoa real, que pode ser usado como substituto em simulações, apresentações, histórias interativas ou experiências personalizadas. A proposta visa permitir que os próprios usuários criem réplicas digitais para usos benignos, como eventos memoriais, simulações de presença e reencenações afetivas. Esse tipo de ferramenta possui especial relevância em ações de marketing emocional e storytelling digital, em que a presença simbólica ou afetiva do indivíduo é central.

Além disso, conforme analisado por Kwok e Koh (2021), a tecnologia *deepfake* tem sido explorada na indústria do turismo e do marketing de destinos, possibilitando a inserção de figuras públicas — reais ou sintéticas — em ambientes virtuais de interação com o público. Um exemplo concreto ocorreu no Japão, onde avatares digitais de celebridades foram utilizados para promover experiências turísticas imersivas. Já na indústria musical, o retorno do grupo ABBA em 2022 foi marcado por um concerto com versões digitais dos integrantes originais, construídas por técnicas de motion capture e IA, numa combinação de inovação tecnológica e nostalgia cultural (ABBA, 2021).

No contexto de redes sociais, a tecnologia também tem sido incorporada em aplicativos como FaceApp e Facebrity, permitindo a geração de vídeos virais personalizados para entretenimento (WESTERLUND, 2019). Embora essas práticas estejam majoritariamente voltadas ao lazer, elas revelam a força das *deepfake*s na economia da atenção, na qual conteúdo altamente personalizado tende a alcançar maior disseminação e engajamento.

Contudo, o uso ético dessa tecnologia no marketing requer transparência, consentimento dos envolvidos e rotulagem adequada dos conteúdos sintéticos. Como alertam Kugler e Pace (2021), a linha entre personalização e manipulação pode ser tênue, especialmente quando o consumidor não é informado de que está interagindo com conteúdo gerado artificialmente. Assim, ainda que legítima, essa aplicação exige a criação de normas específicas de transparência publicitária e de proteção de dados sensíveis, com vistas à preservação da confiança do consumidor e à tutela dos direitos da personalidade. Como sintetiza Amerini *et al*. (2025), a linha entre realidade e ficção nunca foi tão tênue — e tão manipulável.

2.5 CONSIDERAÇÕES ACERCA DO USO LÍCITO

Diante do amplo espectro de aplicações legítimas da tecnologia *deepfake*, torna-se fundamental que o ordenamento jurídico brasileiro reconheça tais possibilidades na formulação de uma regulação adequada. A promoção de usos benéficos, aliados a medidas de supervisão e transparência, deve ser tão central quanto a prevenção dos usos ilícitos. Como bem pontuam Siqueira e Andrade (2024), uma governança eficaz do ciberespaço deve combinar ações repressivas e preventivas, com forte investimento em educação digital e cultura de responsabilidade informacional. Conforme os autores:

> [...] numa perspectiva macro, pode-se discutir o fundamental papel do Estado não sendo apenas um veículo que proporciona a expansão e revisão dos ordenamentos de maneira eficaz e condizente com as necessidades temporais, mas também um agente educacional e informacional, combatendo este fenômeno em todas as suas esfera de atuação, implementando vigilâncias rigorosas que alcancem as plataformas informacionais, concedendo ao usuário a autonomia ao dispor de seus dados, promovendo segurança preventiva, não só reparatória (SIQUEIRA & ANDRADE, 2024, p. 24).

Nesse cenário, é indispensável destacar que a regulação das *deepfake*s deve se orientar também pela tutela de direitos fundamentais. Em primeiro lugar, o direito à privacidade e à proteção de dados pessoais, consagrado no art. 5º, X e XII da Constituição Federal e regulamentado pela Lei Geral de Proteção de Dados (Lei n. 13.709/2018), impõe limites claros ao uso de informações biométricas, imagens e vozes, exigindo consentimento informado e transparência. Da mesma forma, o direito à imagem e à personalidade deve ser resguardado para impedir que identidades sejam exploradas sem autorização, mesmo em contextos aparentemente positivos.

A liberdade de expressão e a liberdade artística, igualmente asseguradas pela Constituição (art. 5º, IV e IX), permitem a utilização da tecnologia em paródias, sátiras, produções culturais e manifestações críticas. No entanto, tais liberdades não são absolutas, encontrando limites quando colidem com a dignidade da pessoa humana e a proteção contra abusos. Por fim, a responsabilização de provedores e plataformas, regulada pelo Marco Civil da Internet (Lei n. 12.965/2014), é peça-chave para equilibrar inovação e segurança, demandando modelos de moderação que respeitem direitos individuais sem cercear indevidamente a inovação tecnológica[8] [5].

como interagimos com a arte, a educação, a cultura e a memória. Quando guiada por princípios éticos, transparência e responsabilidade, a inteligência artificial pode se tornar aliada poderosa na preservação de legados históricos, na

[8] O Supremo Tribunal Federal, ao julgar os Recursos Extraordinários n.º 1.037.396 (Tema 987) e n.º 1.057.258 (Tema 533), reconheceu a inconstitucionalidade parcial e progressiva do art. 19 da Lei n.º 12.965/2014 (Marco Civil da Internet). A Corte entendeu que o modelo de responsabilidade das plataformas digitais, baseado exclusivamente na exigência de ordem judicial prévia para remoção de conteúdo, não garante proteção suficiente a direitos fundamentais e à democracia. Até que sobrevenha nova legislação, o dispositivo deve ser interpretado conforme a Constituição, permitindo a responsabilização civil dos provedores em casos de crimes ou atos ilícitos quando, notificados extrajudicialmente, deixarem de remover o conteúdo, salvo nos crimes contra a honra, em que se mantém a exigência de ordem judicial. A decisão também estabeleceu deveres de cuidado proativo para prevenir a circulação de conteúdos gravemente ilícitos, como terrorismo, pornografia infantil, discurso de ódio, crimes contra mulheres e atos antidemocráticos, cuja omissão pode gerar responsabilidade por falha sistêmica (STF, RE 1.037.396 e RE 1.057.258, Rel. Min. Dias Toffoli e Min. Luiz Fux, j. 26 jun. 2025).

democratização do conhecimento e na expansão das experiências sensoriais e comunicativas. Governar a tecnologia não significa estagná-la, mas orientá-la para que contribua ativamente com os valores humanos e constitucionais que se pretende proteger. Promover o uso lícito da *deepfake* é, portanto, não apenas uma medida de proteção jurídica, mas também um ato de confiança na capacidade da inovação de servir ao bem comum.

2.6 IMPACTOS SOCIAIS E RISCOS JURÍDICOS

Apesar de suas aplicações promissoras, a tecnologia *deepfake* impõe uma série de impactos negativos que atravessam dimensões sociais, jurídicas, políticas e éticas. A produção de conteúdos audiovisuais manipulados de forma hiper-realista coloca em xeque os fundamentos da comunicação digital e gera riscos significativos à integridade democrática, aos direitos da personalidade e à própria noção de verdade no ambiente informacional contemporâneo (CHESNEY; CITRON, 2019).

2.7 DESINFORMAÇÃO E MANIPULAÇÃO DA OPINIÃO PÚBLICA

A disseminação de conteúdos manipulados por inteligência artificial, especialmente *deepfake*s, representa uma ameaça concreta à confiabilidade da informação, à formação da opinião pública e à integridade dos processos democráticos e eleitorais. Ao distorcer visual e auditivamente a realidade, essas tecnologias ampliam a capacidade de desinformação em larga escala, não apenas pelo seu impacto narrativo, mas também pela crescente acessibilidade das ferramentas envolvidas. A popularização de aplicativos, modelos de código aberto e plataformas intuitivas permite que mesmo indivíduos sem conhecimento técnico avançado produzam conteúdos altamente realistas, o que transforma qualquer usuário comum em um potencial “desinformante de elite”.

Ramos (2022) relata dois episódios emblemáticos que ilustram o uso de *deepfakes* como instrumento de engenharia social. O primeiro, de repercussão nacional, envolveu a circulação de um vídeo pornográfico, nas redes sociais durante o pleito eleitoral de 2018, no qual o então governador de São Paulo, João Doria, foi apontado, ainda que sem qualquer comprovação técnica, como um dos supostos envolvidos no vídeo. O segundo ocorreu no cenário internacional e consistiu na veiculação de um vídeo manipulado em que o ex-presidente Barack Obama teria proferido declarações ofensivas ao presidente Donald Trump. Segundo a autora:

> [...] "Um caso conhecido no Brasil envolveu João Doria, Governador de São Paulo, inclusive candidato à Presidência na eleição que se avizinha. Durante o pleito de 2018, circularam nas redes sociais um vídeo de sexo explícito entre seis mulheres e um homem, supostamente Doria. Outro episódio, que gerou repercussão a nível internacional, ocorreu com a divulgação de um vídeo, veiculado no YouTube, em que Barack Obama teria dito que 'o Presidente Trump é um total e completo imbecil'. O rosto, a voz e os movimentos pareciam ser de Obama, porém, tratava-se de conteúdo digital manipulado" (RAMOS, 2022, p. 59).

Mais recentemente, um caso ocorrido na Índia evidencia o uso da inteligência artificial para manipulação de percepções eleitorais de forma ainda mais sofisticada: políticos falecidos foram "revividos" digitalmente por meio de *deepfake*s, proferindo discursos de apoio a determinados candidatos durante o período eleitoral. Esses vídeos, emocionalmente carregados, exploram vínculos afetivos com figuras históricas populares e têm circulado amplamente nas redes sociais, aumentando o risco de manipulação do eleitorado em um contexto altamente sensível e polarizado (AL JAZEERA, 2024).

Esse tipo de manipulação, ao se tornar viral, molda a percepção pública de forma artificial, contribuindo para o fenômeno que Chesney e Citron (2019) denominam de *cascading information hazard*, ou "dinâmica da informação em cascata". Trata-se do ciclo em que o compartilhamento acrítico de conteúdos

manipulados, potencializado pela arquitetura das redes sociais, gera ondas de desinformação de difícil controle. A manipulação audiovisual, nesses casos, compromete a credibilidade de figuras públicas, distorce debates políticos e fragiliza instituições democráticas.

O impacto das *deepfake*s torna-se ainda mais acentuado em contextos eleitorais. Moura (2019) aponta que a desinformação digital tem sido uma constante nas campanhas políticas contemporâneas, não apenas no Brasil, mas também em países como Estados Unidos, Índia e Espanha. Em apresentação no Seminário Internacional Fake News e Eleições, promovido pelo Tribunal Superior Eleitoral (TSE), o autor destacou que, apenas no mês de maio de 2019, houve um aumento de 67% na circulação de notícias falsas no país[9]. Além disso, observou-se que grande parte da população tende a confiar mais em conteúdos compartilhados por familiares e amigos do que nas informações veiculadas pela imprensa profissional, cenário que amplia os efeitos das manipulações audiovisuais com alto potencial de convencimento e impacto comportamental.

Esse cenário revela uma preocupante inversão do eixo de confiança informacional, em que conteúdos enganosos ganham legitimidade pela via relacional, enquanto fontes jornalísticas tradicionais perdem espaço no debate público. Com isso, vídeos e áudios manipulados adquirem poder persuasivo significativo, muitas vezes decisivo para influenciar decisões eleitorais e comprometer a estabilidade democrática.

Diante dessa realidade, torna-se imperioso o desenvolvimento de mecanismos legais, tecnológicos e educacionais para mitigar os efeitos deletérios das

[9] De acordo com pesquisa apresentada por Maurício Moura, em 2019, durante o Seminário Internacional entre estudiosos brasileiros e europeus, organizado pelo TSE, somente no mês de maio daquele ano houve um incremento de 67% de divulgações de notícias falsas em todo o Brasil. Acrescenta o investigador que ‘a fake news perpassa todas as campanhas eleitorais de maneira muito forte. Outro dado que é comum aos Estados Unidos, à Índia, ao Brasil, à Espanha: as pessoas confiam mais no conteúdo recebido por familiares e amigos do que no conteúdo da imprensa tradicional” (RAMOS, 2022, p. 59).

*deepfake*s. O fortalecimento da literacia digital, o incentivo à checagem de fatos e a responsabilização de plataformas e agentes produtores de conteúdo falso figuram como medidas essenciais para a defesa do espaço público informacional.

2.8 *DEEPFAKE*S SEXUAIS E EXPOSIÇÃO INDEVIDA

O uso de tecnologias de inteligência artificial para a criação de *deepfake*s com conteúdo sexual não consentido representa uma grave violação dos direitos da personalidade, afetando diretamente a imagem, honra, privacidade e integridade psicológica das vítimas. Tais práticas configuram uma nova forma de violência digital, especialmente dirigida contra mulheres e figuras públicas.

No Brasil, entre novembro de 2023 e março de 2024, mais de 60 adolescentes foram vítimas da criação e disseminação de vídeos pornográficos falsos, produzidos por colegas de escola utilizando técnicas de *deepfake*. Esses vídeos foram amplamente compartilhados por meio de aplicativos de mensagens e redes sociais, resultando em sérias consequências emocionais e sociais para as vítimas (SIQUEIRA; ANDRADE, 2024).

Casos semelhantes têm ocorrido com figuras públicas nacionais e internacionais. Celebridades como Taylor Swift, Gal Gadot, Emma Watson e Scarlett Johansson tiveram suas imagens manipuladas digitalmente para a criação de conteúdos pornográficos falsos, que foram amplamente divulgados na internet. Mesmo sem a existência de nudez real, o impacto dessas ações é significativo, causando danos à reputação e sofrimento psicológico às vítimas (O GLOBO, 2025a). No Brasil, a atriz Isis Valverde registrou boletim de ocorrência após a circulação de imagens falsas de nudez atribuídas a ela, reforçando que o conteúdo é ilegal mesmo para quem o compartilha (UOL, 2025).

Em resposta a esses casos, a Câmara dos Deputados aprovou, em fevereiro de 2025, o Projeto de Lei nº 3821/24, que tipifica como crime a manipulação, produção

ou divulgação de conteúdos de nudez ou ato sexual falso gerados por inteligência artificial. A proposta prevê pena de reclusão de dois a seis anos, com agravantes caso a vítima seja mulher, criança, adolescente, pessoa idosa ou com deficiência, ou se o crime for cometido mediante disseminação em massa por meio de redes sociais ou plataformas digitais (CNN, 2025). A relatora do projeto, deputada Yandra Moura, destacou que o objetivo da norma é coibir o uso abusivo da tecnologia para macular a imagem das pessoas, enquanto a autora, deputada Amanda Gentil, chamou atenção para o fato de que essa prática afasta mulheres da política e reforça a exclusão dos espaços de poder (CÂMARA DOS DEPUTADOS, 2025).

A sofisticação das redes neurais que replicam rostos e vozes gera um ambiente de incerteza epistêmica, em que a distinção entre verdade e falsidade torna-se cada vez mais tênue. Casos de manipulação de vídeos com conteúdo sexual envolvendo celebridades circulam amplamente na internet, mesmo após solicitações legais de remoção (WESTERLUND, 2019). Essas situações não afetam apenas a imagem pública das vítimas, mas provocam danos emocionais severos e dificuldades jurídicas na responsabilização dos ofensores — especialmente em razão do anonimato e da transnacionalidade dos fluxos digitais (KUGLER; PACE, 2021).

2.8.1 Fraudes, Extorsões e Crimes Cibernéticos

A utilização de tecnologias de inteligência artificial, especialmente as *deepfake*s, tem sido cada vez mais empregada em práticas criminosas sofisticadas, como fraudes financeiras, extorsões e outros delitos cibernéticos. Essas tecnologias permitem a criação de conteúdos audiovisuais falsificados com alto grau de realismo, o que dificulta a detecção e amplia o potencial de dano.

Um exemplo emblemático ocorreu em Hong Kong, onde uma empresa multinacional sofreu um prejuízo de aproximadamente US$25,6 milhões após golpistas utilizarem *deepfake* para simular uma videoconferência com executivos da

empresa. Durante a reunião falsa, um funcionário foi induzido a realizar 15 transferências bancárias para contas dos criminosos, acreditando estar seguindo ordens legítimas (VALOR, 2024).

No Brasil, casos semelhantes têm sido registrados. Criminosos têm utilizado *deepfake*s para criar vídeos falsos de celebridades, como o apresentador Marcos Mion, promovendo falsas promoções e induzindo vítimas a realizar pagamentos indevidos (VERIFACT, 2025). Além disso, a prática conhecida como "sextortion" tem se intensificado, onde criminosos chantageiam vítimas com ameaças de divulgar imagens íntimas manipuladas digitalmente. Em fevereiro de 2025, a Polícia Civil do Rio Grande do Sul prendeu sete pessoas envolvidas em uma quadrilha que extorquia vítimas utilizando esse método (G1, 2025).

A sofisticação dessas práticas criminosas é evidenciada pela capacidade de replicar vozes e imagens com precisão. Damiani (2019) relata um caso em que a tecnologia foi usada para falsificar a voz de um CEO de uma empresa britânica de energia, solicitando uma transferência bancária de €220.000, que foi prontamente realizada devido à verossimilhança da voz falsificada.

Segundo a Europol (2022), o uso malicioso de *deepfake*s pode facilitar diversas atividades criminosas, incluindo assédio, extorsão, fraudes, falsificação de documentos e manipulação de identidades online. A organização destaca a necessidade urgente de adaptação dos marcos regulatórios e desenvolvimento de tecnologias de detecção para enfrentar essas ameaças emergentes.

Diante desse cenário, é imperativo que as autoridades desenvolvam estratégias integradas de prevenção, detecção e repressão a crimes cibernéticos envolvendo *deepfake*s, além de promover a conscientização pública sobre os riscos associados a essas tecnologias.

2.8.2 Uso Indevido de Identidade Digital de Figuras Públicas

A manipulação não consentida da imagem e da voz de pessoas públicas tem ganhado espaço preocupante no cenário nacional, sobretudo em razão da proliferação de conteúdos gerados por inteligência artificial. No Brasil, celebridades têm sido alvos recorrentes de vídeos e áudios falsificados com aparência verossímil, utilizados para fraudes financeiras, promoção indevida de produtos e disseminação de desinformação. Tais práticas representam uma afronta direta aos direitos da personalidade e configuram formas contemporâneas de dano moral coletivo.

Diversas personalidades relataram o uso não autorizado de suas identidades em anúncios fraudulentos nas redes sociais. O jornalista Pedro Bial, vítima de uma dessas montagens, declarou publicamente em entrevista:

> Sigo os passos lentos do processo jurídico, enquanto, na internet, a coisa só piora, se alastrando. Um crime, para se realizar, precisa de três fatores: a motivação para praticá-lo, mais os meios e oportunidades para tal. Os meios e as oportunidades são oferecidos pela Meta, que ainda lucra, ganha 'grana' com esse golpe (O GLOBO, 2024).

A apresentadora Fátima Bernardes também teve sua imagem vinculada, sem autorização, a produtos para emagrecimento, assim como o apresentador Tiago Leifert, que foi inserido em anúncios de loterias e medicamentos falsos. Ambos negaram veementemente qualquer relação com as campanhas em questão, reafirmando a inexistência de consentimento e os riscos de credibilidade associados à prática (O GLOBO, 2024).

O fenômeno se intensificou ao longo de 2025, atingindo nomes como Neymar, Paolla Oliveira, Anitta e William Bonner, cujas imagens e vozes foram manipuladas digitalmente em vídeos utilizados para fins escusos. Em alguns casos, as simulações envolveram conteúdos de teor sexual ou promessas de ganhos financeiros rápidos, com objetivo claro de enganar consumidores ou constranger publicamente os envolvidos (O GLOBO, 2025; UOL, 2025).

Esses episódios escancaram não apenas o potencial disruptivo das *deepfake*s quando empregadas de forma maliciosa, mas também a lacuna regulatória existente quanto à tutela da identidade digital no Brasil. Embora a legislação brasileira contemple dispositivos de proteção à imagem e à honra no plano civil (art. 5º, X, da Constituição Federal e arts. 20 e 21 do Código Civil), bem como normas voltadas à proteção de dados pessoais por meio da LGPD (Lei nº 13.709/2018), ainda são incipientes os mecanismos normativos voltados especificamente à transparência no uso de mídias sintéticas e ao controle sobre ferramentas de clonagem digital. A ausência de obrigações claras quanto à rotulagem de conteúdos gerados por inteligência artificial, somada à ampla disponibilidade e circulação irrestrita dessas tecnologias, dificulta a contenção de seus efeitos nocivos e a construção de uma cultura de uso ético e informado. Nesse contexto, mais do que a responsabilização punitiva ex post, é fundamental que o ordenamento jurídico avance em medidas preventivas e estruturantes, como o incentivo à transparência, à educação digital e à regulação da acessibilidade às tecnologias de manipulação audiovisual.

2.9 DESAFIOS NORMATIVOS E LIMITAÇÕES DO DIREITO VIGENTE

A crescente sofisticação das tecnologias de inteligência artificial, especialmente as *deepfake*s, desafia a efetividade dos marcos jurídicos tradicionais. O ordenamento jurídico brasileiro, assim como diversos outros ao redor do mundo, ainda carece de um arcabouço normativo robusto e específico capaz de enfrentar os múltiplos riscos e impactos sociais associados à manipulação audiovisual sintética (SIQUEIRA; ANDRADE, 2024).

No Brasil, as normas que tangenciam o uso de *deepfake*s encontram-se dispersas, geralmente vinculadas a marcos jurídicos já existentes e voltadas a contextos específicos, como o eleitoral, a proteção de dados e a responsabilidade civil. Embora instrumentos como o Código Civil, o Marco Civil da Internet e a Lei Geral

de Proteção de Dados ofereçam alguma tutela indireta, a literatura aponta para limitações substanciais desses diplomas. O Marco Civil, por exemplo, prioriza a responsabilização apenas mediante ordem judicial, não contemplando mecanismos preventivos ou obrigações específicas de rotulagem de conteúdos sintéticos (MARTINS; LONGHI, 2024). Já a LGPD, apesar de reconhecer imagem e voz como dados sensíveis, não dispõe de instrumentos claros para responsabilizar indivíduos que criam e disseminam conteúdos falsificados (BIONI, 2021).

Nesse sentido, a atuação da Autoridade Nacional de Proteção de Dados (ANPD) também é frequentemente destacada pela doutrina como insuficiente, diante de sua estrutura limitada e da dimensão transnacional dos fluxos de desinformação (SIQUEIRA; ANDRADE, 2024). Do mesmo modo, a ausência de tipos penais específicos para *deepfake*s reforça a dependência de normas genéricas, como os crimes contra a honra ou de falsa identidade, o que, segundo Moura (2019), dificulta a adequação normativa frente à complexidade tecnológica.

Autores como Bioni (2021), Martins e Longhi (2024) e Castells (2015) enfatizam que há uma defasagem estrutural entre a velocidade da inovação tecnológica e a resposta normativa. Como ressalta Castells (2015), os sistemas políticos e jurídicos operam em tempos diferentes do avanço tecnológico, gerando um descompasso que expõe vulnerabilidades sociais e jurídicas. A literatura, portanto, converge para a necessidade de uma abordagem sistêmica que combine regulação normativa, fortalecimento institucional, estratégias educativas e articulação internacional, de modo a proteger os direitos fundamentais na era da informação automatizada.

Assim, a experiência brasileira evidencia tanto os avanços normativos quanto as limitações ainda presentes na regulação das *deepfake*s. A fragmentação legislativa, a falta de tipificação penal específica, a sobrecarga das instituições reguladoras e a dificuldade em alinhar repressão, prevenção e educação digital demonstram que ainda há um longo caminho a percorrer. É justamente diante dessas

lacunas que se revela a relevância desta pesquisa, voltada a examinar como diferentes ordenamentos jurídicos enfrentam os desafios impostos pela manipulação audiovisual sintética. A posterior análise comparativa entre Brasil, Estados Unidos, Europa e China têm como intuito identificar convergências e divergências, bem como avaliar quais soluções internacionais podem servir de inspiração para o contexto brasileiro, contribuindo para o fortalecimento da proteção de direitos fundamentais e para o desenvolvimento de uma governança democrática e eficaz da inteligência artificial.

3 PROCEDIMENTOS METODOLÓGICOS

A presente pesquisa adota uma abordagem qualitativa, de caráter exploratório e analítico-comparativo, orientada à compreensão dos riscos sociais e jurídicos associados à tecnologia *deepfake* e à avaliação crítica das respostas normativas formuladas em diferentes contextos jurídicos. O objetivo central consiste em analisar como o Brasil, a União Europeia, os Estados Unidos e a China vêm regulamentando o uso de inteligência artificial aplicada à síntese de mídias, identificando convergências, divergências e boas práticas regulatórias.

O método empregado é o bibliográfico e documental, fundamentado em três eixos de fontes: (i) literatura jurídica e multidisciplinar sobre inteligência artificial, privacidade, proteção de dados e regulação digital; (ii) documentos normativos, incluindo projetos de lei, legislações vigentes, resoluções administrativas e relatórios técnicos de organismos nacionais e internacionais; e (iii) diretrizes e relatórios institucionais de caráter técnico e político, emitidos por entidades multilaterais como ONU, UNESCO, OCDE e GPAI, que servem de base para compreender tendências regulatórias globais.

A investigação apoia-se em um corpus teórico consolidado, formado por artigos acadêmicos nacionais e internacionais, documentos oficiais e marcos

normativos que refletem a evolução da regulação da inteligência artificial. No âmbito brasileiro, foram analisados o Marco Civil da Internet (Lei nº 12.965/2014), a Lei Geral de Proteção de Dados (Lei nº 13.709/2018) e o Projeto de Lei nº 2.338/2023 (Marco Legal da IA). No plano internacional, destacam-se o *AI Act* europeu, as normas chinesas sobre *deep synthesis* e *generative AI services*, bem como legislações estaduais e municipais norte-americanas relativas à pornografia não consensual e ao uso eleitoral de *deepfake*s.

A metodologia envolve a análise hermenêutica das normas e diretrizes jurídicas, buscando compreender seus fundamentos principiológicos, finalidades e impactos práticos. De forma complementar, utiliza-se o método comparativo, com o propósito de sistematizar as informações normativas a partir de dois critérios principais: (i) a forma como cada ordenamento jurídico enfrenta os riscos sociais e jurídicos da tecnologia — em especial no tocante à desinformação política, pornografia não consensual, fraudes cibernéticas e apropriação indevida de identidade digital; e (ii) as convergências e divergências entre as soluções normativas adotadas nos diferentes contextos analisados.

A pesquisa foi conduzida durante o mês de setembro de 2025, garantindo o uso de dados atualizados até essa data. Considerando o caráter dinâmico da regulação da inteligência artificial, reconhece-se a volatilidade normativa como um desafio metodológico. Alterações recentes — como a entrada em vigor do *AI Act* em 2024, a tramitação de novos projetos de lei no Brasil em 2024 e 2025 e os decretos executivos norte-americanos sobre IA — reforçam a necessidade de constante revisão das fontes e de análise crítica de seus efeitos práticos. Assim, priorizou-se o uso de fontes oficiais, recentes e verificáveis, de modo a assegurar a validade e a relevância dos resultados.

Em síntese, o percurso metodológico adotado busca aliar profundidade teórica e utilidade prática, oferecendo uma análise crítica capaz de evidenciar as fragilidades e potencialidades da regulação brasileira frente às experiências

estrangeiras. A integração entre literatura acadêmica, documentação legislativa e análise comparada visa não apenas compreender o estado atual da regulação das *deepfake*s, mas também contribuir para o debate público e acadêmico acerca da construção de um arcabouço normativo mais justo, eficaz e protetivo dos direitos fundamentais no ambiente digital.

4 PANORAMA GLOBAL DA REGULAMENTAÇÃO DA *DEEPFAKES*

As *deepfake*s consistem em mídias sintéticas — vídeos, áudios ou imagens — geradas por técnicas de inteligência artificial, capazes de reproduzir de forma hiper-realista pessoas reais executando ações ou pronunciando falas que jamais ocorreram (CHESNEY; CITRON, 2019). O uso malicioso dessa tecnologia tornou-se uma preocupação global, considerando seu potencial de aplicação em manipulações políticas, fraudes financeiras e abusos pessoais, com riscos expressivos à privacidade individual, à integridade informacional e à confiança pública nos meios digitais (SIQUEIRA; ANDRADE, 2024).

Nos últimos anos, governos e organismos internacionais têm reconhecido a necessidade de mitigar tais riscos e vêm avançando na formulação de medida segulatórias voltadas à identificação, responsabilização e prevenção de usos indevidos da inteligência artificial generativa. Em linhas gerais, essas iniciativas concentram-se em três eixos principais: (i) a promoção da transparência, mediante a rotulagem obrigatória ou marcação digital de conteúdos gerados por IA; (ii) a criminalização de usos abusivos, especialmente nos casos de *deepfake*s pornográficos não consensuais ou de manipulação em campanhas eleitorais; e (iii) a aplicação extensiva de legislações já existentes, como normas sobre difamação, fraude, proteção de dados e direito de imagem, para permitir a responsabilização civil e penal de infratores (ITU, 2025; G7, 2023).

Contudo, a regulação global das *deepfake*s tem se estruturado de modo multinível e dinâmico, combinando esforços estatais, corporativos e transnacionais. Como explica Archegas (2025), o enfrentamento dos desafios digitais não se limita à atuação estatal, mas envolve também o papel regulador das plataformas privadas, que vêm desenvolvendo mecanismos internos de governança e moderação. O caso do Oversight Board da Meta exemplifica esse movimento, ao instituir um sistema interno de freios e contrapesos para revisão das decisões de conteúdo, funcionando como instância de *accountability* dentro da própria empresa.

Na mesma linha, destaca que a regulação das redes sociais deve buscar a formação de "instituições responsáveis", dotadas de mecanismos de transparência, auditoria e prestação de contas, de modo a equilibrar liberdade de expressão e proteção contra abusos. Em vez de uma intervenção estatal direta, Balkin propõe a criação de incentivos regulatórios que induzam as plataformas a assumir parte da responsabilidade pública pela integridade informacional, tornando-se atores ativos na preservação da saúde democrática da esfera digital.

Além disso, observa que a crise de confiança nas plataformas digitais, agravada pelos episódios de desinformação de 2016 e 2018, desencadeou um movimento global de reavaliação das imunidades jurídicas tradicionais, como o safe harbor, e de fortalecimento da accountability corporativa. Essa "epifania coletiva", como descrevem os autores, marca a transição para uma nova fase da governança digital, na qual a responsabilidade é compartilhada entre Estado, empresas e sociedade civil (ARCHEGAS, 2025, apud ZITTRAIN; BOWERS, 2020).

Assim, o panorama regulatório das *deepfake*s revela uma arquitetura em constante evolução, que conjuga legislações nacionais, políticas internas de plataformas e padrões éticos e técnicos formulados por organizações internacionais. Essa abordagem multinível reflete a busca por equilíbrio entre inovação tecnológica, proteção de direitos fundamentais e segurança informacional. A seguir, serão analisadas as políticas das principais plataformas digitais no enfrentamento de

conteúdos sintéticos, bem como o papel desempenhado por organizações mundiais e padrões globais na consolidação de diretrizes para o uso responsável da inteligência artificial generativa.

4.1 POLÍTICAS DAS PLATAFORMAS DIGITAIS

As grandes empresas de tecnologia e mídia social têm implementado políticas próprias para lidar com os riscos associados aos *deepfake*s, em resposta tanto à pressão pública quanto a potenciais sanções regulatórias. Entre 2023 e 2024, a estratégia predominante consistiu em ampliar mecanismos de rotulagem e sinalização de conteúdos sintéticos, aliados a compromissos de autorregulação coletiva.

A Meta (Facebook e Instagram) anunciou, em 2024, a aplicação automática de rótulos do tipo “Made with AI” em imagens, vídeos e áudios identificados como gerados por inteligência artificial, seja por detecção interna, seja por autodeclaração do usuário. Conteúdos considerados de alto risco, como *deepfake*s políticos, recebem rótulos adicionais que contextualizam a manipulação, além de redução de alcance em casos de desinformação verificada. A empresa, entretanto, privilegia a sinalização em detrimento da remoção, exceto quando há violação de políticas mais amplas, como incitação à violência ou discurso de ódio (KLEGG, 2024).

O YouTube (Google) passou a exigir, em março de 2024, que criadores de conteúdo declarem se seus vídeos contêm material sintético de aparência realista que possa induzir o público a erro. Além disso, a plataforma se reserva o direito de rotular proativamente vídeos não declarados, sobretudo em contextos eleitorais. Também anunciou medidas para facilitar pedidos de remoção de conteúdos ilícitos, como vídeos que simulem identidades pessoais sem consentimento (SATO, 2024).

A TikTok, desde 2020, já proibia o uso malicioso de *deepfake*s. Em 2023-2024, ampliou suas regras, implementando rotulagem também em conteúdos gerados fora da plataforma e aderindo à Coalizão por Proveniência e Autenticidade de

Conteúdo (C2PA), que busca padronizar metadados que atestem a origem de mídias digitais (FORTIS, 2024).

O Twitter/X, por sua vez, mantém desde 2020 a política de “mídia sintética e manipulada”, que prevê a remoção de conteúdos enganosos com potencial de dano. Contudo, diferentemente das concorrentes, a empresa não expandiu sua política em 2023-2024 e, inclusive, retirou-se do Código de Desinformação da União Europeia, gerando críticas sobre sua abordagem mais permissiva. Outras plataformas também avançaram em políticas específicas: o Reddit atualizou suas regras para proibir pornografia *deepfake* não consensual; o LinkedIn vetou a criação de perfis falsos com uso de IA; e empresas como Microsoft e Amazon têm apoiado padrões abertos de certificação de conteúdo digital, além de investir em ferramentas de detecção de *deepfake*s (MIGUEL, 2024).

Além das medidas individuais, observa-se a formação de coalizões voluntárias. Em fevereiro de 2024, cinco grandes plataformas (Meta, Google, TikTok, Microsoft e Amazon) firmaram compromisso de cooperar no combate à desinformação gerada por IA em processos eleitorais nos EUA e na União Europeia. Em julho de 2023, sete empresas líderes em IA — entre elas OpenAI, Google, Meta e Anthropic — comprometeram-se com a Casa Branca a desenvolver mecanismos robustos de watermarking, para garantir rastreabilidade e autenticidade de conteúdos digitais (WHITE HOUSE, 2023). Essas iniciativas indicam que a autorregulação do setor privado desempenha papel essencial no enfrentamento do fenômeno das mídias sintéticas, complementando a atuação legislativa e regulatória estatal.

4.2 ORGANIZAÇÕES MUNDIAIS E PADRÕES GLOBAIS

O combate aos *deepfake*s também mobiliza organizações internacionais, que buscam coordenar padrões e diretrizes entre países. Em julho de 2025, a União Internacional de Telecomunicações (UIT/ITU) publicou um relatório recomendando

medidas mais rigorosas contra o uso malicioso dessas tecnologias, com especial atenção aos processos eleitorais iminentes em diversas jurisdições. O documento enfatizou a necessidade de ferramentas avançadas de detecção e autenticação de mídias digitais, bem como de normas técnicas globais para identificar conteúdos manipulados. A entidade destacou ainda que a ausência de um padrão universal torna a resposta fragmentada e insuficiente diante da natureza transnacional do problema (ITU, 2025).

Outras organizações de caráter normativo têm incorporado o tema em suas agendas. A UNESCO, em sua Recomendação sobre a Ética da Inteligência Artificial (2021), reconheceu os riscos dos *deepfake*s para a integridade informacional e recomendou que os Estados adotem políticas de rotulagem obrigatória e incentivem o desenvolvimento de ferramentas de verificação de autenticidade de vídeos e áudios (UNESCO, 2021). Já o G7, em sua cúpula de Hiroshima (2023), destacou a necessidade de cooperação internacional para enfrentar os usos maliciosos da IA, mencionando expressamente os *deepfake*s e defendendo a criação de códigos de conduta e a troca de informações sobre mecanismos de detecção (G7, 2023). A Parceria Global em Inteligência Artificial (GPAI) também incluiu o fenômeno em sua agenda, instituindo um grupo de trabalho dedicado à desinformação digital e às mídias sintéticas, com foco em fornecer recomendações regulatórias aos países-membros (GPAI, 2023).

Além de organismos multilaterais, ONGs e coalizões da sociedade civil têm desempenhado papel ativo na discussão. Entidades de direitos digitais alertam para a necessidade de calibrar a regulação de *deepfake*s de modo a não restringir indevidamente liberdades criativas, como a paródia e o uso artístico, enquanto pesquisadores defendem o fortalecimento do direito àimagem e a modernização das leis de propriedade intelectual como formas de garantir às pessoas maior controle sobre representações sintéticas de si mesmas. Paralelamente, grupos voltados ao combate à desinformação têm enfatizado a importância da alfabetização midiática

global, como estratégia de capacitar os cidadãos a identificar sinais de manipulação digital. Em resposta, tanto a Comissão Europeia quanto a ONU têm promovido campanhas de conscientização e lançado desafios de inovação tecnológica, como prêmios para melhores detectores de *deepfake*s (MIGUEL, 2024).

Em síntese, o panorama internacional mostra-se dinâmico e multinível, reunindo legislações pioneiras, iniciativas de autorregulação das plataformas e recomendações de organismos internacionais. O objetivo comum é promover transparência na circulação de conteúdo digital e responsabilizar os agentes que utilizam a tecnologia de forma ilícita ou enganosa, ao mesmo tempo em que se busca preservar a inovação e os direitos fundamentais.

5 MODELOSREGULATÓRIOS INTERNACIONAIS EM COMPARAÇÃO

5.1 UNIÃO EUROPEIA

A União Europeia destacou-se ao propor a primeira legislação abrangente sobre inteligência artificial em uma grande jurisdição, o Regulamento Europeu de Inteligência Artificial, conhecido como "*AI Act*". A proposta foi apresentada pela Comissão Europeia em abril de 2021 e, após intensos debates, aprovada pelo Parlamento Europeu em março de 2024 e pelo Conselho da União Europeia em maio do mesmo ano. O texto final foi publicado em 12 de julho de 2024 e entrou em vigor em 1º de agosto de 2024 em todos os 27 Estados-membros. Como regulamento, o *AI Act* possui aplicação direta, mas suas principais obrigações serão exigidas apenas após um período de transição, sendo que algumas disposições passam a valer um ano após a entrada em vigor e outras somente após dois anos, garantindo tempo para a devida adequação (ARTIFICIAL INTELLIGENCE ACT, 2024).

O *AI Act* adota uma abordagem baseada em risco, estabelecendo regras proporcionais ao nível de risco apresentado por cada sistema de inteligência artificial.O regulamento abrange tanto sistemas de IAdesenvolvidos ou utilizados dentro da União Europeia quanto aqueles fornecidos por empresas estrangeiras, dado seu efeito extraterritorial em produtos e serviços oferecidos no mercado europeu. O objetivo central é mitigar riscos à saúde, à segurança e aos direitos fundamentais dos cidadãos, ao mesmo tempo em que busca fomentar a confiança no uso da tecnologia. Para não sufocar a inovação, o regulamento também prevê medidas destinadas a reduzir encargos administrativos, sobretudo para pequenas e médias empresas (UNIÃO EUROPEIA, 2024). Nesse sentido, o documento estabelece quatro categorias principais de risco, com exigências proporcionais ao impacto potencial de cada aplicação de inteligência artificial.

Primeiramente, estão os sistemas de risco mínimo ou baixo, que incluem aplicações cotidianas como filtros de spam e jogos baseados em inteligência artificial. Esses sistemas não estão sujeitos a obrigações específicas no regulamento, sendo apenas incentivadas boas práticas voluntárias, já que representam riscos reduzidos aos usuários.

Em segundo lugar, encontram-se os sistemas de risco limitado, também chamados de risco de transparência. Essa categoria abrange aplicações que interagem diretamente com seres humanos ou que produzem conteúdos artificiais, exigindo o cumprimento de deveres de transparência. Nesses casos, o usuário deve ser claramente informado de que está interagindo com uma máquina — como ocorre com *chatbots* — ou de que determinado conteúdo foi gerado por IA, como os *deepfake*s não maliciosos, que precisam ser devidamente rotulados. Além disso, incluem-se aqui tecnologias como reconhecimento de emoções ou detecção de mentiras, que só podem ser utilizadas mediante a informação prévia e explícita aos indivíduos, permitindo-lhes consentir de forma consciente e informada.

Já em um terceiro nível, o regulamento trata dos sistemas de alto risco, aplicáveis a setores sensíveis nos quais a IA pode afetar de maneira significativa a segurança, a dignidade ou os direitos das pessoas. Exemplos incluem sistemas utilizados em diagnósticos médicos, processos de recrutamento e demissão, avaliação de crédito, educação e gestão de infraestruturas críticas. Embora seu uso seja permitido, ele está condicionado ao cumprimento de exigências rigorosas de conformidade e governança, como avaliações de risco e de qualidade dos dados, garantia de ausência de vieses nos algoritmos, documentação técnica detalhada, manutenção de registos para auditoria, fornecimento de informações claras aos usuários e supervisão humana constante. Além disso, os fornecedores devem registrar esses sistemas em um banco de dados europeu e submeter-se à fiscalização das autoridades competentes, assegurando transparência e responsabilidade no uso da tecnologia.

Por fim, o *AI Act* proíbe expressamente as aplicações classificadas como de risco inaceitável, por representarem ameaças intoleráveis à segurança e aos direitos fundamentais. Nessa categoria estão os sistemas de vigilância biométrica em tempo real em espaços públicos, salvo em hipóteses excepcionais de segurança pública; os sistemas de classificação social de indivíduos (social scoring); e os mecanismos de manipulação comportamental que exploram vulnerabilidades de grupos específicos, como brinquedos com IA capazes de induzir práticas perigosas em crianças. Também são vedadas as técnicas subliminares destinadas a distorcer o comportamento humano de forma prejudicial, bem como os sistemas de identificação biométrica retroativa em imagens e vídeos, exceto quando expressamente autorizados por autoridade judicial em investigações criminais graves (UNIÃO EUROPEIA, 2024).

Ademais, o *AI Act* prevê a criação de um Comitê Europeu de Inteligência Artificial, inspirado no modelo do Comitê Europeu de Proteção de Dados (instituído pelo GDPR), com a finalidade de coordenar a aplicação uniforme das regras entre os Estados-membros. Cada país deverá designar ou instituir autoridades nacionais de

supervisão de IA, dotadas de poderes para fiscalizar, avaliar conformidade e aplicar sanções. Em caso de descumprimento, as multas podem alcançar até 30 milhões de euros ou 6% do faturamento global anual da empresa infratora, prevalecendo o valor mais elevado, em patamar semelhante ao estabelecido pelo GDPR. Em 2025, a Comissão Europeia iniciou consultas públicas para a elaboração de um Código de Conduta voluntário direcionado a fornecedores de IA de propósito geral, como os modelos fundacionais do tipo GPT, antecipando obrigações que se tornarão exigíveis apenas após a plena aplicação do regulamento. Complementarmente, encontra-se em discussão a aprovação de uma Diretiva de Responsabilidade Civil em IA (AI Liability Directive), com o objetivo de facilitar a reparação de danos causados por sistemas defeituosos ou discriminatórios (UNIÃO EUROPEIA, 2024a).

Por fim, explicita-se que o Regulamento Europeu de Inteligência Artificial, *AI ACT*, integra um arcabouço regulatório mais amplo da União Europeia para o setor tecnológico. Em particular, complementa o Regulamento Geral de Proteção de Dados (GDPR), reforçando requisitos de qualidade e transparência no tratamento de dados pessoais. Além disso, articula-se com legislações recentes como a Lei de Serviços Digitais (DSA) e a Lei de Mercados Digitais (DMA), que, embora não se concentrem exclusivamente na IA, impõem obrigações a plataformas digitais, abrangendo aspectos de moderação de conteúdo e sistemas de recomendação baseados em algoritmos. Importa mencionar ainda as Diretrizes de IA Confiável publicadas em 2019, que serviram de fundamento ético e conceitual para a elaboração do regulamento. Com a entrada em vigor do *AI Act*, a União Europeia consolida uma postura regulatória rigorosa: estimular a inovação em inteligência artificial, mas dentro de limites que assegurem valores fundamentais como direitos humanos, democracia, privacidade e segurança. Tal legislação tem sido referência para outros países e blocos econômicos que buscam desenvolver regimes jurídicos semelhantes (UNIÃO EUROPEIA, 2024b).

5.2 ESTADOS UNIDOS

Até a data da presente pesquisa, realizada durante o mês de setembro de 2025, os Estados Unidos não possuem uma lei federal abrangente específica sobre inteligência artificial, adotando uma abordagem setorial e baseada em diretrizes. Em vez de legislação formal, o governo federal tem recorrido a ordens executivas e orientações de agências. Em 30 de outubro de 2023, o presidente Joe Biden assinou uma Ordem Executiva sobre IA, com o objetivo de equilibrar a inovação tecnológica com a segurança nacional e a proteção de direitos dos consumidores (SANTOS, 2023). Essa medida estabeleceu oito diretrizes principais para políticas de IA: criação de padrões de segurança, proteção da privacidade, promoção da equidade e dos direitos civis, defesa de consumidores e estudantes, apoio a trabalhadores, estímulo à inovação e à concorrência, reforço da liderança americana em IA e garantia do uso governamental responsável da tecnologia.

Entre as medidas, determinou-se que desenvolvedores de IA avançada compartilhem resultados de testes de segurança com o governo, que sejam criados padrões técnicos pelo NIST para assegurar a confiabilidade dos sistemas e que conteúdos gerados por IA sejam identificados, a fim de evitar fraudes e desinformação (SANTOS, 2023). A ordem também recomendou ao Congresso a aprovação de uma legislação nacional de proteção de dados, dada a relevância do tema diante do avanço da IA. No âmbito legislativo, diversos projetos de lei federais têm sido apresentados, como o Algorithmic Accountability Act, voltado à transparência algorítmica; contudo, nenhum foi aprovado até 2025, em razão de impasses políticos e da atuação de grupos de interesse (VIANA, 2025). Por fim, merece destaque a publicação, em 2022, do Blueprint for an AI Bill of Rights pela Casa Branca, documento de caráter orientativo que delineia princípios para o uso ético da IA, embora sem força de lei.

Diante da ausência de uma lei federal abrangente, diversos estados e cidades dos Estados Unidos aprovaram suas próprias normas sobre inteligência artificial,

criando um mosaico regulatório. Estima-se que, até 2025, mais de 700 projetos de lei estaduais tratavam de aspectos relacionados à IA, como *deepfake*s, desinformação eleitoral, proteção de crianças e impactos ambientais. Nesse contexto, chegou a tramitar na Câmara dos Deputados uma proposta de moratória de dez anos, com o objetivo de impedir que estados aprovassem legislações específicas sobre IA, de modo a garantir tempo para a formulação de uma política federal unificada. O texto sugeria que nenhum estado poderia aplicar qualquer lei ou regulamento que disciplinasse modelos ou sistemas de inteligência artificial durante esse período. Enquanto esse debate seguia em nível nacional, alguns estados, como a Califórnia e a cidade de Nova York, destacaram-se ao aprovar legislações próprias e pioneiras sobre o tema (VIANA, 2025).

5.3 CALIFÓRNIA

A Califórnia consolidou-se como um dos estados líderes na regulação da inteligência artificial nos Estados Unidos. Em setembro de 2024, foi aprovado um pacote legislativo contendo entre 12 e 17 leis relacionadas à IA, abrangendo desde a gestão de riscos e a transparência sobre os dados de treinamento até questões de privacidade, uso de *deepfake*s, robocalls, aplicação da IA na saúde e na educação, bem como a obrigatoriedade de inserção de marca d'água (watermarking) em conteúdos gerados artificialmente. Entre as medidas, destaca-se a exigência de que desenvolvedores de sistemas de IA generativa divulguem informações detalhadas sobre os conjuntos de dados utilizados no treinamento dos modelos. Outra lei relevante (SB 942) determinou que provedores com mais de um milhão de usuários ofereçam ferramentas de detecção de conteúdo sintético e implementem marca d'água em materiais produzidos. Além disso, passou a ser obrigatória a indicação expressa em propagandas políticas quando houver uso de IA, como forma de reduzir os riscos de desinformação. O governador Gavin Newsom qualificou esse pacote

como a legislação mais abrangente do país sobre o tema, destacando sua importância no combate aos *deepfake*s, na proteção de crianças e trabalhadores e no fortalecimento da confiança pública frente ao avanço da tecnologia. Apesar disso, vetou um projeto que previa regras adicionais para modelos de IA de grande porte, justificando que tais medidas seriam prematuras. A iniciativa californiana reflete a tentativa de preencher a lacuna da ausência de uma lei federal e, ao mesmo tempo, estimular a inovação responsável no Vale do Silício (PwC, 2024).

5.4 NOVA YORK

Em julho de 2023, entrou em vigor na cidade de Nova York a Lei Local nº 144, considerada pioneira nos Estados Unidos na regulação do uso de inteligência artificial em contratações de emprego. Essa legislação municipal exige que empresas que utilizem ferramentas automatizadas de decisão em processos de recrutamento ou promoção realizem auditorias independentes anuais para verificar potenciais vieses algorítmicos, além de tornarem públicos os resultados.

As organizações devem ainda divulgar quais características são avaliadas pelos algoritmos, incluindo eventuais pontuações atribuídas a candidatos com base em raça, etnia ou gênero, de modo a promover transparência e prevenir discriminações. Sem a realização da auditoria e a publicação dos dados, a utilização desses sistemas é proibida, sob pena de multa. A lei se aplica inclusive a vagas remotas destinadas a residentes da cidade de Nova York e surgiu em resposta a casos de viés discriminatório em sistemas de recrutamento por IA, como algoritmos que penalizavam candidaturas de mulheres ou de minorias. Essa iniciativa nova-iorquina inspirou discussões em outras localidades sobre a obrigatoriedade de auditorias algorítmicas em ferramentas de recursos humanos (OLHAR DIGITAL, 2023).

5.5 OUTROS ESTADOS

Além de Nova York e Califórnia, outros estados também avançaram em legislações relacionadas ao uso da inteligência artificial e de dados biométricos. O estado de Illinois, por exemplo, implementou em 2008 a *Biometric Information Privacy Act* (BIPA), considerada pioneira em matéria de privacidade biométrica, exigindo consentimento expresso para a coleta de dados como impressões digitais e reconhecimento facial, além de prever ações judiciais em caso de violação. Já estados como Washington, Colorado e Virgínia aprovaram leis gerais de proteção de dados pessoais inspiradas no GDPR europeu e na CCPA da Califórnia, incluindo dispositivos sobre decisões automatizadas.

Em paralelo, algumas cidades e estados norte-americanos adotaram restrições ao uso de reconhecimento facial por autoridades públicas, como ocorreu em São Francisco e Boston, que proibiram o uso da tecnologia por suas forças policiais. Assim, no vácuo da ausência de uma lei federal, a regulação da IA nos EUA em 2025 combina ações executivas nacionais não vinculantes, projetos legislativos federais pendentes e uma diversidade de normas estaduais e municipais, com tendência de se buscar maior harmonização para evitar a fragmentação regulatória (ILLINOIS, 2008; REUTERS, 2024).

5.6 CHINA

A China tem adotado uma posição proativa e centralizada na regulamentação da inteligência artificial, com foco no controle de conteúdo, na segurança nacional e no alinhamento aos valores estabelecidos pelo Estado. Diferentemente das democracias ocidentais, em que a regulação avança por meio de processos legislativos e debates parlamentares, o modelo chinês temse consolidado

principalmente por regulamentos administrativos e diretrizes emitidas por agências governamentais.

5.7 MEDIDAS ADMINISTRATIVAS PARA IA GENERATIVA (2023)

Em resposta à rápida popularização de modelos generativos — como GPT e Midjourney —, a China foi um dos primeiros países a editar regras específicas para disciplinar o setor. As "*Interim Measures for the Management of Generative AI Services"*, emitidas pela Administração do Ciberespaço da China (CAC) em conjunto com outros órgãos governamentais, entraram em vigor em 15 de agosto de 2023 e estabeleceram diretrizes centrais para o fornecimento de serviços de IA generativa no país.

A priori, as medidas impõem a realização de avaliações de segurança, exigindo que os provedores testem seus sistemas antes de disponibilizá-los ao público, de modo a mitigar riscos associados ao uso da IA generativa. Além disso, determina-se que o conteúdo produzido por essas ferramentas seja verídico e preciso, na medida do possível, com o objetivo de reduzir a disseminação de desinformação e garantir maior confiabilidade nas saídas geradas.

Outro ponto essencial é a responsabilidade editorial atribuída aos provedores, que devem assegurar que os materiais gerados estejam em conformidade com as leis de censura e com os chamados "valores socialistas" da República Popular da China. Fica, portanto, vedada a criação de conteúdos que ameacem a segurança nacional, promovam propaganda subversiva, infrinjam a moral pública ou violem direitos de imagem e de propriedade intelectual.

No campo administrativo, as regras determinam que todas as empresas que desenvolvam ou disponibilizem serviços de IA generativa realizem registro junto às autoridades reguladoras, devendo ainda instituir mecanismos de supervisão humana contínua para monitorar, revisar e filtrar conteúdos considerados problemáticos. O

escopo de aplicação é amplo e abrange tanto empresas nacionais quanto estrangeiras que operem no território chinês, impondo obrigações de transparência, responsabilidade e conformidade regulatória.

Em síntese, esse regulamento reflete uma abordagem que bus caconciliar inovação tecnológica com controle estatal, abordando temas como privacidade, segurança de dados e filtragem de conteúdo, sob a lógica de promover uma inovação “segura e controlada” (CHINA, 2023).

5.8 DISPOSIÇÕES SOBRE *DEEP SYNTHESIS* (2022)

Diante da crescente popularização dos *deepfake*s — imagens, vídeos ou áudios gerados ou alterados por inteligência artificial —, a China promulgou as Disposições sobre a Administração de Serviços de Informação na Internet com Tecnologia de Síntese Profunda (Provisions on the Administration of Deep Synthesis Internet Information Services), que entraram em vigor em 10 de janeiro de 2023. Esse regulamento, pioneiro no cenário global, estabelece que qualquer conteúdo criado ou manipulado por IA deve ser claramente rotulado, a fim de evitar que o público seja induzido em erro. Além disso, impõe às empresas que oferecem serviços de edição de rosto, voz ou criação de mídia sintética a obrigação de obter consentimento expresso dos indivíduos retratados antes de alterar identidades em conteúdos. Os provedores respondem legalmente caso permitam a criação de material falso que cause danos a terceiros ou comprometa a segurança pública. O regulamento também proíbe o uso de *deepfake*s para fins ilícitos — como fraude, difamação, pornografia e propaganda política não autorizada — e exige que sejam implementados algoritmos de detecção e mecanismos de remoção célere de conteúdo malicioso. Assim, a medida buscou coibir falsificações digitais capazes de ameaçar a integridade da informação online e a ordem social, colocando a China na vanguarda do combate jurídico aos *deepfake*s (CHINA, 2022).

5.8.1 Medidas de Revisão Ética em Pesquisa de IA (2023)

Em 1º de dezembro de 2023 entraram em vigor novas disposições voltadas à ética na pesquisa e no desenvolvimento de inteligência artificial. Emitidas pelo Ministério da Ciência e Tecnologia, essas regras determinam que projetos de IA considerados “sensíveis” — isto é, envolvendo seres humanos, animais ou potenciais impactos sociais significativos — sejam submetidos previamente a comitês de revisão ética. As organizações que conduzem pesquisas em IA devem instituir comissões internas de ética e, em determinadas situações, apresentar seus projetos a especialistas externos para avaliação independente. Entre os critérios fixados, destacam-se o respeito à dignidade humana, a proteção da segurança e a prevalência do interesse público. Projetos que não atendam a esses requisitos podem ser suspensos. Embora em sintonia com debates internacionais sobre ética em IA, a medida integra tais preocupações ao aparato estatal chinês, alinhando a inovação tecnológica aos valores e diretrizes governamentais e prevenindo riscos em áreas sensíveis, como neurotecnologia ou manipulação genética (CHINA, 2023).

5.8.2 Regulamento de Algoritmos de Recomendação (2022)

Desde março de 2022, está em vigor na China o Regulamento de Serviços de Recomendação Algorítmica, emitido pela Administração do Ciberespaço da China (CAC). Esse conjunto de regras foi o primeiro do gênero no mundo e tem como objetivo disciplinar algoritmos que personalizam conteúdo online, como feeds de redes sociais, recomendações de vídeos, notícias e plataformas de comércio eletrônico. A norma exige que provedores de serviços de recomendação se registrem junto ao governo e forneçam detalhes de seus algoritmos às autoridades competentes. Também obriga as plataformas a disponibilizar aos usuários a possibilidade de

desativar recomendações personalizadas ou escolher modos alternativos de ordenação de conteúdo (por exemplo, listas não personalizadas).

Inicialmente, o regulamento estabelece exigências de registro e transparência, determinando que as empresas forneçam às autoridades regulatórias informações detalhadas sobre o funcionamento e a lógica de seus algoritmos de recomendação. Além disso, impõe às plataformas a obrigação de disponibilizar aos usuários opções para desativar as recomendações personalizadas ou escolher formas alternativas de visualização, como listas cronológicas ou não personalizadas, conferindo maior autonomia ao usuário sobre sua experiência digital.

Outro aspecto relevante do regulamento diz respeito aos limites de conteúdo, proibindo o uso de algoritmos que fomentem vícios, manipulem comportamentos ou promovam a disseminação de conteúdos ilícitos. Pelo contrário, exige-se que as plataformas priorizem valores considerados “positivos”, como a difusão de informações verificadas e provenientes de fontes oficiais, especialmente em serviços de notícias.

No campo dos direitos individuais, o texto normativo também assegura o direito à reparação, garantindo que os usuários sejam informados quando interagem com conteúdos organizados por sistemas algorítmicos e possam contestar eventuais prejuízos, como casos de discriminação automatizada ou ranqueamento injusto.

Essa regulação busca garantir o uso “responsável e transparente” dos sistemas de recomendação, reconhecendo sua forte influência na opinião pública e no comportamento social na era digital. Em termos práticos, assegura que grandes plataformas como *TikTok/Douyin, WeChat e Alibaba* utilizem seus algoritmos em conformidade com as políticas do Estado e o bem-estar do público (CHINA, 2022).

5.8.3 Outras iniciativas e contexto

A estratégia chinesa de regulação da IA é complementada por leis mais amplas de segurança e privacidade de dados aprovadas em 2021, como a Lei de Segurança de Dados (DSL) e a Lei de Proteção de Informação Pessoal (PIPL), que impõem controles rigorosos sobre exportação de dados, armazenamento local e consentimento no uso de dados pessoais. Além disso, planos como o Plano de Ação de IA de Nova Geração (2017) e as Diretrizes Éticas para a IA (2021) definem metas para a liderança chinesa em IA até 2030, equilibrando incentivo à pesquisa com supervisão estatal.

Nos anos de 2023 e 2024, autoridades chinesas ainda anunciaram a intenção de criar um Conselho Nacional de IA ou de atribuir a coordenação a órgãos como o Ministério da Indústria e Tecnologia da Informação, o que sinaliza a possibilidade de uma futura lei-quadro unificada. Até o momento, entretanto, a governança de IA no país permanece fragmentada, mas avança de maneira rápida e eficiente por meio de regulamentos administrativos e decretos setoriais. Essa postura regulatória, que combina incentivos massivos à inovação com controle estrito sobre o uso das tecnologias, torna a China um caso singular no panorama global (CHINA, 2021; CHINA, 2017).

5.9 BRASIL

Nos últimos anos, o Brasil vem consolidando um arcabouço jurídico para regular a inteligência artificial (IA) e enfrentar os abusos decorrentes dos *deepfake*s, com o objetivo de mitigar riscos sociais e assegurar direitos fundamentais. Em dezembro de 2024, o Senado Federal aprovou o Marco Legal da Inteligência Artificial (PL 2338/2023), que estabelece diretrizes éticas e de governança para sistemas de IA. O texto prevê a criação de um Sistema Nacional de Regulação, coordenado pela Autoridade Nacional de Proteção de Dados (ANPD), e enfatiza princípios como a centralidade da pessoa humana, a proteção da privacidade e a transparência,

garantindo às vítimas meios de rastrear a origem de abusos tecnológicos e buscar reparação. Segundo o relator, o projeto não introduz mecanismos de censura, preservando a liberdade de expressão nas redes sociais, em equilíbrio com a inovação (AGÊNCIA SENADO, 2024a).

Além desse marco geral, diversos projetos de lei específicos têm avançado no Congresso Nacional, voltados a situações concretas em que os *deepfake*s geram riscos mais evidentes, como fraudes, violência de gênero e manipulação política. Entre os principais problemas identificados estão: a desinformação e manipulação da opinião pública, a pornografia não consensual e a exposição indevida, os crimes cibernéticos e fraudes e o uso não autorizado da identidade digital de figuras públicas.

A disseminação de *deepfake*s apresenta riscos sociais e jurídicos de natureza complexa, capazes de afetar tanto a esfera individual quanto a coletiva. Em primeiro lugar, destaca-se a desinformação e manipulação da opinião pública, uma vez que vídeos e áudios sintéticos podem atribuir falas e comportamentos inexistentes a autoridades ou candidatos, distorcendo o debate democrático. No Brasil, episódios de conteúdos falsos atribuídos a figuras públicas já levaram órgãos governamentais a emitir alertas oficiais (BRASIL, 2024). Nesse sentido, propostas legislativas como o PL 3821/2024 e o PL 5931/2023 visam coibir o uso de *deepfake*s eleitorais, prevendo penas severas e multas, além de obrigar a rápida remoção do material ilícito (BORGES, 2025; SOARES, 2023).

Complementarmente, a Resolução TSE nº 23.732/2024 introduziu parâmetros específicos para o uso de tecnologias de inteligência artifi cialdurante o período eleitoral. O texto proíbe a utilização de *deepfake*s que manipulem a imagem ou a voz de candidatos com potencial de enganar o eleitorado, salvo quando houver consentimento expresso e identificação clara de que se trata de conteúdo sintético. Além disso, a resolução obriga a rotulagem de materiais gerados por IA, veda a divulgação de conteúdos artificiais que atentem contra a integridade das eleições e estabelece responsabilidade solidária entre autores, beneficiários e plataformas pela

veiculação de conteúdos manipulados. Ao adotar essas medidas, o TSE reforça o compromisso de equilibrar liberdade de expressão, integridade informacional e proteção do processo democrático, diante do avanço das tecnologias de geração de conteúdo artificial (BRASIL, TSE, 2024).

Outro risco central refere-se aos *deepfake*s sexuais não consensuais, que configuram grave violação da dignidade e da intimidade, sobretudo de mulheres. Em 2023, o caso envolvendo alunas cujas imagens foram manipuladas para criar nudes falsos ilustrou os danos psicológicos e sociais dessa prática (AGUIAR, 2024). Em resposta, o PL 3821/2024 tipificou a conduta como crime, com pena de até seis anos de reclusão, e projetos como o PL 5695/2023 avançam para incluir essa prática como violência digital na Lei Maria da Penha (AGUIAR, 2024; BORGES, 2025).

No campo dos crimes cibernéticos, *deepfake*s têm sido utilizados em fraudes financeiras e extorsões, explorando a credibilidade de figuras públicas ou a voz de familiares para enganar as vítimas. Em 2024, a Polícia Civil desmantelou uma quadrilha que utilizava a imagem sintética de um apresentador para induzir consumidores a acessarem sites fraudulentos e fornecer dados bancários (DE TILIA, 2025). Para enfrentar essa realidade, o PL 146/2024 propõe qualificar o crime de identidade falsa praticado com *deepfake*s, estabelecendo punições mais rigorosas (AGÊNCIA SENADO, 2024b).

Por fim, merece destaque o uso indevido da identidade digital de figuras públicas em contextos comerciais ou publicitários. Casos de exploração da imagem ou voz de celebridades em anúncios de produtos sem consentimento reforçaram a necessidade de regulamentação. O PL 145/2024 exige autorização prévia do titular da imagem ou de seus herdeiros para o uso em publicidade, sob pena de caracterização como propaganda enganosa (AGÊNCIA SENADO, 2024b). Esse projeto soma-se a outras iniciativas, como o PL 3608/2023, que regula as chamadas “ *deepfake*s póstumas” (BRASIL, 2023). Assim, os riscos abarcam desde a

manipulação da democracia até a violação da intimidade, passando por fraudes e apropriação indevida da imagem.

A resposta jurídica a esses riscos tem se estruturado a partir da proteção de direitos fundamentais, de modo a equilibrar repressão a condutas abusivas e preservação das liberdades individuais. O direito à privacidade e à proteção de dados é diretamente afetado, uma vez que imagem e voz são classificados como dados pessoais sensíveis pela Lei Geral de Proteção de Dados (Lei nº 13.709/2018). A manipulação sem consentimento afronta à intimidade e demanda medidas adicionais, já que a LGPD não cobre adequadamente o comportamento de particulares que criam conteúdos ilícitos. O Marco Legal da IA (PL 2338/2023) reforça essa proteção, ao estabelecer a privacidade como princípio central e assegurar às vítimas o direito de rastrear a origem de abusos (AGÊNCIA SENADO, 2024a).

O direito à imagem e à personalidade, garantido pela Constituição (art. 5º, V e X) e pelo Código Civil, também se vê ameaçado. Projetos como o PL 145/2024 e o PL 146/2024 ampliam a tutela, exigindo consentimento expresso para uso publicitário e agravando penas em casos de difamação ou falsa identidade digital (AGÊNCIA SENADO, 2024b). O PL 3608/2023, por sua vez, assegura a dignidade da pessoa falecida, ao exigir consentimento familiar para recriações digitais (BRASIL, 2023).

A liberdade de expressão, por outro lado, demanda especial atenção. Embora a regulação de *deepfake*s seja necessária para conter desinformação, fraude e abusos, ela não pode inviabilizar usos legítimos, como sátiras, paródias e produções artísticas. O Marco Legal da IA ressalta que não cria dispositivos de censura, e projetos tipificam condutas ilícitas exigindo dolo específico de causar dano, preservando, assim, manifestações críticas ou humorísticas (AGÊNCIA SENADO, 2024a).

A responsabilização de provedores, por sua vez, constitui um dos pontos mais sensíveis do debate contemporâneo. O Marco Civil da Internet (Lei nº 12.965/2014) já prevê hipóteses de responsabilização, mas novos projetos buscam expandir esses

deveres. O PL 2630/2020, conhecido como Lei das Fake News, propõe maior responsabilidade das plataformas na contenção de desinformação e *deepfake*s, o que tem suscitado debates sobre o risco de excesso de moderação e possíveis restrições indevidas à liberdade comunicacional (AGÊNCIA SENADO, 2020). Já o PL 145/2024 e o PL 5931/2023 estabelecem deveres específicos de remoção e rotulagem em publicidade e propaganda eleitoral, apontando para um modelo de responsabilidade compartilhada entre produtores, beneficiários e veículos de divulgação (SOARES, 2023).

Nesse contexto, destaca-se a recente decisão do Supremo Tribunal Federal, que, ao julgar os Recursos Extraordinários nº 1.037.396 (Tema 987) e nº 1.057.258 (Tema 533), reconheceu a inconstitucionalidade parcial e progressiva do artigo 19 do Marco Civil da Internet. O Tribunal entendeu que o modelo de responsabilização condicionado exclusivamente ao descumprimento de ordem judicial é insuficiente para proteger direitos fundamentais e a integridade democrática no ambiente digital. Assim, até que sobrevenha nova legislação, o dispositivo deve ser interpretado de modo a permitir a responsabilização civil de provedores também quando, notificados extrajudicialmente, deixarem de remover conteúdos ilícitos, mantendo, entretanto, a exigência de decisão judicial para os crimes contra a honra. A Corte também fixou deveres de cuidado proativo em casos de crimes graves, como terrorismo, pornografia infantil, discurso de ódio e atos antidemocráticos, cuja omissão caracteriza falha sistêmica do provedor. Essa decisão estabelece um modelo escalonado de responsabilidade, que busca equilibrar a liberdade de expressão, a inovação tecnológica e a proteção dos direitos fundamentais no espaço digital (BRASIL, STF, 2025).

Assim, observa-se que os riscos concretos dos *deepfake*s encontram correspondência em respostas jurídicas voltadas à proteção da privacidade, da imagem, da liberdade de expressão e à responsabilização de plataformas. O desafio

central reside em calibrar essas respostas, de modo a prevenir abusos sem inviabilizar a inovação e a liberdade comunicacional.

6 QUADRO COMPARATIVO

A análise da regulamentação de *deepfake*s em diferentes ordenamentos jurídicos apresenta elevada complexidade, dada a variedade de enfoques normativos e institucionais adotados. Nesse contexto, os quadros comparativos surgem como ferramenta metodológica que possibilita sintetizar, de forma sistemática e acessível, os principais pontos de convergência e divergência entre Brasil, Estados Unidos, União Europeia e China.

A opção pela construção de quadros justifica-se por duas razões principais. Em primeiro lugar, permite visualizar de maneira clara como cada jurisdição enfrenta riscos específicos decorrentes dos *deepfake*s, seja no âmbito da desinformação política, da proteção contra conteúdos sexuais não consensuais, da prevenção de fraudes ou da salvaguarda da identidade digital. Em segundo lugar, favorece a identificação de tendências globais e potenciais caminhos de harmonização normativa, objetivo central deste trabalho.

Dessa forma, os quadros comparativos não apenas sintetizam os achados da pesquisa, mas também constituem instrumento crítico de análise, revelando tanto os avanços já consolidados quanto às lacunas ainda existentes. O presente capítulo organiza-se em duas partes: (i) o primeiro quadro aborda os riscos jurídicos e sociais relacionados aos *deepfake*s e as estratégias regulatórias para mitigá-los; (ii) o segundo quadro concentra-se na proteção de direitos fundamentais e exceções, evidenciando como cada sistema jurídico equilibra garantias individuais e liberdades com o combate aos usos abusivos da tecnologia.

6.1 QUADRO I: RISCOS JURÍDICOS E SOCIAIS

A análise da regulamentação das *deepfake*s em diferentes ordenamentos jurídicos apresenta elevada complexidade, dada a diversidade de enfoques normativos e institucionais adotados. Nesse contexto, o Quadro I é apresentado como uma ferramenta metodológica que sintetiza, de forma sistemática e acessível, os principais pontos de convergência e divergência entre Brasil, Estados Unidos, União Europeia e China.

A opção pela construção desse quadro justifica-se por duas razões principais. Em primeiro lugar, permite visualizar de modo claro e comparativo como cada jurisdição enfrenta riscos específicos decorrentes das *deepfake*s — seja no campo da desinformação política, da proteção contra conteúdos sexuais não consensuais, da prevenção de fraudes ou da salvaguarda da identidade digital. Em segundo lugar, favorece a identificação de tendências globais e possíveis caminhos de harmonização normativa, alinhando-se ao objetivo central deste trabalho.

Dessa forma, o Quadro I – Riscos Jurídicos e Sociais sintetiza os achados da pesquisa, e também constitui um instrumento crítico de análise, capaz de evidenciar tanto os avanços consolidados quanto as lacunas ainda existentes. Ao integrar aspectos jurídicos e sociais, o quadro permite compreender de forma abrangente como diferentes sistemas jurídicos estruturam suas estratégias de mitigação dos riscos associados às *deepfake*s, contribuindo para a reflexão sobre modelos de governança digital mais equilibrados e eficazes.

Nos Estados Unidos, a ausência de uma lei federal específica evidencia um modelo fragmentado e desigual, marcado pela proliferação de legislações estaduais que tratam pontualmente de *deepfake*s eleitorais ou sexuais não consensuais. A Ordem Executiva de 2023 buscou preencher parte dessa lacuna ao exigir transparência e *watermarking* em conteúdos sintéticos, mas sua natureza administrativa revela instabilidade, visto que pode ser modificada por gestões

subsequentes. A atuação de agências como a *Federal Trade Commission* (FTC) e do *Federal Bureau of Investigation* (FBI) demonstra capacidade de resposta prática, sobretudo em casos de fraude, mas a dependência de normas gerais de estelionato ou falsa identidade reforça o caráter reativo da tutela. Assim, embora haja resposta institucional, a ausência de uniformidade nacional dificulta a consolidação de um sistema jurídico coeso. Por sua vez, a União Europeia apresenta o modelo mais abrangente e sistemático. O Artificial Intelligence Act (*AI Act*) classifica os conteúdos sintéticos como de "risco limitado", impondo rotulagem obrigatória, e admite sua categorização como de "alto risco" em contextos sensíveis, como segurança biométrica ou sistemas críticos. Complementarmente, o Digital Services Act (DSA) impõe deveres de moderação e transparência às plataformas digitais, enquanto o Regulamento Geral de Proteção de Dados (GDPR) assegura tutela robusta da privacidade, da imagem e da autodeterminação informativa. Embora coerente em teoria, esse modelo depende de forte capacidade administrativa e de cooperação supranacional para que sua aplicação seja efetiva, o que pode representar desafio diante da heterogeneidade institucional entre os Estados-Membros (MARTINS; LONGHI, 2024).

O modelo chinês, por outro lado, distingue-se pelo caráter centralizado, preventivo e rigoroso. O Regulamento de Deep Synthesis (2022/2023) exige rotulagem obrigatória de conteúdos manipulados e impõe consentimento expresso para uso de imagem ou voz, além de proibir explicitamente a manipulação política e pornografia não consensual. As plataformas digitais são responsabilizadas por prevenir, monitorar e remover conteúdos ilícitos, sob pena de sanções severas, em integração com a Lei de Proteção de Dados Pessoais (PIPL) e com a Lei de Segurança Cibernética. Embora se mostre altamente eficaz em curto prazo, a crítica que se impõe é a possibilidade de instrumentalização política e a ausência de garantias plenas de liberdade de expressão, dado o modelo de governança digital pautado pelo controle estatal.

Em síntese, o comparativo revela que: (i) o Brasil avança de forma reativa, com legislações setoriais e incipientes; (ii) os Estados Unidos apresentam resposta fragmentada, com forte atuação prática, mas sem uniformidade normativa; (iii) a União Europeia constrói um modelo abrangente e prospectivo, embora dependente de elevada capacidade institucional para implementação; e (iv) a China adota a regulação mais rígida e imediata, garantindo eficácia, mas com sérios riscos de restrição às liberdades individuais. Dessa forma, pode-se concluir que nenhum dos modelos é isento de limitações: enquanto a União Europeia equilibra melhor inovação e direitos fundamentais, a China privilegia eficácia e controle, os Estados Unidos confiam na descentralização e na autorregulação, e o Brasil ainda busca consolidar uma política nacional integrada.

Risco	**Brasil**	**EUA**	**União Europeia**	**China**
Desinformação e Manipulação da Opinião Pública	PL 2338/2023 (Marco Legal da IA) adota abordagem baseada em risco; Res. TSE nº 23.732/2024 proíbe *deepfakes* em campanhas eleitorais.	Ordem Executiva de 2023 exige transparência e watermarking; leis estaduais (ex.: Califórnia, Texas) obrigam rotulagem de conteúdo sintético em eleições.	*AI Act* exige rotulagem obrigatória (risco limitado), podendo classificar como alto risco em contextos sensíveis; DSA impõe deveres de moderação de plataformas.	Regulamento de Deep Synthesis (2022/2023) exige rotulagem obrigatória e proíbe manipulação política; forte controle estatal.

_deepfake_s Sexuais e Exposição Indevida	PL 3821/2024 criminaliza pornografia *deepfake* não consensual; LGPD protege dados sensíveis; aplicação complementar da Lei Maria da Penha.	Diversos estados já criminalizam pornografia *deepfake*; aplicação de leis de privacidade e assédio.	GDPR protege dados pessoais (imagem/voz); *AI Act* reforça proibição de usos abusivos; legislações nacionais de gênero complement am.	Regulamento s proíbem expressamen te pornografia *deepfake*; provedores devem remover conteúdo sob pena de sanções severas.
Fraudes, Extorsões e Crimes Cibernéticos	Tratados como estelionato, fraude eletrônica e falsa identidade (CP); PL 146/2024 propõe agravantes para *deepfakes* em crimes cibernéticos.	Aplicação de leis federais e estaduais de fraude; FTC e FBI atuam em casos de golpes com *deepfake*.	*AI Act* exige robustez e mitigação de riscos em sistemas de IA de alto risco; Europol orienta sobre detecção de fraudes.	Leis de ciberseguran ça e PIPL impõem auditorias obrigatórias; provedores responsabiliz ados por prevenir golpes com IA.
Uso Indevido de Identidade Digital de Figuras Públicas	Proteção via Código Civil (art. 20), LGPD e Súmula 403 do STJ; PL 145/2024 exige	Aplicação de right of publicity em leis estaduais; indenização	GDPR assegura direitos de personalidad e; *AI Act* impõe	Regulamento de Deep Synthesis exige consentiment o para uso

	consentimento expresso para uso publicitário.	por uso indevido de imagem/voz.	transparência em conteúdos sintéticos.	de imagem/voz; plataformas obrigadas à remoção rápida.

QUADRO 1 – Riscos Jurídicos e Sociais – 2025 Fonte: As autoras (2025)

7 ANÁLISE DOS RESULTADOS

A análise comparada das respostas normativas ao fenômeno das *deepfake*s revela tanto convergências quanto divergências significativas entre Brasil, Estados Unidos, União Europeia e China. Em matéria de desinformação e manipulação da opinião pública, constata-se que todos os ordenamentos buscam enfrentar o impacto da tecnologia no processo democrático, mas a partir de estratégias distintas. O Brasil, por meio da Resolução TSE nº 23.732/2024, inovou ao vedar expressamente o uso de *deepfake*s em campanhas eleitorais, impondo obrigações de remoção imediata e rotulagem de conteúdos sintéticos, ainda que de forma restrita ao período eleitoral. Nos Estados Unidos, a resposta é fragmentada: leis estaduais, como as da Califórnia e do Texas, exigem rotulagem em contextos eleitorais, enquanto a ordem executiva federal de 2023 estabelece diretrizes gerais de transparência e watermarking. Já a União Europeia adota um modelo sistêmico, no qual o *AI Act* impõe a rotulagem obrigatória de conteúdos sintéticos, classificando-os como de "risco limitado", mas permitindo a reclassificação como de "alto risco" em contextos sensíveis, articulando-se com o Digital Services Act (DSA) para atribuir responsabilidades adicionais às plataformas. A China, por sua vez, apresenta o regime mais centralizado, com os Regulamentos de Deep Synthesis (2022/2023), que exigem rotulagem compulsória,

proíbem manipulação política e atribuem aos provedores deveres estritos de monitoramento e remoção, em sintonia com sua lógica de controle estatal.

No tocante aos *deepfake*s sexuais e à exposição indevida, há consenso regulatório quanto à gravidade da violação e à necessidade de criminalização. O Brasil avançou com o PL 3821/2024, que tipifica a pornografia sintética não consensual, em diálogo com a LGPD e com a Lei Maria da Penha, consolidando a perspectiva de gênero na resposta jurídica. Nos Estados Unidos, diversos estados já criminalizaram a prática, articulando-se com remédios civis tradicionais, como os publicity rights e as ações de responsabilidade por invasão de privacidade. Na União Europeia, a proteção decorre principalmente do GDPR, que reconhece imagem e voz como dados pessoais, reforçada por legislações nacionais específicas sobre violência de gênero e pelas diretrizes do *AI Act* que vedam usos abusivos da IA. A China adota posição ainda mais rígida, proibindo expressamente o uso da tecnologia para pornografia sem consentimento e impondo sanções severas a provedores que não removerem conteúdos em tempo hábil. Observa-se, portanto, uma convergência quanto à ilicitude da prática, mas com diferenças relevantes entre regimes de proteção centrados em dados e direitos da personalidade (UE), penalização direta (Brasil e China) e instrumentos híbridos civis e penais (EUA).

No campo das fraudes, extorsões e crimes cibernéticos, a comparação evidencia diferentes pontos de ataque normativo. O Brasil, em grande medida, ainda recorre ao Código Penal, enquadrando condutas como estelionato, fraude eletrônica ou falsa identidade, embora projetos como o PL 146/2024 proponham agravantes específicos para o uso de *deepfake*s em crimes cibernéticos. Os Estados Unidos também não dispõem de tipificação federal específica, mas contam com uma infraestrutura institucional robusta, envolvendo a *Federal Trade Commission* (FTC) e o *Federal Bureau of Investigation* (FBI), para aplicar as normas de fraude existentes a novas modalidades de golpes digitais. Na União Europeia, a abordagem é preventiva, com o *AI Act* impondo requisitos de robustez técnica e mitigação de riscos para

sistemas de alto risco, além de iniciativas de cooperação como as orientações da Europol. Já a China integra a questão aos marcos de cibersegurança e à Lei de Proteção de Informação Pessoal (PIPL), impondo auditorias obrigatórias, deveres preventivos e responsabilidade direta a provedores. Nota-se, assim, que a UE e a China enfatizam mecanismos ex ante, vinculados à robustez e à segurança dos sistemas, ao passo que Brasil e EUA priorizam respostas ex post, por meio de enquadramento penal e investigação repressiva.

Quanto ao uso indevido da identidade digital de figuras públicas, as quatro jurisdições convergem em reconhecer a necessidade de consentimento expresso e transparência, mas divergem na intensidade da tutela. No Brasil, a proteção decorre do Código Civil (art. 20) e da LGPD, complementada pela Súmula 403 do STJ, que presume o dano moral em caso de divulgação não autorizada de imagem, e por projetos legislativos que reforçam a exigência de autorização prévia para usos publicitários. Nos Estados Unidos, prevalece a aplicação dos publicity rights, previstos em legislações estaduais e que asseguram indenização pelo uso indevido de imagem ou voz, embora haja tensões com a Primeira Emenda quando se trata de sátiras ou paródias. A União Europeia combina a proteção de dados pessoais pelo GDPR com a obrigação de transparência imposta pelo *AI Act*, enquanto a China estabelece exigência de consentimento prévio e mecanismos céleres de remoção, responsabilizando solidariamente as plataformas. Nota-se, portanto, que a diferença principal reside no grau de responsabilização de intermediários: enquanto China e UE impõem deveres rigorosos de controle, Brasil e EUA mantêm modelos mais dependentes de acionamento judicial ou de regimes estaduais.

De forma transversal, três eixos se destacam. O primeiro é a transparência técnica, com a rotulagem de conteúdos sintéticos como consenso normativo, mas ainda sem padrões interoperáveis e resistentes a adulterações. O segundo é a responsabilização dos provedores, que oscila entre regimes de responsabilidade limitada (EUA e Brasil) e de co-responsabilidade ativa (UE e China). O terceiro é a

capacidade institucional, fundamental para a efetividade das normas: a União Europeia avança em estrutura regulatória integrada, enquanto Brasil e EUA sofrem com dispersão normativa e a China aposta na centralização como meio de assegurar cumprimento imediato. Em todos os casos, a efetiva proteção de direitos fundamentais — privacidade, imagem, personalidade e liberdade de expressão — depende não apenas da legislação repressiva, mas de instrumentos preventivos, mecanismos céleres de tutela e estratégias de educação digital, capazes de enfrentar a volatilidade tecnológica e os riscos sociais associados às *deepfake*s.

1 CONCLUSÃO

O estudo desenvolvido demonstrou que a tecnologia das *deepfake*s representa, de maneira exemplar, a complexa relação entre inovação tecnológica e proteção dos direitos fundamentais. Embora essa ferramenta seja capaz de produzir ganhos expressivos em áreas como educação, arte, cultura e preservação da memória, seus potenciais abusos — desinformação, pornografia não consensual, fraudes e apropriação indevida da identidade digital — projetam riscos que desafiam os limites tradicionais do Direito. A análise comparativa realizada ao longo deste trabalho evidencia que tais riscos não podem ser enfrentados por soluções isoladas ou meramente repressivas: exigem, ao contrário, um esforço normativo, institucional e social capaz de articular prevenção, responsabilização e governança cooperativa.

No contexto brasileiro, observou-se que a regulação ainda é marcada por fragmentação e por respostas setoriais, notadamente no campo eleitoral e penal. Apesar de avanços relevantes, como a criminalização da pornografia sintética não consensual e a tramitação do Marco Legal da Inteligência Artificial, ainda persiste a ausência de um marco integrado que contemple rotulagem obrigatória, mecanismos de responsabilização preventiva de provedores e fortalecimento das capacidades institucionais. Em contraste, a União Europeia constrói um modelo abrangente, com o

AI Act e o *Digital Services Act*, orientado pela lógica do risco e pela tutela reforçada da transparência. Os Estados Unidos, por sua vez, adotam uma resposta fragmentada, em que ordens executivas federais e legislações estaduais convivem com forte atuação de agências de *enforcement*, revelando um sistema responsivo, mas heterogêneo e pouco uniforme. Já a China implementa o modelo mais centralizado e rigoroso, impondo deveres preventivos e sanções severas a provedores, o que assegura eficácia imediata, mas suscita questionamentos sobre garantias de liberdade de expressão.

Nesse quadro, uma agenda frutífera para futuros estudos consiste em aprofundar a análise da regulação privada, especialmente no papel desempenhado por plataformas digitais na formulação de regras próprias de moderação, transparência e *accountability*. Essa esfera normativa paralela pode revelar tanto virtudes — como a rapidez de resposta e a adaptabilidade técnica — quanto riscos, como a concentração de poder em grandes empresas e a ausência de controle democrático. Ademais, é relevante investigar em que medida as respostas normativas hegemônicas — sobretudo as da União Europeia e dos Estados Unidos — reproduzem lógicas coloniais e eurocêntricas, impondo padrões regulatórios globais que podem não dialogar com as realidades locais de países periféricos. A incorporação de uma perspectiva decolonial permitiria repensar a governança das *deepfake*s de forma mais plural, valorizando arranjos normativos regionais e comunitários, capazes de refletir diferentes contextos culturais, políticos e sociais.

A partir desse panorama, constata-se que nenhum modelo é isento de limitações: a União Europeia se sobressai pela coerência sistêmica, ainda que dependa de elevada capacidade administrativa para sua efetividade; os Estados Unidos garantem agilidade prática, mas padecem de fragmentação normativa; a China apresenta eficácia regulatória imediata, mas ao custo de uma limitação expressiva da pluralidade democrática; e o Brasil ainda se encontra em processo de consolidação de uma política nacional integrada, que consiga equilibrar proteção de direitos e

incentivo à inovação. Nesse contexto, ganha relevo a reflexão de João Victor Archegas (2021, p. 237), ao sustentar que “ao constitucionalismo digital, assim, cabe a desafiadora tarefa de empoderar pessoas, defender direitos humanos e fundamentais e estabelecer limites constitucionais na nova fronteira do poder”.

Essa perspectiva reforça que a efetividade da regulação digital não depende apenas de marcos legais formais, mas da construção de instituições responsáveis, capazes de atuar com transparência, pluralismo e prestação de contas. Como o autor observa, a governança digital contemporânea requer uma lógica de corresponsabilidade, em que Estado, plataformas e sociedade civil assumem papéis complementares na defesa dos direitos fundamentais e na promoção de um ambiente informacional saudável. Essa concepção se mostra especialmente relevante diante das *deepfake*s, cujo enfrentamento ultrapassa os limites tradicionais do direito positivo e exige articulação técnica, ética e institucional.

Em síntese, os resultados desta pesquisa indicam que a regulação das *deepfake*s deve ser concebida de forma integrada e multidimensional: não apenas criminalizar condutas abusivas, mas também fomentar educação digital crítica, estimular inovação ética em inteligência artificial, reforçar mecanismos de cooperação internacional e consolidar arranjos de corregulação entre Estado, sociedade civil e plataformas digitais. Essa visão dialoga diretamente com o que Archegas (2025, p. 180) denomina “nova arquitetura de governança constitucional do digital”, que busca equilibrar liberdade, responsabilidade e transparência nas interações mediadas por tecnologia.

Apenas uma abordagem multifacetada poderá assegurar uma governança democrática, justa e eficaz, capaz de enfrentar os riscos trazidos pelas *deepfake*s sem sufocar a liberdade de expressão e a criatividade. Longe de representar apenas uma ameaça, essa tecnologia também oferece a oportunidade de reafirmar a centralidade da dignidade humana na era da inteligência artificial, transformando o

desafio em ocasião para fortalecer os fundamentos constitucionais da vida digital e consolidar um novo pacto de responsabilidade compartilhada na esfera informacional.

REFERÊNCIAS

ABBA. Abba Voyage. 2021. Disponível em: https://abbavoyage.com/. Acesso em: 20 set. 2025.

AGÊNCIA SENADO. Projeto das fake news: entenda pontos da proposta em análise no Congresso. Senado Notícias, 2020. Disponível em: https://www12.senado.leg.br/noticias/infomaterias/2020/12/projeto-das-fake-news-entenda-pont os-da-proposta-em-analise-no-congresso. Acesso em: 20 set. 2025.

AGÊNCIA SENADO. Projetos buscam restringir manipulação de imagens com inteligência artificial. Senado Notícias, 7 fev. 2024b. Disponível em: https://www12.senado.leg.br/noticias/materias/2024/02/07/projetos-buscam-restringir-manipulac ao-de-imagens-com-inteligencia-artificial. Acesso em: 20 set. 2025.

AGÊNCIA SENADO. Senado aprova regulamentação da inteligência artificial; texto vai à Câmara. Senado Notícias, 10 dez. 2024a. Disponível em: https://www12.senado.leg.br/noticias/materias/2024/12/10/senado-aprova-regulamentacao-da-i nteligencia-artificial-texto-vai-a-camara. Acesso em: 20 set. 2025.

AGUIAR, Victor. Uso de IA e *deepfake*s para constranger mulheres poderá dar até 4 anos de cadeia, prevê projeto. *CNN Brasil*, 24 maio 2024. Disponível em: https://www.cnnbrasil.com.br/politica/uso-de-ia-e- *deepfake*s-para-constranger-mulheres-podera-dar-ate-4-anos-de-cadeia-preve-projeto/. Acesso em: 20 set. 2025.

AL JAZEERA. How AI is resurrecting dead Indian politicians as election looms. 12 fev. 2024. Disponível em:

https://www.aljazeera.com/economy/2024/2/12/how-ai-is-used-to-resurrect-dead-indian-politicia ns-as-elections-loom. Acesso em: 19 maio 2025.

MERINI, I. *et al*. *deepfake* Media Forensics: Status and Future Challenges. *Journal of Imaging*, v. 11, n. 3, p. 73, 2025. DOI: https://doi.org/10.3390/jimaging11030073.

ARCHEGAS, João Victor. Constitucionalismo Digital: limites constitucionais na nova fronteira do poder. Belo Horizonte: Fórum, 2025.

ARTIFICIAL INTELLIGENCE ACT. Artificial Intelligence Act – The EU Artificial Intelligence Regulation. 2024. Disponível em: https://artificialintelligenceact.eu/. Acesso em: 19 set. 2025.

BIONI, Bruno Ricardo. Proteção de Dados Pessoais: a função e os limites do consentimento. 3. ed. Rio de Janeiro: Forense, 2021.

BORGES, Rebeca. Câmara aprova projeto que pune divulgação de " *deepfake*" com conteúdo sexual. *CNN Brasil*, 19 fev. 2025. Disponível em: https://www.cnnbrasil.com.br/politica/camara-aprova-projeto-que-pune-divulgacao-de- *deepfake*-com-conteudo-sexual/. Acesso em: 20 set. 2025.

BRASIL. Agência Câmara de Notícias. Projeto exige consentimento prévio para uso de *deepfake* de pessoa falecida. Rep.: Emanuelle Brasil, 7 ago. 2023. Disponível em: https://www.camara.leg.br/noticias/983623-projeto-exige-consentimento-previo-para-uso-de-dee pfake-de-pessoa-falecida/. Acesso em: 20 set. 2025.

BRASIL. Código Civil. Lei nº 10.406, de 10 de janeiro de 2002.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD).

BRASIL. Senado Federal. Projeto de Lei nº 2338/2023. Dispõe sobre o uso de sistemas de Inteligência Artificial. Brasília, 2023. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/162439. Acesso em: 20 set. 2025.

BRASIL. Supremo Tribunal Federal. Informação à sociedade: decisão sobre a inconstitucionalidade parcial e progressiva do artigo 19 do Marco Civil da Internet (Lei nº 12.965/2014). Recursos Extraordinários nº 1.037.396 (Tema 987) e nº 1.057.258 (Tema 533). Relatores: Ministros Dias Toffoli e Luiz Fux. Julgado em 26 jun. 2025. Brasília, DF: STF, 2025. Disponível em: https://portal.stf.jus.br/processos/detalhe.asp?incidente=5160549. Acesso em: 24 out. 2025.

BRASIL. Tribunal Superior Eleitoral. Resolução nº 23.732, de 27 de fevereiro de 2024. Altera a Resolução-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-d e-2024. Acesso em: 24 out. 2025.

CAPORUSSO, N. *deepfake*s for the good: a beneficial application of contentious artificial intelligence technology. In: *International Conference on Applied Human Factors and Ergonomics*. Orlando: Springer, 2020. p. 235–241.

CASTELLS, Manuel. A sociedade em rede. 9. ed. São Paulo: Paz e Terra, 2015.

CHESNEY, Robert; CITRON, Danielle Keats. Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, v. 107, n. 6, p. 1753–1820, 2019. DOI: https://doi.org/10.2139/ssrn.3213954.

CHINA. Cyberspace Administration of China (CAC). *Provisions on the Administration of Algorithmic Recommendation for Internet Information Services*. Pequim, 2022. Disponível em: http://www.cac.gov.cn/2022-01/04/c_1642894606364257.htm. Acesso em: 19 set. 2025.

CHINA. Cyberspace Administration of China (CAC). *Provisions on the Administration of Deep Synthesis Internet Information Services*. Pequim, 2022. Disponível em: http://www.cac.gov.cn/2022-12/11/c_1672221536394086.htm. Acesso em: 19 set. 2025.

CHINA. *Data Security Law of the People's Republic of China (DSL).* Congresso Nacional do Povo, Pequim, 2021. Disponível em: http://www.npc.gov.cn/. Acesso em: 19 set. 2025.

CHINA. *Ethical Norms for New Generation Artificial Intelligence.* Ministério da Ciência e Tecnologia (MOST), Pequim, 2021. Disponível em: http://www.most.gov.cn/. Acesso em: 19 set. 2025.

CHINA. *Interim Measures for the Management of Generative Artificial Intelligence Services.* CAC; Conselho de Estado; Ministério da Indústria e TI, 2023. Disponível em: http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm. Acesso em: 19 set. 2025.

CHINA. *Measures for Ethical Review of Science and Technology Related to Humans.* MOST, Pequim, 2023. Disponível em: https://www.most.gov.cn/. Acesso em: 19 set. 2025.

CHINA. *New Generation Artificial Intelligence Development Plan.* Conselho de Estado da RPC, Pequim, 2017. Disponível em: http://www.gov.cn/. Acesso em: 19 set. 2025.

CHINA. *Personal Information Protection Law of the People's Republic of China (PIPL).* Congresso Nacional do Povo, Pequim, 2021. Disponível em: http://www.npc.gov.cn/. Acesso em: 19 set. 2025.

CLEGG, Nick. Labeling AI-generated images on Facebook, Instagram and Threads. *About.fb.com*, 6 fev. 2024. Disponível em: https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-th reads/. Acesso em: 20 set. 2025.

CNN BRASIL. Câmara aprova projeto que pune divulgação de *deepfake* com conteúdo sexual.

10 set. 2024. Disponível em: https://www.cnnbrasil.com.br/politica/camara-aprova-projeto-que-pune-divulgacao-de- *deepfake*-com-conteudo-sexual/. Acesso em: 20 set. 2025.

EUROPOL. Facing reality? Law enforcement and the challenge of *deepfake*s. Europol Innovation Lab, 2022. Disponível em: https://www.europol.europa.eu/cms/sites/default/files/documents/Europol_Innovation_Lab_Faci ng_Reality_Law_Enforcement_And_The_Challenge_Of_ *deepfake*s.pdf. Acesso em: 7 maio 2025.

ESTADOS UNIDOS. California Consumer Privacy Act (CCPA). Califórnia, 2018.

ESTADOS UNIDOS. Children's Online Privacy Protection Act (COPPA). 1998.

ESTADOS UNIDOS. Executive Order 14110 – Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 30 out. 2023.

ESTADOS UNIDOS. Executive Order 14179 – Removing Barriers to American Leadership in Artificial Intelligence. 23 jan. 2025.

STADOS UNIDOS. Health Insurance Portability and Accountability Act (HIPAA). 1996.

ESTADOS UNIDOS. *Tennessee.* Ensuring Likeness Voice and Image Security (ELVIS) Act. Tenn. Code Ann. §47-25-1101 et seq., 2024.

ESTADOS UNIDOS. *Take It Down Act.* Congress Bill no. 118-78, 2025.

ORBES. Celebridades brasileiras são vítimas da IA: aprenda a reconhecer conteúdos falsos. São Paulo, jan. 2025. Disponível em: https://forbes.com.br/forbes-tech/2025/01/celebridades-brasileiras-sao-vitimas-de-ia-aprenda-a-reconhecer-conteudos-falsos/. Acesso em: 6 maio 2025.

FORTIS, Savannah. TikTok to automatically label AI-generated content. *CoinTelegraph*, 9 mai. 2024. Disponível em: https://cointelegraph.com/news/tiktok-auto-labeling-ai-generated-content. Acesso em: 20 set. 2025.

G1 PERNAMBUCO. Alunas de colégio particular do Recife denunciam divulgação de "nudes" falsos criados com inteligência artificial; polícia investiga. *G1*, nov. 2023. Disponível em: https://g1.globo.com/pe/pernambuco/noticia/2023/11/07/alunas-de-colegio-particular-do-recife-d enunciam-divulgacao-de-nudes-falsos-criados-com-inteligencia-artificial-policia-investiga.ghtml. Acesso em: 3 maio 2025.

G1 RIO GRANDE DO SUL. Adolescentes são suspeitos de criar falsos vídeos com inteligência artificial de alunas nuas no RS; polícia investiga. *G1*, mar. 2024. Disponível em: https://g1.globo.com/rs/rio-grande-do-sul/noticia/2024/03/19/policia-civil-investiga-videos-de-nud ez-de-adolescentes-produzidos-por-inteligencia-artificial-em-porto-alegre.ghtml. Acesso em: 3 maio 2025.

G7. Hiroshima Leaders' Communiqué. Hiroshima: G7 Summit, 2023. Disponível em: https://www.g7hiroshima.go.jp/en/documents/. Acesso em: 20 set. 2025.

GEORGE WASHINGTON UNIVERSITY. US states struggle to define *deepfake*s and related terms as technically complex legislation proliferates. GW Law, 2023. Disponível em: https://www.law.gwu.edu/us-states-struggle-define- *deepfake*s-and-related-terms-technically-co mplex-legislation-proliferates. Acesso em: 19 maio 2025.

GPAI. Working Group on Misinformation and Synthetic Media. Paris: GPAI, 2023. Disponível em: https://gpai.ai/workstreams/. Acesso em: 20 set. 2025.

ILLINOIS. Biometric Information Privacy Act (BIPA), 740 ILCS 14. Springfield, 2008. Disponível em: https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57. Acesso em: 19 set. 2025.

INTERNATIONAL TELECOMMUNICATION UNION (ITU). *Report on AI-driven deepfakes and global elections.* Genebra: ITU, 2025.

KARNOUSKOS, S. Artificial Intelligence in Digital Media: The Era of *deepfake*s. *IEEE Transactions on Technology and Society*, v. 1, n. 3, p. 138–147, set. 2020. DOI: 10.1109/TTS.2020.3001312.

KOBISCHKO, Mariane; HEISS, Robert. Regulating *deepfake*s in the EU and Germany: Approaches and Perspectives. *Journal of Data Protection & Privacy*, v. 6, n. 2, p. 130–145, 2023.

KWOK, A. O.; KOH, S. G. *deepfake*: a social construction of technology perspective. *Current Issues in Tourism*, v. 24, n. 13, p. 1798–1802, 2021.

MAGNO, Jeaniel Carlos; MAGELA PIERONI, Geraldo. Os perigos do *deepfake* para a democracia brasileira. *Anais de Artigos do Seminário Internacional de Pesquisas em Midiatização e Processos Sociais*, v. 1, n. 6, 2024. ISSN 2675-4290.

MIGUEL, Raquel. Platforms AI Policy Updates in 2024: Labelling as the Silver Bullet. *Disinfo.eu*, 2024. Disponível em: https://www.disinfo.eu/publications/platforms-ai-policy-updates-in-2024-labelling-as-the-silver-bu llet/. Acesso em: 20 set. 2025.

MIT SLOAN SCHOOL OF MANAGEMENT. *deepfake*s explained. *Ideas Made to Matter*, 2020. Disponível em: https://mitsloan.mit.edu/ideas-made-to-matter/*deepfake*s-explained. Acesso em: 19 maio 2025.

MOLINA, Adriano Cezar; BERENGUEL, Orlando Leonardo. *deepfake*: a evolução das fake news. *Research, Society and Development*, v. 11, n. 6, e56211629533, 2022. DOI: http://dx.doi.org/10.33448/rsd-v11i6.29533. Acesso em: 19 maio 2025.

MOURA, Tarcísio. Direito e desinformação: fake news, *deepfake*s e a erosão da esfera pública digital. *Revista Brasileira de Políticas Públicas*, v. 9, n. 2, p. 289–312, 2019.

OLHAR DIGITAL. Nova York aprova lei pioneira que regula o uso de IA em contratações. 5 jul. 2023. Disponível em: https://olhardigital.com.br/2023/07/05/pro/nova-york-aprova-lei-pioneira-que-regula-o-uso-de-ia-em-contratacoes/. Acesso em: 19 set. 2025.

PATEL, Y. *et al*. *deepfake* Generation and Detection: Case Study and Challenges. *IEEE Access*, v. 11, p. 143296–143323, 2023. DOI: 10.1109/ACCESS.2023.3342107.

REUTERS. Illinois governor approves business-friendly overhaul of biometric privacy law. 5 ago. 2024. Disponível em: https://www.reuters.com/legal/government/illinois-governor-approves-business-friendly-overhaul-biometric-privacy-law-2024-08-05/. Acesso em: 19 set. 2025.

SANTOS, Pedro Henrique. EUA adotam novas diretrizes para IA. *Data Privacy Brasil*, 31 out. 2023. Disponível em: https://www.dataprivacybr.org/eua-adotam-novas-diretrizes-para-ia/. Acesso em: 19 set. 2025.

SATO, Mia. YouTube adds new AI-generated content labeling tool. *The Verge*, 18 mar. 2024. Disponível em: https://www.theverge.com/2024/3/18/24104743/youtube-ai-generated-content-disclosure-label. Acesso em: 20 set. 2025.

SIQUEIRA, João Pedro; ANDRADE, Camila. *deepfake* e Privacidade: uma análise jurídica acerca da manipulação da imagem dos usuários. *Revista de Direito, Tecnologia e Sociedade*, v. 10, n. 2, p. 55–78, 2024.

SIQUEIRA, M. de; ANDRADE, E. J. de. *deepfake* e privacidade: uma análise jurídica acerca da manipulação da imagem dos usuários. *Revista Foco*, v. 17, n. 8, e5679, 2024. DOI: https://doi.org/10.54751/revistafoco.v17n8-014.

TECH POLICY PRESS. US states struggle to define *deepfake*s and related terms as technically complex legislation proliferates. 2023. Disponível em: https://www.techpolicy.press/us-states-struggle-to-define- *deepfake*s-and-related-terms-as-technically-complex-legislation-proliferates/. Acesso em: 19 maio 2025.

UNIÃO EUROPEIA. Diretrizes Éticas para uma IA Confiável. High-Level Expert Group on Artificial Intelligence. Bruxelas, 2019. Disponível em: https://digital-strategy.ec.europa.eu/. Acesso em: 19 set. 2025.

UNIÃO EUROPEIA. Regulamento (UE) 2022/1925 do Parlamento Europeu e do Conselho, de 14 de setembro de 2022, relativo a mercados digitais contestáveis e equitativos (Digital Markets Act). *Jornal Oficial da União Europeia*, Bruxelas, 12 out. 2022. Disponível em: https://eur-lex.europa.eu/. Acesso em: 19 set. 2025.

UNIÃO EUROPEIA. Regulamento (UE) 2022/2065 do Parlamento Europeu e do Conselho, de 19 de outubro de 2022, relativo a um mercado único de serviços digitais (Digital Services Act). *Jornal Oficial da União Europeia*, Bruxelas, 27 out. 2022. Disponível em: https://eur-lex.europa.eu/. Acesso em: 19 set. 2025.

UNIÃO EUROPEIA. Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 12 de julho de 2024, relativo à Inteligência Artificial (*AI Act*). *Jornal Oficial da União Europeia*, Bruxelas, 12 jul. 2024. Disponível em: https://eur-lex.europa.eu/. Acesso em: 19 set. 2025.

UNIÃO EUROPEIA. *AI Liability Directive* – proposta de Diretiva de Responsabilidade Civil em IA. Bruxelas, 2024. Disponível em: https://eur-lex.europa.eu/. Acesso em: 19 set. 2025.

UNESCO. *Recommendation on the Ethics of Artificial Intelligence.* Paris: UNESCO, 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137. Acesso em: 2 maio 2025.

UOL. De Ísis Valverde a Anitta: quem já foi vítima de nudes falsos feitos em IA. 30 abr. 2025. Disponível em: https://www.uol.com.br/splash/noticias/2025/04/30/de-isis-

valverde-a-anitta-famosas-sao-vitima s-de-nudes-falsos-feitos-em-ia.htm. Acesso em: 7 maio 2025.

UOL. De William Bonner a Anitta: veja famosos vítimas de *deepfake*s criminosas. São Paulo, 26 mar. 2025. Disponível em: https://www.uol.com.br/splash/noticias/2025/03/26/de-william-bonner-a-anitta-veja-famosos-viti mas-de- *deepfake*s-criminosas.htm. Acesso em: 6 maio 2025.

VALOR. *deepfake*: criminosos usam IA para roubar R$ 129 milhões. *Valor Econômico*, 26 fev. 2024. Disponível em: https://valor.globo.com/patrocinado/dino/noticia/2024/02/26/ *deepfake*-criminosos-usam-ia-para-r oubar-r-129-milhoes.ghtml. Acesso em: 7 maio 2025.

VEJA. Casos de falsos nudes expõem lado sombrio da inteligência artificial. 3 nov. 2023. Disponível em: https://veja.abril.com.br/brasil/casos-de-falsos-nudes-expoem-lado-sombrio-da-inteligencia-artifi cial/. Acesso em: 7 maio 2025.

VERIFACT. Crimes com uso de *deepfake* disparam no Brasil. 14 abr. 2025. Disponível em: https://www.verifact.com.br/crimes-com-uso-de- *deepfake*/. Acesso em: 7 maio 2025.

WESTERLUND, Mika. The Emergence of *deepfake* Technology: A Review. 2019.

WHITE HOUSE. FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI. Washington, D.C., 21 jul. 2023. Disponível em: https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-har ris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/. Acesso em: 20 set. 2025.

VIANA, Natalia. Inteligência Artificial: EUA podem virar terra sem lei para IA. *Agência Pública*, São Paulo, 19 maio 2025. Disponível em: https://apublica.org/2025/05/inteligencia-artificial-eua-podem-virar-terra-sem-lei-para-ia/. Acesso em: 19 set. 2025.

Y. PATEL *et al*. *deepfake* Generation and Detection: Case Study and Challenges. *IEEE Access*, v. 11, p. 143296–143323, 2023. DOI: 10.1109/ACCESS.2023.3342107.