

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
FACULDADE DE DIREITO



Trabalho de Conclusão de Curso

**A PROLIFERAÇÃO DOS DEEPPFAKES CRIADOS A PARTIR DE
INTELIGÊNCIA ARTIFICIAL: DESAFIOS À PROTEÇÃO DOS
DIREITOS FUNDAMENTAIS NO CONTEXTO DO
CONSTITUCIONALISMO DIGITAL.**

Orientada: Laura Zieri Mesquita (RA00306672)

Orientadora: Prof. Dra. Flavia de Campos Pinheiro

SÃO PAULO

2025

LAURA ZIERI MESQUITA

**A PROLIFERAÇÃO DOS DEEPPAKES CRIADOS A PARTIR DE INTELIGÊNCIA
ARTIFICIAL: DESAFIOS À PROTEÇÃO DOS DIREITOS FUNDAMENTAIS NO
CONTEXTO DO CONSTITUCIONALISMO DIGITAL.**

Trabalho de Conclusão de Curso apresentado como
requisito para a obtenção do título de Bacharel em Direito
pela Pontifícia Universidade Católica de São Paulo - PUC
SP.

Orientadora: Prof. Dra. Flavia de Campos Pinheiro

SÃO PAULO

2025

Agradecimentos

Agradeço, em primeiro lugar, a Deus, ao meu anjo da guarda e a Vó Joaquina, por terem me guiado com sabedoria e me proporcionado tantos aprendizados nessa jornada.

Agradeço à Pontifícia Universidade Católica de São Paulo – PUC-SP, minha faculdade, por ter me acolhido e me proporcionado tantas realizações e aprendizados.

Agradeço a todos os seus funcionários, que foram essenciais no meu dia a dia na universidade, e ao corpo docente, que foram a base da minha formação jurídica.

Agradeço, em especial, à Professora Doutora Flávia de Campos Pinheiro, por sua orientação neste trabalho de conclusão de curso e por todas as contribuições prestadas durante a minha graduação. Sua presença desde o meu primeiro dia na universidade foi fundamental para a minha formação como profissional do Direito.

As minhas amigas da faculdade, deixo um agradecimento especial: a caminhada foi muito mais leve, rica e significativa com a companhia e o apoio de vocês.

Aos meus colegas de trabalho, expresso o meu agradecimento pela constante colaboração, apoio e encorajamento ao longo dessa jornada.

Ao meu namorado, expresso minha gratidão pelo companheirismo e por estar ao meu lado em todos os momentos, compartilhando comigo cada etapa dessa trajetória.

E, por fim, à minha família e aos meus familiares, que são a minha base, meu alicerce e minha maior bênção. Agradeço pelo amor incondicional, pelo apoio em todas as fases da minha vida e por me proverem tudo o que tenho e sou.

Resumo

O direito constrói suas próprias realidades. Com isso, os direitos fundamentais foram elaborados e incorporados no texto constitucional à luz de contínuas transformações sociais, as quais permanecem em constante evolução, sendo indubitavelmente necessários para garantir o Estado democrático de direito e impedir retrocessos legislativos. Assim, com o desenvolvimento tecnológico e econômico, novos desafios a sua proteção vem surgindo, como a criação e proliferação em massa dos *deepfakes*, conteúdos digitais produzidos com elevado grau de sofisticação e precisão. Nesse sentido, o uso malicioso desta tecnologia tem gerado uma série de prejuízos e violações legais, afrontando diretamente os direitos fundamentais à intimidade, à vida privada, à honra e à imagem, previstos no artigo 5º, inciso X, da Constituição da República Federativa do Brasil de 1988. Diante desse panorama, impõe-se a necessidade de compreender quais seriam os caminhos possíveis para regulamentar as tecnologias *deepfakes*, tanto no que diz respeito à responsabilidade do usuário, quanto das plataformas, a partir de um estudo jurídico internacional e nacional. Portanto, busca-se compreender como o direito vem acompanhado tais mudanças sociais e tecnológicas, permitindo, por um lado, a evolução das tecnologias digitais, e por outro, a manutenção e garantia dos direitos fundamentais alinhados à proteção da dignidade da pessoa humana.

Palavras-chave: Intimidade; Vida Privada; Honra; Imagem; Deepfake; Constitucionalismo Digital.

Abstract

Law constructs its own realities. Accordingly, fundamental rights were conceived and incorporated into the constitutional text in response to ongoing social transformations, which remain in constant evolution and are undeniably essential to safeguarding the democratic rule of law and preventing legislative setbacks. With technological and economic developments, however, new challenges to their protection have emerged, such as the creation and mass proliferation of *deepfakes*, digital content produced with a high degree of sophistication and precision. The malicious use of this technology has caused a series of harms and legal violations, directly infringing upon the fundamental rights to intimacy, privacy, honor, and image, as enshrined in Article 5, section X, of the 1988 Constitution of the Federative Republic of Brazil. In this context, it becomes imperative to explore possible pathways for regulating *deepfake* technologies, addressing both user and platform liability, through an analysis grounded in national and international legal frameworks. This study, therefore, seeks to understand how law has been keeping pace with such social and technological changes, enabling, on the one hand, the advancement of digital technologies while, on the other, ensuring the protection of fundamental rights consistent with the safeguarding of human dignity.

Keywords: Intimacy; Privacy; Honor; Image; Deepfake; Digital Constitutionalism.

Sumário

Introdução.....	7
1. Direitos Fundamentais.....	9
1.1. Evolução Histórica dos Direitos Fundamentais.....	9
1.2. A Proteção dos Direitos Fundamentais na Constituição Federal de 1988.....	10
1.3. Características e Princípios Intrínsecos dos Direitos Fundamentais.....	14
1.4. Constitucionalismo Digital e a Proteção de Direitos Fundamentais na Era da Informação.....	18
2. Deepfakes e Inteligência Artificial: Aspectos Técnicos e Jurídicos.....	20
2.1. A Aprendizagem das Máquinas: Como a Inteligência Artificial Processa e Materializa as <i>Deepfakes</i>	20
2.2. Conceito Teórico de <i>Deepfake</i> e os Efeitos Decorrentes de sua Propagação na Internet.....	22
2.3. Os Impactos Negativos da Proliferação dos Materiais de <i>Deepfakes</i> na Nação, na Sociedade e nos Indivíduos.....	26
2.4. Impactos Positivos do desenvolvimento das Tecnologias de <i>Deepfake</i>	33
3. Dos Caminhos para Regular o Uso de <i>Deepfakes</i>.....	36
3.1. A Necessidade de Regulação: Análise da Violação dos Direitos Fundamentais à Intimidade, Vida Privada, Honra e Imagem pelas <i>Deepfakes</i>	36
3.2. Uma Análise Jurídica Internacional sobre a Responsabilização de Usuários pela Criação e Propagação Indevida de <i>Deepfakes</i>	38
3.3. Os Desafios da Responsabilização Jurídica das Plataformas sob uma Ótica Internacional.....	42
3.4. Como o Direito Brasileiro Enfrenta o Desafio da Regulamentação dos <i>Deepfakes</i> ...	47
Considerações Finais.....	60
Referências Bibliográficas.....	64

Introdução

Apesar do cenário jurídico atual ser marcado por inúmeras divergências doutrinárias e até mesmo jurisprudências, não há o que se questionar quanto à importância dos direitos fundamentais, e a necessidade de sua preservação em prol do Estado democrático de direito. Previstos na Constituição de 1988, tais direitos são cláusulas pétreas, fruto de um processo histórico complexo influenciado por necessidades econômicas, governamentais e culturais de cada época, sendo desenvolvidos à luz das constantes transformações que irradiavam-se outrora, e que ainda se irradiam na contemporaneidade.

Nesse sentido, por meio deste trabalho de conclusão de curso (“TCC”), busca-se, *a priori*, investigar se os postulados que refletem nos direitos fundamentais, como o princípio da máxima efetividade (interpretação que assegure maior eficácia possível aos direitos fundamentais constitucionais), conseguem ainda ter força e relevância ante os avanços tecnológicos e a transformação da sociedade digital. Dentre os inúmeros direitos fundamentais existentes, para fins deste projeto, destaca-se o inciso X do artigo 5º da Constituição Federal, que garante a inviolabilidade da intimidade, vida privada, honra e imagem.

A proteção desses direitos tem se tornado cada vez mais complexa diante das novas tecnologias desenvolvidas diariamente, incluindo os chamados *deepfakes*. A expressão *deepfake* resulta da combinação de “*deep*” (derivada de “*deep learning*”, isto é, aprendizado profundo) e “*fake*” (“falso” em inglês). Em outras palavras, são conteúdos falsificados criados com um alto nível de elaboração e sofisticação tecnológica. Assim, com o uso da tal inteligência artificial (ou “IA”), são produzidos conteúdos fraudulentos e muitas vezes maliciosos, mediante a manipulação de elementos: (i) visuais (substituição de rostos ou alteração de ambientes); (ii) auditivos (substituição ou sobreposição de vozes); e/ou (iii) audiovisuais.

Nesse sentido, não há qualquer dúvida que a criação e disseminação de *deepfakes* tem um impacto direto nos direitos fundamentais, englobando ainda a afronta em um nível estatal - comprometendo a credibilidade das instituições públicas, interferindo no processo democrático e colocando em risco a segurança nacional -, social - disseminando a desinformação coletiva, complicando ainda mais o tecido social já fragilizado - e individual - relacionando-se diretamente à proteção dos direitos fundamentais de cada indivíduo.

Posto isso, como resposta a essa e tantas outras novas interações sociais cibernéticas, surge o fenômeno do constitucionalismo digital. Esse conceito baseia-se fundamentalmente

em um conjunto de iniciativas político-jurídicas destinadas a articular e sistematizar direitos, normas, mecanismos de governança e regras específicas, com o objetivo primordial de aplicar ao ambiente digital os limites e princípios estabelecidos pela ordem constitucional vigente.

Sob este prisma, o objetivo final deste TCC será examinar a frenética e incontrolável proliferação dos *deepfakes* produzidos por meio de inteligência artificial e os desafios que essa tecnologia apresenta à proteção dos direitos fundamentais no contexto do constitucionalismo digital, bem como, vislumbrar qual a resposta que o direito vem dado a essa realidade, contemplando um estudo sobre o cenário internacional e nacional que aduz sobre os mecanismos criados e utilizados em prol da proteção do direito à inviolabilidade da intimidade, vida privada, honra e imagem, de acordo com os princípios e garantias constitucionais e o preceito fundamental da dignidade da pessoa humana.

1. Direitos Fundamentais

1.1. Evolução Histórica dos Direitos Fundamentais

Os direitos fundamentais, expostos na Constituição da República Federativa do Brasil de 1988 (“CRFB/88”), não surgiram repentinamente, uma vez que foram fruto de uma evolução histórica, pautando o seu desenvolvimento nas necessidades econômicas, sociais e culturais que se intensificaram ao decorrer dos anos. Nesse sentido, Paulo Bonavides afirma que *“os direitos fundamentais passaram na ordem institucional a manifestar-se em três gerações sucessivas, que traduzem sem dúvida um processo cumulativo e quantitativo...”*.¹

A priori, para Edson Saleme os direitos de primeira dimensão surgem no final do século XVIII e caracterizam-se como uma resposta direta ao absolutismo, exigindo do Estado uma postura não interventiva na vida do indivíduo. Assim, foram frutos de revoluções liberais, francesas e norte-americanas, pela luta da burguesia em prol da liberdade individual, e a respectiva limitação do poder absoluto do Estado. Os direitos de primeira dimensão tem como rol exemplificativo as liberdades convencionais, como o direito à vida, à propriedade, à liberdade e à igualdade.

Ainda, o mesmo autor alude que os direitos de segunda dimensão estão relacionados diretamente com uma liberdade positiva, ou concreta, sendo marcados pela Revolução Industrial, a qual, a partir do século XIX pleiteou pela luta em prol do proletariado, visando a defesa dos direitos sociais, essenciais e básicos, como a alimentação, saúde e educação. Assim, nascem da necessidade de intervenção Estatal na vida social, a fim de coibir certos abusos do liberalismo, viabilizando a intervenção do Estado na economia.²

Por sua vez, os direitos de terceira dimensão são aqueles que dizem respeito aos princípios da fraternidade e solidariedade, pensados para atender a coletividade social e proteger os interesses dos seus titulares de forma coletiva ou difusa. Nesse sentido, Paulo Bonavides explica que esses direitos representam um novo patamar na evolução histórica das liberdades humanas, somando-se aos direitos de liberdade e igualdade. Dotados de forte carga humanitária e caráter universal, não se destinam a um sujeito específico, mas ao ser humano em sua totalidade, elevando a humanidade como destinatária principal e reconhecendo-a como valor fundamental em sua concretude existencial.³

¹ BONAVIDES, Paulo. **Curso de Direito Constitucional**. 19ª Edição, São Paulo: Malheiros, 2006.

² SALEME, Edson R. **Direito Constitucional**. 5ª Edição. Barueri: Manole, 2022. *E-book*. ISBN 9786555766370. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9786555766370/>. Acesso em: 17 mai. 2025.

³ Ob. Cit - Pág. 09.

Ainda, não é pacificado jurisprudencialmente o conceito de direitos de quarta geração, mas os doutrinadores que defendem a sua existência asseguram que eles estariam ligados aos desenvolvimentos humanos contemporâneos. Para Norberto Bobbio “*tratam-se dos direitos relacionados à engenharia genética.*”⁴; já para o doutrinador Marcelo Novelino, “*Os direitos fundamentais de quarta dimensão compreendiam o futuro da cidadania e correspondem à derradeira fase da institucionalização do Estado social sendo imprescindíveis para a realização e legitimidade da globalização política.*”⁵.

Nesse sentido seriam aqueles direitos que tem como objetivo acompanhar os avanços tecnológicos por meio de valores morais e culturais. De tal forma, que o novo cenário cibernético globalizado acaba exigindo do direito uma nova construção de princípios, regras e valores que possam adaptar as premissas já concebidas.

Por fim, os direitos de quinta dimensão, também debatidos a título doutrinário, seriam aqueles relacionados à paz mundial e à solidariedade entre os povos. Advindos de diferentes tragédias que acompanharam o desenvolvimento dos povos, como o atentado de 11 de setembro, tais direitos irradiar-se-iam em uma ótica plural e coletiva, protegendo a paz como um bem geral.

Posto isso, considerando a evolução histórica e doutrinária dos direitos fundamentais, fica evidente que a proteção da dignidade humana, da liberdade individual e dos direitos coletivos representa um alicerce central do eixo constitucional. Cada geração de direitos surge como resposta a novos desafios sociais, econômicos, tecnológicos e culturais, consolidando um modelo normativo que busca, de forma progressiva, atender às transformações da sociedade e garantir o pleno exercício da cidadania. É a partir dessa perspectiva evolutiva que se impõe uma análise sobre o tratamento conferido aos direitos fundamentais pela Constituição da República Federativa do Brasil de 1988.

1.2. A Proteção dos Direitos Fundamentais na Constituição Federal de 1988

Os direitos fundamentais, por sua vez, foram incorporados de forma sistemática aos textos constitucionais a partir do período pós Segunda Guerra Mundial, momento em que a comunidade internacional vivenciava um processo de reconstrução social, política e jurídica, marcado por uma crescente preocupação com a proteção efetiva da dignidade da pessoa humana, uma vez que essa foi fortemente violada pela força de regimes autoritários, como o

⁴ BOBBIO, Norberto. **A Era dos Direitos**. 1ª Edição. Rio de Janeiro: Campus, 1992.

⁵ NOVELINO, Marcelo. **Direito Constitucional**. 2ª Edição. São Paulo: Método, 2008.

nazismo, o fascismo e o stalinismo. Além disso, o cenário global enfrentava uma ameaça constante à estabilidade e à paz mundial, evidenciada não apenas pelos conflitos armados, mas também pela propagação de ideologias totalitárias que negligenciaram os direitos fundamentais.⁶

Segundo a interpretação de José Afonso da Silva, os direitos fundamentais possuem caráter dinâmico e não se encerram em um rol fixo, pois acompanham o desenvolvimento da sociedade, isto é, a cada avanço civilizatório, novas realidades reverberam, e com isso, surgem novos direitos que necessitam de proteção. Assim, aduz que o processo de reconhecimento desses direitos, por meio de declarações formais, é relativamente recente e está longe de alcançar sua completude. Trata-se, portanto, de um processo contínuo de afirmação da dignidade humana, que exige constante atualização e expansão do conceito de direitos fundamentais.⁷

O Brasil, por sua vez, iniciou um processo de redemocratização em 1985, após vivenciar 21 anos de um regime ditatorial (iniciado em 1964). Nesse sentido, tal movimentação acaba por desencadear na elaboração da CRFB/88, sendo essa caracterizada pelo grande avanço que propôs a proteção e garantia dos direitos fundamentais. Tal afirmação, inclusive, é aludida em seu preâmbulo, que garante o compromisso que se irradia ao longo de todo o texto constitucional:

*“... para instituir um Estado democrático, destinado a assegurar o exercício dos direitos sociais e individuais, a liberdade, a segurança, o bem-estar, o desenvolvimento, a igualdade e a justiça, como valores supremos de uma sociedade fraterna, pluralista e sem preconceitos, fundada na harmonia social e comprometida, na ordem interna e internacional, com a solução pacífica das controvérsias...”*⁸

O Título II da Magna Carta, intitulado “Dos Direitos e Garantias Fundamentais”, consagra direitos que são classificados como cláusulas pétreas a luz do ordenamento, nos termos do artigo 60, § 4º, inciso IV da CRFB/88, de tal forma que lhes é conferida uma

⁶ ABREU, Neide Maria Carvalho. **Direitos Fundamentais na Constituição Federal de 1988**. Anais do Conselho, 2007. Disponível em: <http://publicadireito.com.br/conpedi/manaus/arquivos/anais/XIVCongresso/055.pdf>. Acesso em 01 jul. de 2025.

⁷ SILVA, José Afonso da. **Curso de Direito Constitucional Positivo**. 19ª Edição. São Paulo: Malheiros Editores, 2000.

⁸ BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em 01 jul. de 2025.

proteção reforçada, uma vez que não pode nem se quer ser projeto de deliberação uma proposta legislativa tendente a aboli-los. Dentre esses direitos, merece destaque o inciso X do artigo 5º, que assegura a inviolabilidade da intimidade, da vida privada, da honra e da imagem das pessoas, garantindo, ainda, o direito à indenização por danos materiais ou morais decorrentes de sua violação.

A priori, um indivíduo teria os direitos à vida privada e à intimidade violados a luz da: (i) invasão de sua vida pessoal, familiar ou doméstica; (ii) atentado à sua integridade física ou mental, bem como à sua liberdade moral e intelectual; (iii) ofensa à sua reputação; (iv) apresentação distorcida de sua imagem ou contexto; (v) divulgação de informações irrelevantes ou constrangedoras ligadas à sua vida íntima; (vi) utilização indevida de seu nome, identidade ou imagem; (vii) práticas de vigilância não autorizada ou perseguição; (viii) Interferência em sua correspondência pessoal; (ix) uso impróprio de informações fornecidas de forma oral ou escrita; e/ou (x) compartilhamento indevido de dados obtidos sob obrigação de sigilo profissional.⁹

Com isso, a fim de iniciar uma análise sobre a diferenciação dos institutos (vida privada e intimidade), pode-se compreender que, a princípio, a existência de um indivíduo se desdobra em duas esferas fundamentais: a vida exterior e a vida interior. A vida exterior corresponde ao conjunto de manifestações públicas do sujeito, envolvendo sua atuação no ambiente social (como no trabalho, em eventos públicos, em interações profissionais ou comunitárias), isto é, todos os aspectos da vivência em sociedade que, por sua natureza, estão mais expostos ao olhar de terceiros. Já a vida interior diz respeito à dimensão mais reservada da existência, o que abarca os vínculos familiares e as interações afetivas mais íntimas, bem como os dados e informações que o indivíduo deseja manter fora do alcance do público, como sua orientação sexual, crenças, rotina pessoal, histórico de saúde ou movimentações financeiras.

Assim, para José Cavero e Vidal Serrano, dentro da vida interior temos ainda a "vida privada" e a "intimidade". Segundo essa análise, a vida privada diz respeito ao âmbito da existência individual que se estrutura na ausência da exposição pública, ou seja, corresponde à esfera de conforto do indivíduo, permanecendo restritas ao ambiente doméstico e familiar, com aversão a qualquer tipo de intrusão externa. Essa dimensão compreende, por exemplo, a

⁹ GUERRA, Sidney. **Direito fundamental à intimidade, vida privada, honra e imagem**. 2º Volume. Rio de Janeiro: Freitas Bastos, 2006. Disponível em: http://www.publicadireito.com.br/conpedi/manaus/arquivos/anais/recife/direitos_fundam_sidney_guerra.pdf. Acesso em 01 jul. 2025.

rotina doméstica, as preferências pessoais, os hábitos, e outros aspectos do cotidiano que o sujeito deseja manter sob discrição e domínio próprio.

Já o conceito de intimidade, para os mesmos autores, é ainda mais restrito e específico, referindo-se à dimensão interpessoal da vida privada, isto é, aos aspectos mais profundos, subjetivos e sensíveis da experiência pessoal, compartilhados com um número muito limitado de pessoas com vínculos de extrema confiança. A intimidade, nesse sentido, representa o último reduto da privacidade, cuja violação pode atingir diretamente o núcleo da dignidade e da integridade moral do indivíduo.¹⁰ Assim, diz respeito aos e sentimentos, medos, angústias e segredos. Apesar dessa distinção conceitual, não há consenso doutrinário ou jurisprudencial claro sobre a diferenciação entre a privacidade e a intimidade, o que evidencia a complexidade e a sensibilidade envolvidas no tema.

Por sua vez, o direito à honra diz respeito à proteção dos atributos que compõem a dignidade individual, como o reconhecimento social, a boa reputação, o apreço dos demais e a integridade moral. Trata-se do direito de ser respeitado, de não ser alvo de ofensas, humilhações ou qualquer forma de ataque. A honra está profundamente ligada aos princípios éticos e morais que orientam a conduta de um indivíduo, refletindo-se na possibilidade de manter-se respeitado perante a sociedade, preservar um nome limpo e cultivar a autopercepção de ser uma pessoa íntegra e digna de respeito.¹¹

O direito à honra envolve a forma como o indivíduo é percebido socialmente, sendo essa imagem essencial para sua valorização pessoal e inserção nos diversos contextos da vida em sociedade. Quando esse direito é violado, a pessoa afetada sente-se rebaixada, desrespeitada e exposta a situações constrangedoras, o que pode acarretar danos significativos, tanto emocionais quanto patrimoniais. Diante disso, torna-se fundamental garantir a proteção da honra, uma vez que, como observa Carlos Alberto Bittar, a opinião pública reage com grande sensibilidade a informações negativas sobre alguém.¹² O ordenamento jurídico, portanto, deve resguardar esse valor não apenas para assegurar o bem-estar e a autoestima do indivíduo, mas também para garantir que ele possa se desenvolver plenamente nos âmbitos social, econômico, profissional e político.

¹⁰ CAVERO, José Martinez De Pisón; NUNES JÚNIOR, Vidal Serrano. **A proteção Constitucional da Informação e o Direito à Crítica Jornalística**. São Paulo: FTD, 1997.

¹¹ BUENO, Francisco da Silveira. **Dicionário Escolar da Língua Portuguesa**. Rio de Janeiro: FENAME, 1956.

¹² BITTAR, Carlos Alberto. **Os Direitos da Personalidade**. Rio de Janeiro: Forense Universitária, 1995.

Já o direito a imagem, para José Afonso da Silva, refere-se à proteção da representação da personalidade física - incluindo características faciais, postura, gestos, expressões, vestimentas, entre outros - bem como os aspectos morais que permeiam a vida do indivíduo - como a reputação, fama e a identidade subjetiva do indivíduo. Também é considerado um direito de personalidade quando envolve a reprodução de traços visuais, da voz ou dos gestos que identificam a pessoa. Esse direito é reconhecido como um bem jurídico inviolável, destinado à garantia da integridade da figura humana, assegurando que ninguém utilize a imagem de outrem sem o seu prévio e exposto consentimento.

Com isso, o uso legítimo da imagem depende necessariamente da autorização do titular, de tal forma que, na ausência desta, a sua divulgação é considerada como indevida, impondo-se a necessidade de reparar os danos causados. A proteção conferida ao direito à imagem é abrangente, alcançando inclusive a vedação de modificações não autorizadas em materiais audiovisuais, como filmes, vídeos, discos, publicações impressas e eletrônicas, transmissões televisivas, e conteúdos digitais em geral.¹³

Posto isso, conclui-se que os direitos à intimidade, à vida privada, à honra e à imagem constituem dimensões essenciais da personalidade humana, interligados de forma indissociável e protegidos como garantias fundamentais que asseguram a dignidade da pessoa. Esses direitos, ao preservarem a esfera mais íntima e o reconhecimento social do indivíduo, representam pilares indispensáveis para o pleno exercício da cidadania e do respeito à autonomia pessoal. No contexto digital, sua proteção torna-se ainda mais premente, diante das particularidades do ambiente virtual, que potencializam a sua exposição indevida, a disseminação rápida de informações e a vulnerabilidade das pessoas a ataques e violações cibernéticas. Assim, a regularização desses direitos fundamentais na esfera online é imperativa, exigindo a adoção de mecanismos eficazes, considerando, inclusive, as características e princípios essenciais que irradiam o postulado dos direitos fundamentais.

1.3. Características e Princípios Intrínsecos dos Direitos Fundamentais

A fim de compreender até que ponto se chega a extensão e aplicabilidade prática dos direitos fundamentais, vale salientar a respeito de suas principais características e princípios norteadores. Nesse sentido, partindo da premissa de que os direitos fundamentais consistem em prerrogativas que visam proteger a dignidade da pessoa humana em suas diversas

¹³ Ob. Cit - Pág. 12.

dimensões, esses se manifestam em diferentes esferas: por meio dos direitos de liberdade, assegurando a autonomia individual frente ao Estado; pelos direitos sociais ou de igualdade, protegendo o indivíduo em suas necessidades básicas; e pelos direitos de solidariedade, buscando a preservação coletiva da humanidade.

Nesse sentido, entre as principais características que qualificam um direito como fundamental, destaca-se, em primeiro lugar, a historicidade. Assim, os direitos fundamentais não surgiram de maneira instantânea, mas sim são fruto de um conjunto sistêmico de lutas sociais e conquistas gradativas da humanidade. Por conseguinte, cada um desses direitos advém de necessidades que foram pleiteadas em determinadas épocas. Ainda, é interessante analisar que um mesmo direito fundamental pode ter diferentes aplicabilidades com o passar do tempo, de acordo com o contexto em que se encontra. Por exemplo, os direitos expostos no meio do artigo 5, inciso X da CRFB/88, que, diferentemente de outrora, hoje são pensados inclusive em um âmbito digital, com o avanço tecnológico e exposição excessiva da vida privada e da intimidade de cada um.

Assim, em decorrência de tais mudanças sociais é possível observar uma quebra de paradigmas na coletividade em que a inviolabilidade da intimidade, da vida privada, da honra e da imagem entram com ainda mais força no mundo digital. Nesse sentido, em 1988, quando tal artigo foi reverberado no texto constitucional, não existiam redes sociais, e nem se quer sabia-se sobre os denominados influenciadores digitais. Hoje em dia há uma diferença clara entre pessoas públicas e anônimas nas redes sociais. Com a mudança no mundo tangível surgem diferentes questionamentos jurídicos, como se seria válido que pessoas notoriamente conhecidas ou "famosas" tivessem uma menor proteção jurídica desses direitos em comparação a cidadãos anônimos, podendo as plataformas e usuários utilizarem suas imagens de forma livre e indiscriminada.

Para responder a este questionamento, levanta-se uma segunda característica de tais direitos: a universalidade. Isso significa que, no geral, os direitos fundamentais não podem ser restritos a determinados grupos, categorias ou classes sociais. Embora alguns direitos sejam específicos - como os direitos trabalhistas -, a universalidade garante que, dentro do grupo alcançado pela norma, todos os seus membros gozarão dos mesmos direitos, sem qualquer forma de discriminação.

O que ocorre é que a invasão de privacidade de determinadas pessoas famosas acaba sendo justificada pela existência de um interesse público nestas vidas. No entanto, não parece correto que tal premissa possibilite a livre circulação e manipulação de imagens de pessoas

públicas em redes sociais. Assim, caso a caso precisa ser analisado a fim de compreender-se o ocorrido, o que irradia outra característica dos direitos fundamentais, a qual seria a sua limitabilidade. Nenhum direito fundamental possui caráter absoluto. Isso significa que o exercício de um direito encontra limites quando colide com outros direitos de igual hierarquia. Nesses casos, o intérprete deve realizar uma ponderação para assegurar que nenhum dos direitos seja totalmente anulado. O critério normalmente utilizado é o da cedência recíproca, pelo qual cada direito cede parcialmente para que ambos coexistam de forma equilibrada, preservando seus respectivos núcleos essenciais.

Mesmo assim, na paralela, deve-se considerar a irrenunciabilidade de tais direitos, o que significa que não se pode abrir mão deles de forma definitiva. Um indivíduo não pode, por exemplo, renunciar de modo absoluto à sua honra, imagem ou privacidade. Isso não impede, contudo, que ele deixe de exercer certos direitos ou ainda, que os limite temporariamente, especialmente quando o faz voluntariamente e em contextos juridicamente tolerados, como ocorre todos os dias quando milhares de pessoas desprovidas, ou não, de fama dispõem de suas imagens, privacidade e intimidade voluntariamente na internet.

Outra característica relevante dos direitos fundamentais é sua concorrência, ou seja, uma mesma conduta pode ser protegida por mais de um direito simultaneamente. Por exemplo, isso é observado quando se produz um conteúdo fraudulento sobre outrem, o colocando em uma situação prejudicial a sua reputação, e com isso, abala-se tanto o direito à honra, quanto a imagem daquele indivíduo.

Além da conformidade com suas características, a aplicação prática e concreta desses direitos fundamentais no cotidiano normativo, por sua vez, deve ser norteadas por seus respectivos princípios estruturantes, sendo postulados de interpretação constitucional. Tais postulados constituem um conjunto sistemático de diretrizes interpretativas que orientam o trabalho do intérprete jurídico no processo de atribuição de significado às normas constitucionais.

O primeiro deles, o princípio da unidade da Constituição, parte da premissa de que a Constituição é um sistema orgânico: nenhuma de suas disposições existe isoladamente, pois todas se integram num todo harmônico estruturado por princípios. Por isso, a interpretação deve evitar contradições internas, reconhecendo que a Constituição, enquanto fundamento máximo da ordem jurídica, goza de supremacia e deve ser compreendida de maneira sistemática.

Além disso, também deve ser observado o princípio da concordância prática (ou harmonização), o qual garante que quando dois ou mais direitos fundamentais colidem caberia ao intérprete jurídico ponderar valores à luz do caso concreto, reduzindo a incidência de um deles apenas na medida estritamente necessária para preservar o núcleo do outro, sempre com observância dos critérios de necessidade, adequação e proporcionalidade em sentido estrito.

Já o princípio do efeito integrado determina que se busque conciliar e articular os diversos campos políticos e sociais refletidos no texto constitucional, assegurando coesão às políticas públicas e à atuação estatal. Em complemento, o princípio da máxima efetividade orienta a interpretação no sentido de conferir às normas constitucionais relativas aos direitos fundamentais a mais ampla e intensa eficácia possível, de modo a assegurar que esses direitos sejam plenamente contemplados na prática, alcançando sua função protetiva e garantindo sua aplicabilidade concreta em todas as esferas da vida social.

Por fim, o princípio da justeza volta-se de maneira específica à atuação do Supremo Tribunal Federal, estabelecendo que, no exercício da função de guardião da Constituição, esse tribunal deve respeitar rigorosamente os limites institucionais e funcionais que competem ao Poder Judiciário dentro da estrutura do Estado democrático de direito. Esse postulado atua como um fator moderador, orientando a Corte a manter o equilíbrio entre a necessária aplicação da força normativa da Constituição e o respeito à separação de poderes, evitando excessos que possam invadir competências dos demais órgãos constitucionais.

Assim, observa-se que tais princípios moldam a forma com que tais direitos devem ser observados e garantidos. Nesse sentido, quando tratamos dos direitos à inviolabilidade da intimidade, da vida privada, da honra e da imagem, estamos tratando de um rol de garantias que deve ser interpretado de acordo com todo o texto constitucional e as máximas que o irradiam, além disso, devem ser considerados cotidianamente na atuação do poder público, em todas as suas esferas, tendo a mais ampla e intensa eficácia possível, sendo ainda rigorosamente respeitados pelo Supremo Tribunal Federal. Logo, essa noção é o ponto de partida básico que faz construir o olhar que deve-se ter ao analisar esses direitos. No entanto, com a consciência de que ao mesmo tempo, eles podem ser ponderados em detrimento do caso concreto e não são absolutos.

Com isso, na era moderna surgem preocupações e debates, entre a proteção de tais direitos e a desenfreada circulação de conteúdos que os abalam, uma vez que para Celso Bastos o avanço das tecnologias de comunicação e vigilância nos meios eletrônicos ampliou

de maneira significativa, inclusive, os riscos de exposição indevida da vida pessoal dos indivíduos. Recursos como câmeras de longo alcance, equipamentos de captação sonora e, mais recentemente, ferramentas digitais sofisticadas, permitem que aspectos antes restritos ao âmbito privado sejam facilmente acessados e divulgados sem autorização.¹⁴

Em busca de analisar como deve se dar a aplicação dos direitos fundamentais no nosso atual cenário tecnológico, surge então o constitucionalismo digital, o qual parte das características e princípios supramencionados para disciplinar a garantia dos direitos fundamentais na contemporaneidade. Assim, o objetivo de tal mobilização é entender como tornar a proteção de tais direitos eficaz ante as enormes transformações sociais que são vivenciadas diariamente.

1.4. Constitucionalismo Digital e a Proteção de Direitos Fundamentais na Era da Informação

A era digital, fundamental para o desenvolvimento humano, adveio com uma abrupta mudança no comportamento das sociedades, a qual passou por uma transformação em sua própria forma de se comunicar, por meio do compartilhamento instantâneo e frenético de informações. Tais novas tecnologias passaram por grandes transformações na última década, se tornando essenciais e até mesmo habituais no cotidiano da população. Para Camila Takano, foi possível, inclusive, observar impactos positivos dessas tecnologias, com a modernização dos governos, os quais puderam aprimorar a sua forma de gestão, nacional e internacionalmente; a maior autonomia administrativa das instituições, em relação aos seus produtos e serviços; a independência iminente de grande parte das redes sociais; e a facilidade do acesso à informação aos indivíduos.

Ainda, a mesma autora alude que no final do século XX tal crescimento tecnológico atingiu, inclusive, a seara da biotecnologia, mercado financeiro e a quebra de barreiras entre lugares, impactando significativamente as relações de trabalho. Desse modo, a tecnologia da informação pode ser considerada a principal ferramenta do Estado moderno, assim como a energia foi para o período da Revolução Industrial (séc. XVIII). O mundo se tornou essencialmente digital, sendo a internet onipresente na vida social.¹⁵

¹⁴ BASTOS, Celso Ribeiro. **Curso de Direito Constitucional**. São Paulo: Saraiva, 1997.

¹⁵ TAKANO, Camila Cardoso; DA SILVA, Lucas Gonçalves. **O Constitucionalismo Digital e as Novas Tecnologias da Informação e Comunicação (TIC)**. 6º Volume. Revista de Direito, Governança e Novas Tecnologias, 2020.

O processo de transformação tecnológica, por sua vez, revolucionou o paradigma até então vigente, introduzindo novas dinâmicas nas esferas econômica, cultural, geopolítica e nas formas de existência humana. Essa profunda mudança, por consequência, impôs ao direito o desafio de repensar seus fundamentos e adaptar-se aos novos modelos que passaram a estruturar a sociedade contemporânea, uma vez que ao lado dos benefícios da revolução digital, vieram também os malefícios, como os cibercrimes e as inúmeras violações de direitos que são realizadas em âmbito digital.

Com isso, adveio o constitucionalismo digital, o qual consiste em um conjunto de iniciativas político-jurídicas voltadas à articulação e sistematização de direitos, normas, mecanismos de governança e regras específicas, cujo escopo primordial é trazer ao âmbito digital os limites e princípios estabelecidos pela ordem constitucional vigente.¹⁶ Tal necessidade é particularmente sensível no que diz respeito aos direitos à intimidade, à vida privada, à honra e à imagem das pessoas, previstos no inciso X do artigo 5º da Constituição Federal de 1988.

Esses direitos, ao serem transpostos para o ambiente digital, tornam-se ainda mais vulneráveis, uma vez que podem ser violados ou manipulados com grande facilidade por meio do uso de tecnologias avançadas, como algoritmos de inteligência artificial, que são capazes de coletar, reproduzir, distorcer e divulgar informações pessoais em larga escala. Essas transformações revelam as limitações do modelo constitucional tradicional, que foi originalmente elaborado para uma realidade analógica.

As mudanças provocadas pela revolução digital - em nossa relação conosco mesmos, com os outros e com a sociedade como um todo - ocorrem sob um conjunto de normas constitucionais que foram concebidas para comunidades analógicas. Assim, a fim de que essas tecnologias continuem gerando impactos positivos na sociedade, é fundamental que seu desenvolvimento e uso estejam orientados pelo interesse público e pela garantia dos direitos fundamentais. Essa necessidade torna-se ainda mais urgente diante de práticas como a criação de vídeos falsos - os chamados *deepfakes* -, que tornam possível e cada vez mais comum a manipulação da imagem de pessoas, fazendo-as parecer dizer ou fazer algo que jamais ocorreu.¹⁷

¹⁶ GILL, Lex; REDEKER Dennis; GASSER., Urs. **Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights**. The Berkman Center for Internet & Society at Harvard University. n° 2015-15. Disponível em: <http://dx.doi.org/10.2139/ssrn.2687120>. Acesso em 02 jul. 2025.

¹⁷ CELESTE, Edoardo. **Digital Constitutionalism: The Role of Internet Bills of Rights**. Nova Iorque: Routledge, 2023.

Posto isso, é evidente que o direito, especialmente em sua dimensão constitucional, precisa acompanhar os desafios impostos pela nova realidade digital. O surgimento de práticas como os *deepfakes* evidencia não apenas a potência da inteligência artificial, mas também os riscos que ela representa à integridade dos direitos fundamentais. Assim, torna-se indispensável aprofundar a compreensão sobre o funcionamento dessas tecnologias - especialmente da inteligência artificial e sua aplicação na criação de conteúdos falsificados - para que se possa construir mecanismos normativos mais eficazes e condizentes com a complexidade do mundo digital.

2. *Deepfakes* e Inteligência Artificial: Aspectos Técnicos e Jurídicos

2.1. A Aprendizagem das Máquinas: Como a Inteligência Artificial Processa e Materializa as *Deepfakes*

A manipulação de vídeos e imagens na internet não é uma prática recente. Todavia, no início do desenvolvimento digital, a manipulação de tais conteúdos se pautava nos seus cortes e na habilidade de copiar e colar esses *frames* com precisão. Atualmente, com o advento de novas tecnologias, é possível criar novos códigos e manipulá-los sem que seja realizada a elaboração do tradicional método de edição. Em suma, foi aperfeiçoada a forma de elaborar conteúdos falsos via internet, sendo que um dos pilares desse desenvolvimento foi a concepção de um novo processo denominado “Auto-Encoder Network”, o qual pode ser subdividido em duas etapas: a codificação e a decodificação de dados.¹⁸

Em resumo, a codificação é responsável por transformar dados em um código específico, ou seja, no caso dos *deepfakes*, por exemplo, ele comprimiria imagens faciais, extraindo suas características, e as transformaria em valores vetoriais em um espaço latente. Por sua vez, o decodificador tem o papel de traduzir esse código de volta a sua forma original, isto é, ele reconstruiria o rosto original, com base nos dados e nas características faciais, objetivando fazer com que esses dados sejam o mais próximo possível dos extraídos pelo codificador.¹⁹

¹⁸ RAMADHANI, K. N.; MUNIR R. **A Comparative Study of Deepfake Video Detection Method**. Yogyakarta, Indonesia: International Conference on Information and Communications Technology (ICOIACT), 2020.

¹⁹ YUZH, Zhang; RUIFANG, Wang; LIANG, Zhu. **The Review of Generation and Detection Techniques for Deepfakes**. Journal of Information Security Research, 2022. Disponível em: <https://www.atlantis-press.com/proceedings/ic-icaie-22/125981029>. Acesso em 02 jul. 2025.

Nesse sentido, em primeira instância, dois codificadores distintos extraem as características faciais da imagem de origem e da imagem de destino, respectivamente. Em seguida, dois decodificadores distintos reconstroem as expressões faciais correspondentes. Assim, com essa troca de informações entre o codificador e o decodificador a imagem falsa é gerada, uma falsificação tão convincente que, muitas vezes, é praticamente indistinguível de uma imagem real.

Outra tecnologia surgiu, o que demonstra a evolução da produção de conteúdos de *deepfake*, de tal modo que, por sua vez, a “Rede Generativa Adversarial” (GAN), proposta por Ian J. Goodfellow e colaboradores em outubro de 2014²⁰, introduziu um mecanismo baseado na contraposição, que consiste na geração e detecção repetitiva de dados internos pelo próprio algoritmo, em um processo de competição contínua entre duas redes neurais. Assim, dentro da estrutura GAN, há dois componentes principais que operam de forma simultânea e independente: o gerador e o discriminador.

Em suma, o gerador é responsável por criar imagens ou vídeos forjados que sejam convincentes. Por sua vez, o discriminador avalia a autenticidade das amostras geradas, comparando-as com imagens reais, assim, ele envia os resultados da análise de volta ao gerador, indicando quais elementos precisam ser ajustados. Esse processo ocorre de forma contínua: o gerador ajusta suas criações com base no feedback recebido pelo discriminador, corrigindo imperfeições e aumentando o grau de realismo. Ambos os componentes operam em ciclos sucessivos de aperfeiçoamento até que estejam plenamente ajustados, isto é, até que o discriminador não consiga mais diferenciar, com precisão, as imagens geradas das imagens reais.²¹

Ou seja, ainda que o resultado final do trabalho seja uma imagem falsa, a produção de um vídeo, áudio e/ou imagem *deepfake* requer, necessariamente, a utilização de imagens reais de indivíduos, para que os dados deles sejam extraídos e, de alguma forma, replicados na *deepfake*. Essa prática é especialmente sensível sob uma perspectiva jurídica, isso porque, ao se codificar a imagem de uma pessoa para gerar outra visualmente realista, baseada na análise

²⁰ GOODFELLOW, Ian. **Generative Adversarial Nets**. Advances in Neural Information Processing Systems, 2014. (apud. UZHI, Zhang; RUIFANG, Wang; LIANG, Zhu. **The Review of Generation and Detection Techniques for Deepfakes**. Journal of Information Security Research, 2022.)

²¹ LIU, Min; XIJIN, Zhang. **Deepfake Technology and Current Legal Status of It**. 3rd International Conference on Artificial Intelligence and Education (IC-ICAIE 2022), 2022. (apud. UZHI, Zhang; RUIFANG, Wang; LIANG, Zhu. **The Review of Generation and Detection Techniques for Deepfakes**. Journal of Information Security Research, 2022.)

e processamento de seus dados faciais, há uma intervenção profunda em sua esfera pessoal e infração aos seus direitos fundamentais.

Posto isso, torna-se necessário compreender a fundo o conceito formal de *deepfake*, bem como as suas distintas formas de expressão, e, inclusive, como esses materiais podem se perpetuar, facilmente, por meio das redes sociais durante a era digital, estando cada vez mais presentes no cotidiano, com potencial de influenciar comportamentos, decisões e realidades de maneira preocupante.

2.2. Conceito Teórico de *Deepfake* e os Efeitos Decorrentes de sua Propagação na Internet

A expressão *deepfake* é concebida pela combinação da palavra *deep* (extraída de deep learning, “aprendizado profundo”) e *fake* (“falso” em inglês). Ou seja, são conteúdos falsos produzidos com um alto grau de elaboração e sofisticação. No contexto contemporâneo, a partir do emprego da inteligência artificial, produzem-se imagens, áudios ou vídeos fraudulentos mediante a adulteração de elementos: (i) visuais, tais como substituição de rostos ou alteração de ambientes; (ii) auditivos, incluindo a substituição ou sobreposição de vozes, bem como a criação fictícia de diálogos; e/ou (iii) audiovisuais preexistentes.

Há ainda quem as denomine de “novas fake news”, considerando que o objetivo subjacente desses materiais, inevitavelmente, é induzir terceiros a erro, fazendo-os acreditar na veracidade de fatos ou situações que, na realidade, jamais ocorreram. Ou seja, em razão de seu elevado grau de verossimilhança e da capacidade de reproduzir, com extrema fidelidade, a aparência, a voz e os gestos de pessoas reais, além de simular contextos e eventos inexistentes, tais conteúdos constituem instrumentos de engano potencialmente lesivos e de difícil detecção, ensejando sérios riscos à tutela da dignidade da pessoa humana, à honra, à imagem e à própria segurança da informação no ambiente digital, de acordo com o “Guia Ilustrado Contra as *Deepfakes*” publicado pelo governo brasileiro.²²

As *deepfakes* podem se manifestar por meio de distintas modalidades, e entre as formas mais recorrentes, destaca-se a denominada de *face swap*, isto é, a substituição de rostos, a qual se dá quando a fisionomia de um indivíduo é sobreposta ao corpo de outro,

²² BRASIL. **Guia Ilustrado Contra as Deepfakes**. Data Privacy Brasil. Brasília: STF, Coordenadoria de Combate à Desinformação, 2024. Disponível em: <https://noticias.stf.jus.br/postsnoticias/supremo-lanca-guia-ilustrado-contra-as-deepfakes/>. Acesso em 09 jul. 2025.

simulando situações nunca ocorridas.²³ Inclusive, a título exemplificativo, foi constatado que na América Latina, especialmente no Brasil, houve um crescimento alarmante de casos de fraudes bancárias cometidas por meio da falsificação de imagens para burlar sistemas de autenticação por meio do reconhecimento facial

Inclusive, de acordo com a matéria, tal preocupação se irradia também em diferentes países ao redor do mundo, gerando um impacto global, de tal forma que um caso que tornou-se notadamente conhecido foi o ocorrido na China, no qual um executivo foi convencido de realizar uma transferência bancária de US\$20 milhões (vinte milhões de dólares norte-americanos), por meio da falsificação de uma videochamada. Tais práticas configuram modalidades sofisticadas de crimes cibernéticos, envolvendo falsidade ideológica, estelionato digital e violação de dados pessoais, e revelam que os mecanismos tradicionais de controle e segurança virtual ainda restam-se insuficientes.

Assim, a reportagem alude que, em resposta, empresas de tecnologia têm desenvolvido soluções preventivas voltadas à proteção de sistemas de identificação biométrica. Este foi o caso da Appdome, que lançou novas ferramentas capazes de detectar tentativas de fraude com uso de vídeos manipulados por câmeras virtuais. As referidas soluções utilizam métodos que pautam-se na análise de textura, profundidade e iluminação facial para identificar a veracidade da imagem do indivíduo. Além disso, também trabalham na geração de relatórios técnicos que subsidiam medidas corretivas e ações preventivas por parte das instituições afetadas. De acordo com Tom Tovar, CEO e cofundador da Appdome: *“Não conseguimos impedir que deepfakes sejam criados, mas conseguimos barrar o uso deles dentro dos aplicativos móveis”*.²⁴

Além do abalo econômico, também é possível vislumbrar os abalos reputacionais que podem ser desencadeados pela proliferação das *deepfakes*, o que chegou até o próprio Vaticano. No caso, houve a proliferação de imagens sintéticas geradas por inteligência artificial do Papa Francisco vestindo uma jaqueta acolchoada de alto luxo (**“Figura 1”**). Embora se tratasse de uma renderização produzida por software de IA (Midjourney), a aparência realista da imagem levou milhares de usuários a acreditarem em sua autenticidade, criticando o sacerdote pela suposta “ostentação”, o que, por si, demonstra o potencial das

²³ Ob. Cit. - Pág. 22.

²⁴ DINO. **Deepfake está clonando rostos para roubar contas bancárias**. Estado de Minas. 2025. Disponível em: <https://www.em.com.br/mundo-corporativo/2025/03/7077961-deepfake-esta-clonando-rostos-para-roubar-contas-bancarias.html>. Acesso em 09 jul. 2025.

deepfakes em induzir os indivíduos ao erro em larga escala e o abalo à honra e à imagem de suas vítimas.²⁵



Figura 1: Imagem gerada pela inteligência artificial Midjourney do Papa Francisco vestido com uma jaqueta acolchoada de alto luxo.

Ademais, segundo o "Guia Ilustrado Contra as *Deepfakes*" publicado pelo governo brasileiro há também a clonagem de voz, aparência e expressões faciais, técnica que permite a geração de vídeos em que o indivíduo artificialmente reproduz, de forma precisa, falas, entonações e gestos realizados originalmente por um terceiro. Tal prática, conhecida no meio técnico como *puppet-master* (ou "jogo do ventríloquo"), permite simular a atuação real de um indivíduo em uma circunstância falsa.²⁶

Um outro recente caso ocorrido em Hong Kong tratou que um funcionário do setor financeiro de uma multinacional foi induzido a realizar uma transferência de aproximadamente US\$ 25 milhões (vinte e cinco milhões de dólares norte americanos) após participar de uma videoconferência forjada, na qual todos os participantes (supostamente colegas de trabalho, inclusive o diretor financeiro da empresa) eram, na verdade, criações sintéticas produzidas com o uso de inteligência artificial.

Segundo a polícia local, o golpe se baseou na manipulação audiovisual de imagens públicas, com o uso de *deepfakes* para simular com alta verossimilhança rostos e vozes de pessoas reais, enganando inclusive sistemas de videoconferência e reconhecimento facial. O

²⁵ ELLERY, Simon. **Fotos falsas do Papa Francisco com uma Jaqueta Acolchoada Viralizam, Destacando o Poder e o Perigo da IA.** CBS News. 2023. Disponível em: <https://www.cbsnews.com/news/pope-francis-puffer-jacket-fake-photos-deepfake-power-peril-of-ai/>. Acesso em 09 jul. 2025.

²⁶ Ob. Cit - Pág. 22.

funcionário, embora inicialmente suspeitasse de tentativa de *phishing*, foi convencido da legitimidade da solicitação após a videochamada, na qual reconheceu os supostos colegas. Além desse episódio, as autoridades relataram a existência de diversas outras fraudes semelhantes no mesmo período, incluindo o uso de identidades roubadas combinadas com imagens *deepfake* para abrir contas bancárias e solicitar empréstimos.²⁷

Ademais, segundo o guia publicado pelas autoridades brasileiras, outra técnica amplamente utilizada é a modificação da região labial (*lip-sync*), que sincroniza os movimentos dos lábios de uma figura humana com um áudio previamente inserido, conferindo ao conteúdo manipulado a aparência de autenticidade.²⁸ Um exemplo emblemático desse cenário foi o caso do médico Drauzio Varella, que se tornou vítima da divulgação de vídeos falsos nas redes sociais. Em um dos casos, um vídeo publicado na plataforma TikTok simulava o médico promovendo um suposto produto cosmético, com voz clonada e imagens manipuladas, com o objetivo de induzir consumidores a erro e favorecer práticas comerciais enganosas.²⁹

Tais conteúdos *deepfake*, inclusive, não se restringem à superfície visível da internet (*surface web*), sendo disseminados também na *dark web* e na *deep web*. A *surface web* é a camada da internet acessível ao público em geral, composta por informações armazenadas em servidores e que podem ser encontradas por meio de mecanismos de busca comuns, como o Google, uma vez que suas páginas são indexadas. Já a *deep web* compreende conteúdos não indexados, acessíveis apenas mediante URL direto e autenticação. Por fim, a *dark web*, porção restrita da *deep web*, exige softwares específicos para acesso e está frequentemente associada a práticas ilícitas.³⁰

Nesses ambientes mais obscuros, a gravidade do material da *deepfakes* se intensifica, de tal forma que são disseminados não apenas vídeos relacionados as fraudes ou campanhas de desinformação, mas também observa-se a produção de conteúdos de pornografia infantil,

²⁷ CHEN, Heather; MAGRAMO, Kathleen. **Funcionário de Multinacional Paga US\$ 25 Mi a Golpista que Usou "Deepfake" para Simular Reunião.** CNN Brasil. 2024. Disponível em: <https://www.cnnbrasil.com.br/internacional/funcionario-de-multinacional-paga-us-25-mi-a-golpista-que-usou-de-epfake-para-simular-reuniao/>. Acesso em 09 jul. 2025.

²⁸ Ob. Cit. - Pág. 22.

²⁹ MACHADO, Simone. **Drauzio Varella Vendendo Colágeno? Como Deep Fakes Estão Sendo Usados para Golpes.** BBC News Brasil. 2023. Disponível em: <https://www.bbc.com/portuguese/articles/c8788pv7z7jo>. Acesso em 09 jul. 2025.

³⁰ MEAD, Esther; AGARWAL, Nitin. **Surface Web Vs Deep Web Vs Dark Web.** Encyclopedia of Big Data. Springer, Cham, 2020.

assassinato em massa ou execução assistida, o que configura graves violações aos direitos da personalidade e crimes de altíssima reprovabilidade social e jurídica. Nesse contexto, evidencia-se que o problema não reside apenas na tecnologia em si, mas se agrava significativamente conforme a profundidade e o grau de anonimato da rede em que esses conteúdos circulam.

Dessa forma, os conteúdos cuja indexação não é permitida por buscadores comuns (*surface web*), tendem a migrar para camadas mais profundas da rede, como a *deep web* e, subsequentemente, a *dark web*. Essa migração se dá, muitas vezes, como uma estratégia para dificultar sua rastreabilidade e responsabilização jurídica, uma vez que tratam-se de ambientes propícios para a veiculação de *deepfakes* em sua faceta mais nociva, dada a combinação entre o anonimato dos usuários e a baixa capacidade de fiscalização estatal. Esse deslocamento digital reflete um agravamento das condutas ilícitas associadas aos *deepfakes*, tornando ainda mais desafiadora a sua repressão e responsabilização jurídica.

Diante do exposto, verifica-se que os *deepfakes* representam uma ferramenta tecnológica de alta complexidade e poder persuasivo, capaz de reproduzir com notável fidelidade traços visuais e sonoros de indivíduos reais, bem como simular situações completamente fictícias. Sua utilização em contextos fraudulentos, comerciais, financeiros e reputacionais evidencia os inúmeros riscos associados à sua circulação, tanto na superfície da internet quanto em suas camadas mais profundas e menos reguladas. Nesse sentido, a fim de compreender o seu impacto social como um todo, é indubitavelmente relevante compreender, sob uma perspectiva prática, as externalidades negativas decorrentes do uso indevido dos *deepfakes* para a nação, a sociedade e os indivíduos.

2.3. Os Impactos Negativos da Proliferação dos Materiais de *Deepfakes* na Nação, na Sociedade e nos Indivíduos.

A partir de uma estrutura dedutiva, o impacto decorrente do uso indevido e da proliferação de *deepfakes* pode ser sistematicamente analisado em três esferas distintas e complementares. Em primeiro lugar, destaca-se a esfera estatal, representativa dos interesses da nação, cuja vulnerabilidade diante da manipulação de informações compromete a credibilidade das instituições públicas, interfere no processo democrático e ameaça a segurança nacional. Em segundo lugar, observa-se a esfera social, onde *deepfakes* intensificam a desinformação, corroem a confiança coletiva na mídia e nos discursos públicos

e fomentam conflitos, agravando o já delicado tecido social. Por fim, há a esfera individual, diretamente relacionada à proteção dos direitos fundamentais do ser humano.

A priori, vale salientar que a tecnologia *deepfake* representa ameaças concretas à estabilidade e à segurança da própria nação, em uma esfera estatal. Ao possibilitar a criação de vídeos falsos altamente realistas, os *deepfakes* podem ser utilizados para manipular a opinião pública, disseminar desinformação em larga escala, comprometer a credibilidade de autoridades e instituições, e até mesmo influenciar processos eleitorais. Em contextos de polarização política ou crises sociais, esses conteúdos falsificados têm o potencial de agravar tensões, gerar instabilidade institucional e enfraquecer a confiança da população no Estado de Direito, colocando em risco os pilares democráticos do país.

A obra "Democracia em Xequê: Inteligência Artificial e Deep Fake" aduz que o Estado Democrático de Direito, portanto, acaba vendo por fragilizado os seus princípios basilares, sendo que, dentre esses, destaca-se a *accountability*, isto é, o dever de prestação de contas dos agentes públicos à sociedade, em consonância com o postulado constitucional de que *“todo o poder emana do povo, que o exerce por meio de representantes eleitos”* (art. 1º, parágrafo único, da CRFB/88). Ou seja, tal dever implica a obrigação do Estado de atuar com transparência, de modo a viabilizar o controle social e o acompanhamento da gestão pública pelos cidadãos, especialmente quanto à administração de recursos e à formulação de políticas públicas.

Nesse contexto, o texto alude que a confiança nas informações veiculadas por canais oficiais (tais como pronunciamentos de autoridades, comunicados institucionais e plataformas governamentais) revela-se imprescindível para a consolidação da legitimidade democrática. Afinal, o exercício pleno da cidadania pressupõe o acesso a informações verídicas, que permitam ao eleitorado tomar decisões políticas de maneira consciente e informada. Todavia, os *deepfakes* comprometem gravemente esse cenário, ao viabilizar a disseminação de conteúdos falsificados de alta complexidade e verossimilhança.

Assim, sustenta que neste caso tal tecnologia trata-se de uma forma sofisticada de desinformação, que não apenas falseia a realidade, mas simula com precisão falas, comportamentos e declarações de figuras públicas, podendo ser instrumentalizada para manipular conteúdos políticos, forjando uma aparência de autenticidade que induz o receptor ao erro, o que desencadeia na saturação do ambiente informacional com conteúdos não verificáveis e enfraquece a credibilidade das comunicações institucionais, promovendo um ceticismo generalizado. Com isso, não apenas se dificulta a identificação do que é verdadeiro,

mas também se abala a própria noção de verdade, o que compromete diretamente o exercício da cidadania e a estabilidade do regime democrático.

Ademais, o mesmo texto também disciplina que outro pilar fundamental do Estado democrático de direito consiste na realização de eleições livres e periódicas, sendo inerente ao processo democrático a incerteza quanto ao resultado do pleito. Contudo, o uso indevido da inteligência artificial, em especial por meio de tecnologias de manipulação audiovisual como *deepfakes*, representa uma ameaça concreta à integridade do processo eleitoral. Isso porque a inteligência artificial pode disseminar conteúdos falsos, aparentemente verídicos, sobre resultados eleitorais, o que, por sua vez, pode ser chave fundamental para alterar todo o rumo de uma determinada eleição.³¹

Mais grave ainda é a possibilidade de que tais ferramentas tecnológicas sejam utilizadas para criar narrativas artificiais por meio da manipulação da imagem de figuras políticas, o que pode impactar diretamente os resultados das urnas. A título ilustrativo, de acordo com matéria da revista The Journalist Resource destaca-se o episódio ocorrido em fevereiro de 2024, no estado norte-americano de New Hampshire, no qual diversos eleitores receberam uma chamada telefônica automatizada (*robocall*) supostamente realizada pelo ex-presidente dos Estados Unidos, Joe Biden.

Na gravação, cuja voz apresentava notável semelhança com a do ex-mandatário, era veiculada a seguinte mensagem: “*guarde o seu voto para a eleição de novembro*”, sugerindo aos eleitores que não participassem das eleições primárias presidenciais daquele estado, sob o falso argumento de que tal participação poderia impedir o exercício do voto nas eleições gerais subsequentes. Diante da gravidade do caso, a Federal Communications Commission (FCC) - agência reguladora das comunicações nos Estados Unidos - decidiu, em caráter normativo, tornar ilegal o uso de vozes geradas por inteligência artificial em chamadas automáticas (*robocalls*), reconhecendo o potencial lesivo desse tipo de manipulação para o ambiente eleitoral.³²

Ainda, os *deepfakes* já demonstraram capacidade concreta de impactar processos eleitorais em diversas partes do mundo, levantando sérias preocupações quanto à sua

³¹ CARVALHO JÚNIOR, O. L. de .; CARVALHO, S. S.; SOUSA, B. A. de. **Democracia em Xequê: Inteligência Artificial e Deep Fake**. 7º Volume. São Paulo: Revista JRG de Estudos Acadêmicos, 2024. Disponível em: <https://revistajrg.com/index.php/jrg/article/view/1743>. Acesso em: 26 jul. 2025.

³² MIRZA, R. **How AI Deepfakes Threaten the 2024 Elections**. 2º Volume. The Journalist Resource, 2024. Disponível em: <https://journalistsresource.org/home/how-ai-deepfakes-threaten-the-2024-elections/>. Acesso em: 26 jul. 2025.

influência sobre a legitimidade democrática. A mesma matéria da revista *The Journalist Resource* alude que em recentes eleições ocorridas na Eslováquia, por exemplo, circularam nas redes sociais gravações comprometedoras geradas por inteligência artificial, simulando a voz de um candidato liberal, o qual, supostamente, estaria declarando afirmações que dizem respeito ao aumento nos preços de bebidas alcoólicas e ainda, sobre a sua intenção de fraudar o pleito. De forma semelhante, durante as eleições presidenciais na Nigéria, em fevereiro de 2023, um áudio manipulado por IA implicava falsamente um dos candidatos em um plano de adulteração de votos.³³ Nesse contexto, os *deepfakes* não apenas colocam em risco a lisura das eleições, mas também ameaçam a estabilidade institucional, a coesão social e a confiança da população nas instituições democráticas, abalando, assim, a nação e a estrutura do Estado Democrático de Direito.

Ademais, a disseminação de *deepfakes* representa uma ameaça concreta à própria estrutura social. A possibilidade de se criar conteúdos audiovisuais altamente verossímeis - mas inteiramente falsos - compromete a percepção coletiva de realidade, elemento essencial para o convívio democrático e o funcionamento das instituições.³⁴ A tecnologia *deepfake* pode ser instrumentalizada para disseminar desinformação em larga escala, fomentar conflitos sociais e instigar distúrbios, contribuindo para a erosão da confiança nas fontes oficiais de informação e na própria ideia de verdade factual.

A título exemplificativo, destaca-se um caso ocorrido nos Estados Unidos, no qual a utilização de tecnologia *deepfake* resultou em sérias consequências reputacionais e institucionais. Uma gravação falsificada, que atribuía ao diretor da escola Pikesville, Sr. Eric Eiswert, a emissão de declarações racistas e antissemitas, levou à sua colocação em licença administrativa enquanto se procedia à apuração dos fatos. Durante o período de investigação, o diretor recebeu diversas ameaças de violência, inclusive uma mensagem afirmando que “*o mundo seria um lugar melhor se você estivesse do outro lado da terra (enterrado)*”, para além da sua difamação social, conforme relatado pelas autoridades policiais.

Desde o início, Eiswert negou a autoria das falas, sustentando que a voz constante no áudio não era sua e que, possivelmente, tratava-se de um conteúdo manipulado por meio de

³³ Ob. Cit. - Pág. 28.

³⁴ JOANGUETE, Celestino. **O poder das mentiras digitais: a ameaça das deepfakes para a sociedade**. Editora Dialética, 2023. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=dtrPEAAAQBAJ&oi=fnd&pg=PT4&dq=impacto+social+deepfake&ots=1TefYZ7zew&sig=TwpmY9YOifhK1fHxTbtXjBJpPQ&redir_esc=y#v=onepage&q=impacto%20social%20deepfake&f=false. Acesso em 26 de jul. de 2025.

inteligência artificial. Por sua vez, a polícia de Baltimore confirmou que a gravação era, de fato, falsa. O responsável pela falsificação foi um funcionário da própria instituição de ensino que anteriormente teria se desentendido com o diretor, ele foi preso, sendo acusado, entre outros delitos, de perturbação do funcionamento regular da escola. Com isso, afirmou o chefe executivo do condado de Baltimore, Johnny Olszewski: *"Hoje nos sentimos aliviados por termos esclarecido um pouco a origem deste áudio (...) No entanto, está claro que também estamos entrando em uma nova fronteira profundamente preocupante"*.³⁵

Além disso, um outro exemplo foi o incidente envolvendo uma das maiores empresas de publicidade do mundo, o grupo WPP. O CEO da companhia, Mark Read, foi alvo de uma sofisticada tentativa de fraude conduzida por meio de *deepfake* vocal. A tentativa foi formalmente comunicada por Read em mensagem enviada à liderança da organização, na qual reforçou o alerta quanto ao aumento de ataques baseados em engenharia social com uso de tecnologias avançadas, recomendando especial cautela com comunicações que aparentam ser oriundas da alta administração.

Conforme detalhado no comunicado, os fraudadores criaram uma conta no aplicativo *whatsapp* utilizando uma imagem pública do executivo e agendaram, em nome dele, uma reunião virtual, via Microsoft Teams, com outro alto dirigente do grupo. Durante o encontro, os criminosos utilizaram um clone de voz (gerado com base em conteúdos disponíveis em plataformas como o youtube) e interagiram por meio do chat da reunião, simulando a presença do CEO fora das câmeras. O objetivo era induzir o interlocutor, um líder de agência, a abrir um novo negócio, buscando extrair informações confidenciais e recursos financeiros.

Apesar da complexidade do golpe, a tentativa foi prontamente identificada e frustrada, sem prejuízos à companhia. Em seu comunicado interno, Read enfatizou: *"Todos precisamos estar atentos às técnicas que vão além dos e-mails, explorando reuniões virtuais, inteligência artificial e deepfakes"*. A WPP informou ainda que tem enfrentado a proliferação de sites fraudulentos utilizando indevidamente sua identidade corporativa e que está colaborando com as autoridades competentes para mitigar riscos e coibir práticas criminosas desse tipo.³⁶

³⁵ LAKE, Thomas. **Gravação Falsa com Deepfake gera Ameaças a Diretor de Colégio nos EUA**. CNN. 2024. Disponível em: <https://www.cnnbrasil.com.br/internacional/gravacao-falsa-com-deepfake-gera-ameacas-a-diretor-de-colegio-no-s-eua/>. Acesso em 26 jul. 2025.

³⁶ ROBINS-EARLY, Nick. **CEO da Maior Empresa de Publicidade do Mundo é Alvo de Golpe de Deepfake**. The Guardian. 2024. Disponível: <https://www.theguardian.com/technology/article/2024/may/10/ceo-wpp-deepfake-scam>. Acesso em 26 jul. 2025.

Nesse contexto, observa-se que, em última instância, é a própria esfera individual de cada um que se vê comprometida diante da crescente disseminação de *deepfakes*, especialmente quando tais tecnologias violam direitos fundamentais consagrados no ordenamento jurídico brasileiro. Destacam-se, nesse sentido, os direitos à intimidade, à vida privada, à honra e à imagem, expressamente tutelados pelo artigo 5º, inciso X, da CRFB/88, cuja salvaguarda constitui pilar essencial da dignidade da pessoa humana.

Inclusive, um estudo conduzido pelo Oxford Internet Institute (OII), vinculado à Universidade de Oxford, no Reino Unido, revelou um aumento expressivo na criação de imagens manipuladas por inteligência artificial (IA), notadamente *deepfakes* de conteúdo íntimo não consensual, majoritariamente direcionados contra mulheres. A pesquisa identificou a existência de aproximadamente 35 mil softwares acessíveis ao público para download, aptos a gerar imagens falsas com base em apenas 20 fotografias de referência, em um intervalo inferior a 15 minutos.

Segundo a análise publicada, cerca de 96% dessas ferramentas são utilizadas para a produção de imagens falsas envolvendo mulheres (tanto figuras públicas quanto indivíduos anônimos) com maior incidência em países como China, Coreia do Sul, Japão, Reino Unido e Estados Unidos. Esses conteúdos são frequentemente classificados com marcadores como “pornô”, “sexy” ou “nu”, configurando, em muitos casos, violação à privacidade, exposição indevida da imagem e prática de violência de gênero digital.

Assim, o principal autor do estudo, Will Hawkins, concluiu: *“Existe uma necessidade urgente de proteções técnicas mais robustas, políticas de plataforma mais claras e aplicadas de forma mais proativa, além de novas abordagens regulatórias para lidar com a criação e a distribuição desses modelos de IA prejudiciais”*³⁷. Além dos dados quantitativos alarmantes, um outro estudo revelou impactos concretos que tais produções tiveram sobre diferentes vítimas, como no caso de Heejin, uma estudante universitária sul-coreana.

Heejin recebeu uma mensagem via Telegram e ao abri-la se deparou com uma fotografia sua (tirada quando ainda cursava o ensino médio) seguida de uma imagem manipulada com conteúdo sexualmente explícito. Ao se deparar com tais imagens, ficou apavorada e não enviou qualquer resposta, mas continuou recebendo outros conteúdos. No entanto, constatou-se que este não foi um caso isolado. Paralelamente, a jornalista

³⁷ CHRIST, Giovana. **IA Aumenta a Criação de Deepfakes de Pornografia com Mulheres, diz Estudo**. CNN Brasil. 2025. Disponível em: <https://www.cnnbrasil.com.br/tecnologia/ia-aumenta-a-criacao-de-deepfakes-de-pornografia-com-mulheres-diz-estudo/>. Acesso em 26 jul. 2025.

sul-coreana Ko Narin havia publicado uma investigação de grande repercussão, revelando que a polícia local estava conduzindo inquéritos sobre redes criminosas de pornografia *deepfake* em funcionamento dentro de duas das principais universidades do país.

Convencida de que o problema era mais disseminado, Ko aprofundou sua apuração em redes sociais e identificou dezenas de grupos ativos na plataforma de mensagens Telegram, nos quais usuários compartilhavam imagens de mulheres que conheciam pessoalmente e solicitavam, em tempo real, que fossem transformadas em conteúdo pornográfico falso. Segundo a jornalista, esses grupos atuavam de maneira sistemática e contínua: *“A cada minuto, as pessoas estavam enviando fotos de garotas que conheciam e pedindo que fossem transformadas em deepfakes”*.³⁸

O fenômeno não se limitava ao ambiente universitário. Foram identificadas salas de bate-papo com foco específico em escolas de ensino médio e até mesmo do ensino fundamental. Um dos grupos mais perturbadores contava com mais de 2 mil membros e era dedicado à disseminação de imagens falsas de alunas menores de idade. A publicação da reportagem provocou mobilização de ativistas e organizações de defesa dos direitos das mulheres, que passaram a monitorar a atividade desses grupos de forma autônoma.

Até o final daquela semana, mais de 500 instituições de ensino (entre escolas e universidades) haviam sido identificadas como alvos dessas redes. Embora o número exato de vítimas não tenha sido determinado, há indícios de que muitas são menores de 16 anos, o que configura, sob a legislação sul-coreana, violação direta à idade mínima de consentimento legal. A arquitetura do próprio aplicativo Telegram tem contribuído para o agravamento do problema: os usuários operam de forma anônima, as salas de bate-papo podem ser configuradas como “chats secretos” e o conteúdo é frequentemente apagado sem deixar registros, o que torna a coleta de provas e a responsabilização penal mais difíceis.

O Centro de Defesa das Vítimas de Abuso Sexual Online da Coreia do Sul (ACOSAV) já havia detectado um aumento substancial de casos envolvendo vítimas menores de idade antes mesmo da repercussão pública do escândalo. Além de prestar apoio psicológico e jurídico às vítimas, o centro atua na detecção e notificação de conteúdos ilícitos, trabalhando em cooperação com plataformas digitais para viabilizar sua remoção. Representantes do centro confirmaram que, embora não seja comum, houve situações em que

³⁸ MACKENZIE, Jean; LEEHYUN, Choi. **Pornografia Deepfake se Espalha por Escolas da Coreia do Sul e Vítimas são Jovens Estudantes.** BBC News Brasil. 2024. Disponível em: <https://www.bbc.com/portuguese/articles/crlrwrk929po>. Acesso em 26 jul. 2025.

o Telegram atendeu aos pedidos de exclusão de conteúdo. A própria plataforma, em nota à imprensa, declarou que utiliza ferramentas de inteligência artificial, moderação ativa em canais públicos e sistemas de denúncia para remover milhões de conteúdos que violam seus termos de uso diariamente.³⁹

Posto isso, é notável os inúmeros impactos negativos decorrentes da proliferação de conteúdos manipulados por *deepfakes*, que afetam desde a nação em sua totalidade, até a dinâmica em sociedade e a individualidade de cada ser humano. Assim, ante uma perspectiva jurídica plural, vale ainda considerar se existiriam impactos positivos do desenvolvimento da tecnologia *deepfake*, e como o uso de tal tecnologia poderia gerar benefícios à sociedade quando utilizada com responsabilidade.

2.4. Impactos Positivos do desenvolvimento das Tecnologias de *Deepfake*.

A priori, vale salientar que a tecnologia de *deepfake*, por si só, não constitui um problema. Trata-se de uma ferramenta avançada de inteligência artificial capaz de gerar conteúdos audiovisuais sintéticos com alto grau de realismo. No entanto, assim como ocorre com diversas inovações tecnológicas, seu uso inadequado por indivíduos mal-intencionados é o que tem gerado preocupações legítimas, especialmente quando empregada na criação de vídeos falsos com conteúdos maliciosos. É fundamental, portanto, distinguir a tecnologia de seu uso indevido. Diante desse cenário, surge uma questão relevante: afinal, haveria aspectos positivos no uso dos *deepfakes*? Até o momento, foi possível observar alguns casos interessantes que contemplam o uso da tecnologia *deepfake* para fins não maliciosos.

No campo das artes, destaca-se a iniciativa do Dalí Museum, localizado em Saint Petersburg, Estados Unidos, que em 2019 apresentou a exposição “*Dalí Lives (via Artificial Intelligence)*”. Por meio de uma experiência conduzida com o uso de inteligência artificial, o artista espanhol Salvador Dalí foi recriado digitalmente, em tamanho real, para interagir com os visitantes. A reprodução *deepfake* recepcionava o público, narrava passagens de sua vida e explicava detalhes sobre as obras expostas. Os visitantes ainda podiam tirar selfies com a representação virtual do artista. A iniciativa é considerada uma das primeiras a empregar a tecnologia *deepfake* com propósitos simultaneamente artísticos, institucionais e educativos.⁴⁰

³⁹ Ob. Cit. - Pág. 32.

⁴⁰ DALÍ MUSEUM. **Behind the scenes: Dalí Lives, St. Petersburg.** The Dalí. 2019. Disponível em: thedali.org/exhibit/dali-lives. Acesso em 30 jun. 2025.

Além disso, a tecnologia *deepfake* tem demonstrado potencial significativo na promoção da acessibilidade, inclusive, no cuidado com pessoas que perderam a capacidade de se comunicar verbalmente, seja por doenças ou acidentes. Um exemplo emblemático é o publicado no World Intellectual Property Organization (WIPO), o qual trata do caso do ator Val Kilmer, que perdeu a voz em decorrência de um câncer na garganta diagnosticado em 2015. Com o apoio da empresa Sonantic, especializada em inteligência artificial aplicada à síntese de som e voz, foi possível recriar digitalmente sua voz original. A tecnologia permitiu que Kilmer “voltasse a falar”, utilizando recursos de *voice cloning* baseados em registros vocais anteriores.

Além disso, a WIPO ainda alude que no cenário global, a tecnologia *deepfake* também se apresenta como um facilitador que acompanha o desenvolvimento da globalização, contribuindo para a superação de barreiras linguísticas e culturais. Um exemplo marcante ocorreu na campanha internacional de combate à malária, na qual o ex-jogador David Beckham participou de um vídeo em que, com o auxílio da tecnologia *deepfake*, foi possível transmitir sua mensagem em nove idiomas diferentes. A iniciativa ampliou significativamente o alcance da comunicação, permitindo que públicos diversos fossem impactados por uma única peça audiovisual personalizada - algo que reforça o potencial da tecnologia como ferramenta estratégica em campanhas globais de conscientização e engajamento.⁴¹

No setor da moda, a tecnologia *deepfake*, aliada à inteligência artificial, tem impulsionado soluções de hiper personalização que transformam a experiência de compra online. Por meio da criação de avatares personalizados, os consumidores podem verificar com maior precisão quais roupas lhes servem ou combinam, mesmo sem experimentá-las fisicamente. Diversas empresas já oferecem provadores virtuais que permitem o escaneamento corporal e a simulação digital das peças, tornando o processo de decisão mais interativo e assertivo. Um exemplo é a empresa japonesa de inteligência artificial DataGrid, que permite aos clientes inserir seus próprios rostos em modelos virtuais por meio de *deepfakes*, proporcionando uma visualização realista e personalizada dos produtos.⁴²

⁴¹ LALLA, Vejay; MITRANI, Adine; HARNED, Zach; **Inteligência Artificial: os Deepfakes na Indústria do Entretenimento**. World Intellectual Property Organization (WIPO). 2022. Disponível em: <https://www.wipo.int/pt/web/wipo-magazine/articles/artificial-intelligence-deepfakes-in-the-entertainment-industry-42620>. Acesso em 16 ago. 2025.

⁴² CHOW, Millie. **What are the Positive Applications of Deepfake?** Jumpstart Magazine. Retrieved. 2022. Disponível em: <https://www.jumpstartmag.com/what-are-the-positive-applications-of-deepfakes/>. Acesso em 16 ago. 2025.

Na área da educação, a tecnologia *deepfake* também tem sido explorada como uma ferramenta inovadora para tornar o processo de aprendizagem mais dinâmico e atrativo. Um exemplo é a Udacity, organização educacional norte-americana que iniciou a integração dessa tecnologia em seus cursos online. Tradicionalmente, os alunos tinham acesso apenas a slides com conteúdo textual ou a aulas em formato de áudio. Contudo, a Udacity vem testando uma nova estrutura baseada em aprendizado de máquina que permite a geração automática de vídeos de aula a partir de textos ou narrações. Essa solução proporciona uma experiência de ensino mais interativa e envolvente, superando o modelo passivo de leitura de slides e promovendo maior engajamento dos alunos com os conteúdos apresentados.⁴³

Os *deepfakes*, portanto, podem representar uma ferramenta valiosa para a sociedade, seja no campo artístico, cultural ou científico, como também em qualquer iniciativa voltada ao seu aprimoramento. No entanto, para que seu uso seja efetivamente benéfico, é fundamental que esteja acompanhado de ações educativas e afirmativas direcionadas tanto à comunidade quanto ao setor empresarial.⁴⁴

Em um primeiro momento, esses impactos positivos podem parecer pouco relevantes, em comparação aos malefícios do uso pejorativo da tecnologia. No entanto, considerando a constante evolução da internet e o avanço acelerado das tecnologias digitais, é inegável que uma ferramenta de alta resolução como a *deepfake* possui potencial para gerar benefícios ainda mais amplos. O direito, nesse contexto, assume um papel fundamental ao buscar assegurar que ferramentas tecnológicas como as *deepfakes* sejam utilizadas de forma regulamentada em sociedade. Diante dos riscos associados ao uso indiscriminado dessa tecnologia, torna-se essencial analisar quais seriam os caminhos jurídicos e regulatórios possíveis para disciplinar o uso das *deepfakes*, equilibrando inovação tecnológica e segurança jurídica.

⁴³ Ob. Cit - Pág. 34.

⁴⁴ ASSIS JR., Fabio de Paula; HESSEL, Ana Maria Di Grado. **Entre Ver e Crer: Deepfake e Criação para Arte e Entretenimento**. Revista Digital de Tecnologias Cognitivas. 2021. Disponível em: <https://revistas.pucsp.br/index.php/teccogs/article/view/55980>. Acesso em 16 ago. 2025.

3. Dos Caminhos para Regular o Uso de *Deepfakes*

3.1. A Necessidade de Regulação: Análise da Violação dos Direitos Fundamentais à Intimidade, Vida Privada, Honra e Imagem pelas *Deepfakes*

A tecnologia de *deepfake*, mesmo que apresente aspectos positivos, ao permitir a manipulação hiper-realista de imagens, vídeos e áudios, possibilita a elaboração de diversos conteúdos maliciosos - com base na imagem de terceiros - que circulam intensamente nas redes, o que desafia diretamente a proteção de direitos fundamentais, sendo que, dentre eles, destacam-se os direitos à intimidade, à vida privada, à honra e à imagem das pessoas, previstos no inciso X do artigo 5º da CRFB/88.

No que tange os direitos à intimidade e à vida privada, é importante destacar que os *deepfakes* possuem o potencial de divulgar, em larga escala e por meio da internet, conteúdos extremamente íntimos envolvendo determinada pessoa. Ainda que tais materiais sejam falsos, a ausência de consentimento e a ampla disseminação desses conteúdos resultam em exposição indevida e em constrangimento profundo à vítima, violando gravemente sua esfera privada e dignidade. A título exemplificativo, citam-se os casos em que os rostos de indivíduos são inseridos artificialmente em conteúdos de conotação sexual.

Assim, como exposto por Tércio Sampaio Ferraz, não há um conceito absoluto de intimidade, mas ele inclui, em suas diferentes esferas, o “direito de estar só”, “direito a ser deixado em paz”, “direito à liberdade de fazer e de não fazer”.⁴⁵ Quando uma pessoa é vítima da manipulação de um *deepfake*, acaba sendo artificialmente inserida em contextos que simulam situações de intimidade, expondo-a perante a sociedade de forma indevida e ofensiva à sua dignidade, mesmo que tais eventos jamais tenham ocorrido na realidade. É desrespeitado o seu direito à não expor algo que nem sequer chegou a ser concretizado no mundo factual.

Por sua vez, ao atribuírem falas ou comportamentos falsos a determinada pessoa, os *deepfakes* podem violar diretamente a honra do indivíduo, tanto no que diz respeito a sua auto imagem quanto em relação a sua reputação perante terceiros. Trata-se de uma afronta à integridade moral, com potencial significativo de comprometer a sua vida em diversas esferas, o que inclui mas não se limita ao abalo de sua imagem pública ou ainda os prejuízos que podem vir a ser desencadeados em sua vida pessoal e profissional.

⁴⁵ FERRAZ, Tércio Sampaio. **Sigilo de dados: Direito à Privacidade e os Limites à Função Fiscalizadora do Estado**. 88º Volume. Revista da Faculdade de Direito da USP. 1999.

Dessa forma, o impacto gerado por esse tipo de violação torna-se imensurável, uma vez que a honra constitui um elemento fundamental de validação social. Trata-se de um bem imaterial cuja preservação ou abalo interfere diretamente no modo como o indivíduo é percebido, podendo resultar tanto em prestígio e reconhecimento quanto em humilhação, marginalização e exclusão do convívio social e profissional.⁴⁶

Por sua vez, o uso não autorizado da imagem de uma pessoa em montagens produzidas por meio de inteligência artificial configura violação ao seu direito personalíssimo de dispor sobre a própria representação visual. Ainda que a imagem tenha sido obtida em ambiente público, sua manipulação e inserção em contextos diversos, especialmente aqueles de cunho ofensivo ou pejorativo, sem o devido consentimento, caracteriza uso abusivo e ilícito, afrontando diretamente a dignidade e a autonomia do indivíduo.

O direito à imagem assegura ao indivíduo o poder exclusivo de dispor sobre a captação, utilização e reprodução de sua própria representação visual, com ou sem finalidade comercial. Ainda que a internet se constitua como um espaço de liberdade de expressão, a proteção da imagem permanece como um direito inviolável e indispensável.⁴⁷ É certo que existem exceções, como nos casos em que o próprio titular opta por comercializar ou divulgar amplamente sua imagem, especialmente em se tratando de figuras públicas, cuja exposição tende a ser mais ampla, no entanto, a manipulação de conteúdos mediante inteligência artificial para a disseminação de informações falsas não se enquadra entre tais exceções, configurando, ao contrário, grave violação aos direitos da personalidade.

Ainda, a dignidade da pessoa humana (cláusula pétrea e núcleo axiológico dos direitos fundamentais) é reiteradamente violada quando indivíduos são transformados em objetos de manipulação digital, reduzidos à condição de simulacro, muitas vezes com intenções abusivas, ofensivas ou fraudulentas. Assim, é observada uma afronta direta ao que seria a dignidade para Immanuel Kant, o qual a enxergava como uma forma de tratar cada indivíduo como um fim em si mesmo, e não como um meio para atender as necessidades de terceiros.

Posto isso, conclui-se que o uso malicioso de *deepfakes* vem sendo amplamente utilizado como instrumento de violação de direitos fundamentais, em diferentes frentes, além de representar sérias ameaças à sociedade e à própria estabilidade do Estado democrático de

⁴⁶ TEIXEIRA, Carla Costa. **O Preço da Honra**. Universidade de Brasília Departamento de Antropologia, 1999.

⁴⁷ DE TEFFÉ, Chiara Spadaccini. **Considerações sobre a proteção do direito à imagem na internet**. 54º Volume. Revista de Informação Legislativa. 2017.

direito, sendo necessário a sua respectiva regulamentação, a qual, pode ser considerada, a partir de uma análise jurídica internacional sobre a responsabilização dos usuários pela criação e propagação indevida de *deepfakes*.

3.2. Uma Análise Jurídica Internacional sobre a Responsabilização de Usuários pela Criação e Propagação Indevida de *Deepfakes*

A responsabilização dos usuários pela criação e propagação indevida dos *deepfakes* surge como uma resposta necessária à prevenção e à reparação dos riscos inerentes à utilização dessa tecnologia. Entre os danos frequentemente associados ao uso indevido de *deepfakes*, destacam-se o assédio cibernético, a extorsão digital e a prática de fraudes eletrônicas, bem como o roubo de identidade online e a violação de direitos de propriedade intelectual. Ademais, observa-se a crescente preocupação com a disseminação não consensual de conteúdos pornográficos, a falsificação ou manipulação de provas eletrônicas, e a utilização de mídias falsas para a propagação de desinformação.⁴⁸ Assim, os ordenamentos jurídicos internacionais vêm se desenvolvendo de forma heterogênea, refletindo as distintas prioridades legislativas e os diferentes contextos culturais e sociais de cada país.

Na América, mais especificamente, América do Norte, os Estados Unidos apresentam uma abordagem legislativa descentralizada em relação à regulação dos *deepfakes*, com diversos estados adotando normas específicas conforme suas prioridades locais. No estado da Virgínia, por exemplo, a legislação concentra-se especialmente na repressão à chamada *revenge porn* (pornografia de vingança) produzida por meio de tecnologia de *deepfake*, tipificando tal prática como ilegal. O *Virginia Code Annotated S.182-386.2* proíbe expressamente a disseminação de conteúdos pornográficos manipulados digitalmente, quando esses têm por finalidade retratar o rosto ou a imagem de uma pessoa real sem o seu consentimento.⁴⁹

No Texas, evidencia-se uma clara preocupação com os efeitos prejudiciais dos *deepfakes* no cenário político, especialmente no que se refere à integridade dos processos eleitorais, visto que foi promulgada a lei *TXSB751*, que estabelece que constitui infração penal a produção, publicação ou distribuição de vídeos manipulados digitalmente nos trinta

⁴⁸ RIEHLE, Cornelia. **EuroPol Report Criminal Use of Deepfake Technology**. Eucrim. 2022. Disponível em: <https://eucrim.eu/news/eupol-report-criminal-use-of-deepfake-technology/>. Acesso em 16 ago. 2025.

⁴⁹ LOOMIS, Abigail. **Deepfakes and American Law**. Davis Political Review. 2022. Disponível em: <https://www.davispoliticalreview.com/article/deepfakes-and-american-law?format=amp>. Acesso em 16 ago. 2025.

dias que antecedem uma eleição, quando realizados com o intuito de prejudicar um candidato ou de influenciar o resultado do pleito eleitoral. Diferentemente de outros estados, até a presente data, o ordenamento texano não reconhece a pornografia de vingança como uma modalidade de *deepfake* passível de enquadramento legal sob a norma vigente.⁵⁰

Por sua vez, a Califórnia está buscando uma abordagem legislativa mais abrangente no enfrentamento aos *deepfakes*, focando tanto na erradicação de conteúdos manipulados digitalmente de natureza política, quanto aqueles de origem pornográfica. Nesse contexto, destacam-se dois marcos normativos: o Projeto de Lei nº 602 e o Projeto de Lei nº 730. O Projeto de Lei nº 602 trata especificamente dos *deepfakes* de conteúdo sexual explícito, prevendo a possibilidade de ajuizamento de ação civil privada contra qualquer pessoa que, de forma intencional, crie ou divulgue esse tipo de material, ciente de que o indivíduo retratado não consentiu. Já o Projeto de Lei nº 730 regula o uso de *deepfakes* no contexto eleitoral, alterando o Código Eleitoral da Califórnia para proibir que qualquer pessoa, comitê ou entidade distribua, intencionalmente, mídia audiovisual materialmente enganosa dentro do período de 60 dias anteriores a uma eleição. A norma estabelece que tal material será considerado ilícito se for suficientemente verossímil a ponto de uma pessoa razoável acreditar que se trata de conteúdo autêntico.⁵¹

No continente africano, verifica-se que diversos países ainda não possuem uma legislação específica voltada ao fenômeno dos *deepfakes*. No entanto, algumas nações têm se valido da adaptação de normativas já existentes para enfrentar os desafios impostos por essa tecnologia. É o caso da Nigéria, cuja Lei de Crimes Cibernéticos, promulgada em 2015, contempla certos aspectos relacionados ao uso malicioso de *deepfakes*. Essa legislação criminaliza condutas fraudulentas praticadas por meio de mensagens eletrônicas enviadas ao destinatário, quando tais mensagens contenham deturpação material de fatos ou de um conjunto de fatos, e cuja confiança nelas depositada resulte em prejuízo ou dano. Além disso, a norma estabelece como infração penal a conduta de qualquer indivíduo que, utilizando rede pública de comunicação eletrônica, transmita mensagem contendo ameaça de causar dano à

⁵⁰ VARGHESE, Benson. **Deep Fakes in Texas: What are they and are they Illegal?** Versus Texas. 2021. Disponível em: <https://versustexas.com/deepfakes/>. Acesso em 16 ago. 2025.

⁵¹ SIERRA, Anthony David. **California Deepfake Law First in Country to Take Effect.** Lexology. 2020. Disponível em: <https://lexology.com/library/detail.aspx?g=4700f977-4845-417b-834d-b3c06390ee27>. Acesso em 16 ago. 2025.

propriedade ou à reputação do destinatário, de terceiros, ou mesmo à reputação de pessoa falecida.⁵²

No Egito, a Dar Al-Ifta (organização governamental e sem fins lucrativos responsável por fornecer orientações e pareceres religiosos islâmicos) posiciona-se contrária ao uso da tecnologia *deepfake*. Segundo a instituição, tal prática constitui uma forma de mentira e engano, contrariando a realidade e os ensinamentos do Profeta Maomé. Acrescenta-se que a utilização de *deepfakes* com o intuito de enganar ou prejudicar terceiros é considerada incompatível com os princípios do Islã. Embora o Egito ainda não disponha de uma legislação específica voltada à regulamentação dos *deepfakes*, certas condutas relacionadas à referida tecnologia podem ser enquadradas na Lei de Combate aos Crimes Cibernéticos e de Tecnologia da Informação, norma que se concentra na segurança cibernética e na proteção de dados pessoais, penalizando, inclusive, condutas que contrariem os valores familiares da sociedade egípcia.⁵³

A África do Sul também não possui uma legislação específica que regule a criação e a divulgação de conteúdos manipulados por meio de *deepfakes*, sendo que a responsabilização por tais condutas pode ser estabelecida com base em princípios gerais do direito ou por outras legislações, como a Lei de Crimes Cibernéticos da África do Sul nº 19, de 2020. A norma penaliza atos praticados com a intenção de fraudar ou de produzir representações falsas por meio da manipulação de dados ou programas de computador, sempre que tais condutas causem, ou tenham potencial de causar, prejuízo a terceiros. A legislação também criminaliza a divulgação de mensagens eletrônicas, por meio de serviços digitais, quando realizadas com a finalidade de incitar ou ameaçar a integridade física ou patrimonial de indivíduos ou grupos, ou de incitar à violência contra esses.⁵⁴

Na Europa, destaca-se a legislação adotada pelo Reino Unido, que trata especificamente dos *deepfakes*, com foco principal na conduta conhecida como *revenge porn*

⁵² UBA, Josephine. **Deepfakes in Nigeria: Protection and Legal Framework against Deepfake Attacks in Nigeria.** Olisa Agbakoba Legal. 2021. Disponível em: <https://www.mondaq.com/nigeria/security/1114750/deepfakes-in-nigeria-protection-and-legal-framework-against-deepfake-attacks-in-nigeria> Acesso em 17 ago. 2025.

⁵³ AHARAM ONLINE. **Egypt 's Dar Al-Ifta Prohibits Deepfake Video and Audio Clips.** Ahram Online. 2022. Disponível em: <https://english.ahram.org.eg/NewsContent/1/64/454765/Egypt%E2%80%99s-Dar-AlIfta-prohibtsdeepfake-video-and-au.aspx>. Acesso em 17 ago. 2025.

⁵⁴ MASHININI, Nomalanga. **The Impact of Deepfakes on the Right to Identity: A South African Perspective.** South African Mercantile Law Journal. 2020. Disponível em: <https://doi.org/10.47348/SAMLJ/v32/i3a5>. Acesso em 17 ago. 2025.

(pornografia de vingança). A principal norma sobre o tema é o *Online Safety Bill* (Projeto de Lei de Segurança Online). Diferentemente do que ocorria anteriormente no país - quando a punição dependia da comprovação de que o autor do conteúdo manipulado digitalmente tinha a intenção de causar sofrimento à vítima, elemento subjetivo e de difícil demonstração - a nova legislação prevê sanções mesmo na ausência dessa comprovação. Assim, ainda que não se comprove a intenção de causar dano, a simples produção de um vídeo *deepfake* enganoso já é suficiente para ensejar punição. Caso a intenção de causar sofrimento seja comprovada, a pena pode chegar a dois anos de prisão. Na hipótese de não haver essa comprovação, mas sendo demonstrada a criação de conteúdo *deepfake* enganoso, a pena prevista é de até seis meses de prisão.⁵⁵

A Austrália, na Oceania, por sua vez, não possui uma legislação específica sobre *deepfakes*. No entanto, existem algumas legislações vigentes e entendimentos jurisprudenciais que podem ser aplicadas a casos concretos, como a *Australian Defamation Act* (Lei Australiana de Difamação), a qual trata da proteção de vítimas que tiveram suas imagens digitalmente alteradas. No caso *Gilbert v. Nationwide News Pty Ltd*, por exemplo, um jornal publicou uma foto manipulada digitalmente do autor da ação, e tal imagem foi considerada difamatória pelo tribunal. Já no âmbito dos negócios, a Seção 29(1)(a) da *Australian Consumer Law* (Lei Australiana do Consumidor) proíbe declarações falsas ou enganosas, no âmbito de atividades comerciais, quanto à existência de patrocínio ou aprovação por parte de determinada pessoa em relação a bens ou serviços.⁵⁶

Nos Emirados Árabes Unidos (EAU), no Oriente Médio, Ásia, o uso malicioso de *deepfakes* tem aumentado. Em 2020, o advogado de família Byron James, radicado em Dubai, relatou que um vídeo apresentado em juízo, no qual seu cliente aparecia ameaçando violentamente a esposa, havia sido manipulado. Após perícia técnica, constatou-se que se tratava de um *deepfake*.⁵⁷ No mesmo ano, outro caso ganhou notoriedade: o uso de tecnologia de *deepfake* de voz para aplicar um golpe bancário de US\$ 35 milhões, também nos EAU.

⁵⁵ Ob. Cit - Pág. 34.

⁵⁶ WHITE KNIGHT LAWYERS. **Deepfake Technology: Current Remedies and Possible Legal Consequences.** White Knight Lawyers. 2023. Disponível em: <https://www.wkls.com.au/deepfake-technology-current-remedies-and-possible-legal-consequences/>. Acesso em 17 ago. 2025.

⁵⁷ PATRICK, Ryan. **Deepfake Audio Evidence used in U.K.** The National News. 2023. Disponível em: <https://www.thenationalnews.com/uae/courts/deepfake-audio-evidence-usedin-uk-court-to-discredit-dubai-dad-1.975764>. Acesso em 17 ago. 2025.

Como resposta, os Emirados lançaram um Guia sobre Tecnologia *Deepfake*, por meio de seu Programa Nacional de Inteligência Artificial.⁵⁸

O país elaborou este guia com o objetivo de orientar os seus cidadãos sobre a importância da proteção de seus dados pessoais e a detecção de conteúdos falsificados por meio de ferramentas baseadas em inteligência artificial. O documento classifica os conteúdos manipulados em duas categorias: *Shallow fakes* (manipulações simples), *Deepfakes* (manipulações avançadas com uso de IA). Além disso, apresenta: os problemas associados aos deepfakes; os diferentes tipos de conteúdo forjado; métodos de detecção; estratégias de proteção; e as diretrizes do Conselho de Bem-Estar Digital dos EAU. Ainda, destacam os impactos negativos desvendados pelos *deepfakes*, dentre eles: danos à reputação pessoal; manipulação da opinião pública com o intuito de causar desordem; aumento da desconfiança generalizada e produção de provas falsas para influenciar decisões judiciais.⁵⁹

Posto isso, conclui-se que o uso malicioso de *deepfakes* têm gerado crescente preocupação em âmbito global. Independentemente da existência - ou não - de uma legislação específica que regule a matéria, é imprescindível que os ordenamentos jurídicos contem com normas voltadas aos *cybercrimes*, prevendo sanções adequadas para aqueles que, no ambiente digital, criam ou disseminam conteúdos manipulados. Ainda, para além do autor da *deepfake*, é necessário avaliar os limites e deveres das plataformas digitais utilizadas para a criação e circulação desse tipo de conteúdo, de modo a garantir um ambiente virtual mais seguro e responsável.

3.3. Os Desafios da Responsabilização Jurídica das Plataformas sob uma Ótica Internacional

No plano internacional, observa-se que existem diferentes abordagens legislativas quanto à responsabilidade jurídica das plataformas que viabilizam a produção de conteúdos por meio de *deepfakes*. A forma como uma nação opta por regular - ou não - os aplicativos e sites de internet constitui, em grande medida, um reflexo claro do modelo de governança

⁵⁸ WAM. UAE Launches 'Deepfake Guide'. Khaleej Times. 2023. Disponível em: <https://www.khaleejtimes.com/tech/uae-launches-deepfake-guide?amp=1>. Acesso em 17 ago. 2025.

⁵⁹ ZAWYA. The UAE National Programme for Artificial Intelligence Launches a Guide to illustrate Uses of Deepfake Technology. Zawya. 2023. Disponível em: <https://www.zawya.com/en/press-release/the-uaenational-program-for-artificial-intelligence-launches-a-guide-to-illustrate-uses-of-deepfake-jb67ph5i>. Acesso em 17 ago. 2025.

adotado pelo seu Estado, refletindo os princípios sociais, econômicos e políticos que lhe orientam.

A China constitui um exemplo emblemático de Estado cujo controle governamental sobre o acesso à internet é exercido de forma e centralizada, marcado pela sua forte censura à liberdade de expressão individual. Assim, uma vez que trata-se de um ambiente de monitoramento sistemático das atividades individuais, tal cenário impulsiona a preocupação do governo em controlar o conteúdo que se propaga nas redes digitais, o que, inevitavelmente, favorece a formulação, implementação e fiscalização de normas específicas voltadas à regulação da concepção e circulação de *deepfakes*.⁶⁰

Assim, o país foi pioneiro ao estabelecer uma legislação abrangente e bem definida sobre *deepfakes*, por meio do órgão *Cyberspace Administration of China (CAC)* (Administração do Ciberespaço da China). A legislação entrou em vigor em 10 de janeiro de 2023, com dois objetivos centrais: (i) intensificar os mecanismos de controle e regulação do conteúdo, resultando em um aumento da censura e restrição à liberdade de expressão; e (ii) atualizar-se frente ao avanço tecnológico, buscando antecipar desafios ou, ao menos, estar preparada para enfrentá-los quando surgirem. A lei denominada *Deep Synthesis Technology Regulation* (“DSTR”) estabelece normas para disciplinar duas categorias principais: os *Deep Synthesis Service Providers* (“Provedores de Serviços de Síntese Profunda”, ou seja, plataformas utilizadas para criar *deepfakes*) e os *Deep Synthesis Service Users* (“Usuários de Serviços de Síntese Profunda”, isto é, usuários que utilizam das plataformas para criar, duplicar, publicar ou transferir informações geradas por *deepfakes*).⁶¹

De todo modo, vale destacar que, apesar de apresentarem restrições aos usuários e plataformas, as normas previstas na lei são bem mais restritivas para os *Deep Synthesis Service Providers* (“DSSP”), isto é, as plataformas. Isso se justifica pelo fato de que o governo chinês entende que uma vez que os DSSP criam o meio digital necessário para a geração dos *deepfakes*, elas têm o dever de possuir o controle sobre o que é manipulado por meio de seus serviços aos usuários, o que seria uma condição natural de sua própria

⁶⁰ HAMRAJANI, Asha. **China’s New Legislation on Deep Fakes: Should the Rest of Asia Follow Suit?** The Diplomat. 2023. Disponível em: <https://thediplomat.com/2023/03/chinas-new-legislation-on-deepfakes-should-the-rest-of-asia-follow-suit/>. Acesso em 17 ago. 2025.

⁶¹ KHARPAL, Arjun. **China is about to get Tougher on Deep Fakes in an Unprecedented Way. Here’s what the Rules Mean.** CNBC. 2023. Disponível em: <https://www.cnbc.com/2022/12/23/china-is-bringing-in-first-of-its-kind-regulation-on-deepfakes.html>. Acesso em 10 ago. 2025.

tecnologia. Com isso, eles deveriam ter uma maior responsabilidade e o dever de censurar a criação de determinados materiais. Um reflexo claro disso é a própria DSTR, a qual é composta por quatro capítulos distintos que disciplinam tais responsabilidades.

Nesta legislação,⁶² o capítulo I baseia-se em leis chinesas de cibersegurança, proteção de dados e regulamentação de serviços de informação na internet. O objetivo deste capítulo é disciplinar sobre como deve-se regular a gestão de serviços de síntese profunda (*deepfakes*), alinhando-os aos valores socialistas, à segurança nacional e ao interesse público, bem como aludindo que os serviços devem respeitar as leis, costumes e ética, mantendo orientação política adequada. Ainda, define que o Departamento Estatal de Informação da Internet será o responsável por exercer essa supervisão nacionalmente, enquanto órgãos de telecomunicações e segurança pública atuam conforme suas atribuições. Ademais, estimula que plataformas do setor estabeleçam padrões e sistemas de autorregulação, promovendo conformidade à supervisão social.

O capítulo II proíbe o uso de *deepfakes* para criar, publicar ou transmitir conteúdos ilegais, falsos ou prejudiciais à segurança, à ordem ou aos direitos de terceiros. Impõe aos provedores o dever de adotar sistemas de gestão e segurança, o que inclui a verificação da identidade real dos usuários e uma revisão constante dos conteúdos produzidos por esses. Ainda, determina que cabe às plataformas informarem aos canais de denúncia respectivos qualquer atividade inadequada praticada por determinado usuário. Outrossim, impõe às plataformas de distribuição a obrigação de verificar, previamente, a conformidade legal do conteúdo a ser disponibilizado em sua rede, antes de torná-lo acessível aos demais usuários, com dever de suspender ou remover conteúdos que se encontrem em situação irregular.

O capítulo III estabelece que os provedores e apoiadores técnicos devem adotar medidas para proteger os dados pessoais coletados por meio de seus sistemas, observando integralmente as normas aplicáveis, especialmente quando envolverem informações biométricas, cujo armazenamento exige o consentimento prévio e expresso dos titulares. Determina, ainda, que tais plataformas implementem uma gestão técnica capaz de: (i) revisar, avaliar e verificar periodicamente os mecanismos algorítmicos; e (ii) realizar avaliações de segurança - internas ou conduzidas por terceiros - para ferramentas que gerem ou editem informações biométricas ou outros conteúdos sensíveis com potencial impacto sobre a

⁶² CHINA LAW TRANSLATE. **Provisions on the Administration of Deep Synthesis Internet Information Services.** China Law Translate. 2023, Disponível em: <https://www.chinalawtranslate.com/deep-synthesis>. Acesso em 10 ago. 2025.

segurança nacional, a imagem do país ou os interesses estatais. Ademais, impõe que o conteúdo gerado pela plataforma seja acompanhado de rotulagem clara e visível indicando que foi alterado, abrangendo textos, vozes, imagens, vídeos e cenas imersivas, de modo a alertar o público quanto à sua natureza sintética, sendo vedada a exclusão, modificação ou ocultação desses rótulos.

O capítulo IV define os mecanismos de supervisão e as consequências pelo descumprimento das regras sobre serviços de *deepfake*. Em síntese, estabelece que os provedores devem realizar registro formal junto às autoridades, sendo obrigados a divulgar o número de registro e link correspondente em seus sites, aplicativos e serviços, devendo passar por avaliações de segurança conforme as normas estatais aplicáveis. Além disso, estabelece que órgãos de informação na internet, telecomunicações e segurança pública podem inspecionar os serviços das plataformas, cabendo aos provedores cooperar e fornecer apoio técnico, dados e outras informações necessárias, cabendo às autoridades banir conteúdos que sejam considerados ameaças. Em caso de haver infração à segurança pública, aplicam-se sanções administrativas; se configurado crime, haverá responsabilização penal.

Já o capítulo V define termos como “tecnologia de síntese profunda”, “provedor”, “usuário”, “apoio técnico”, “dados de treinamento” e “cenas imersivas”, listando exemplos de tecnologias reguladas, incluindo geração/edição de texto, áudio, música, imagens, vídeos, características biométricas e ambientes virtuais, os denominados *deepfakes*. Por fim, estabelece que, se os provedores atuarem também em publicações online, atividades culturais ou programação audiovisual, devem cumprir as regras específicas desses setores.

Posto isso, verifica-se que o governo chinês adota um controle rigoroso sobre os conteúdos gerados e as eventuais manifestações no ambiente digital, atribuindo às plataformas uma ampla responsabilidade pela supervisão e filtragem do material produzido por seus usuários. Tal modelo, por sua vez, contrasta significativamente com o adotado pelos Estados Unidos, que, por muitas vezes, não tem a prerrogativa de exercer tamanho controle em ambiente digital com base no princípio da liberdade de expressão.

Nos Estados Unidos, a regulação de conteúdos digitais adota um modelo descentralizado, no qual a definição e a aplicação de normas recaem predominantemente sobre os próprios provedores de serviços. Nesse contexto, conforme já mencionado, embora o governo americano tenha o dever/poder de criar leis e sanções para regulamentar as

deepfakes, estas, em geral, tendem a incidir com maior frequência sobre o usuário que cria e dissemina o conteúdo adulterado, e não sobre as plataformas que os hospedam.⁶³

O principal desafio para conter tal prática decorre da ampla proteção à liberdade de expressão garantida pela Constituição norte-americana. Um dos dispositivos mais representativos dessa proteção é a Seção 230 do *Communication Decency Act* (Lei de Decência nas Comunicações), que estabelece: “*Nenhum provedor ou usuário de um serviço interativo de computador deve ser tratado como editor ou responsável pelo conteúdo fornecido por outro provedor de conteúdo informativo.*”.

Tal disposição impede que o governo exerça regulação direta sobre sites e aplicativos, transferindo aos próprios provedores e desenvolvedores a responsabilidade de elaborar, implementar e fiscalizar suas políticas de uso, conforme suas estratégias e modelos de governança. Como consequência, os *deepfakes* acabam sendo menos reguladas, o que gera problemas que se propagam para o mundo real.⁶⁴ Por outro lado, os usuários desfrutam de maior liberdade de expressão na internet e não são monitorados de forma constante e invasiva, como ocorre sob o rígido controle exercido pelo governo chinês.

Como um meio-termo, pode-se citar o modelo adotado no Reino Unido, cuja legislação atribui responsabilidades também às plataformas, mas sem um controle excessivo pelas suas atitudes e usuários. O *Online Safety Bill*, por exemplo, estabelece obrigações específicas para provedores de serviços de *deepfake*, incluindo a verificação da identidade dos usuários para possibilitar o rastreamento de conteúdos manipulados maliciosamente até seus criadores. Além disso, impõe avaliações de risco voltadas a crianças e adultos, visando evitar que os conteúdos disponíveis facilitem a exploração ou o abuso sexual infantil, bem como atentem contra a dignidade de adultos, como em casos de *revenge porn* (pornografia de vingança). Em especial, o artigo 13(4) determina que conteúdos prejudiciais de caráter adulto sejam: (i) removidos; (ii) restritos quanto ao acesso por parte dos usuários; e (iii) limitados em termos de recomendação e promoção.⁶⁵

Com isso, conclui-se que a percepção sobre a responsabilidade das plataformas pela criação e disseminação de *deepfakes* varia conforme as políticas públicas e as estruturas

⁶³ NNAMDI, Nmesoma; ONIYINDE, Omolade Adeyemi; BABALOLA, Abegunde. **An Appraisal of the Implications of Deep Fakes: The Need for Urgent International Legislations**. 8º Volume. American Journal of Leadership and Governance. 2023. Disponível em: <https://www.researchgate.net/publication/372611418>. Acesso em 16 ago. 2025.

⁶⁴ Ob. Cit. - Pág. 45.

⁶⁵ Ob. Cit. - Pág. 45.

sociais de cada Estado, refletindo, assim, a configuração institucional e os valores sociais de cada país. Nesse sentido, para situar o presente estudo no contexto nacional, faz-se necessário um aprofundamento específico sobre a forma como o ordenamento jurídico brasileiro tem enfrentado o desafio da regulamentação dos *deepfakes*, analisando normas, projetos legislativos e decisões judiciais pertinentes.

3.4. Como o Direito Brasileiro Enfrenta o Desafio da Regulamentação dos *Deepfakes*

O Brasil ainda não dispõe de legislação específica voltada à regulamentação dos *deepfakes*, o que torna imprescindível analisar como esse cenário vem se desenvolvendo diante da ameaça que tal desregulamentação representa aos direitos fundamentais, especialmente ao direito à intimidade, à vida privada, à honra e à imagem, expressamente protegidos pelo artigo 5º, inciso X, da CRFB/88. Nesse contexto, deve-se compreender como a responsabilidade das plataformas e dos usuários vem sendo tratada diante desse desafio tecnológico.

No que concerne à responsabilidade das plataformas, até então, o Marco Civil da Internet dispõe, de forma clara, em seu artigo 18, que “*o provedor de conexão à internet não será responsabilizado civilmente por danos decorrentes de conteúdo gerado por terceiros*”.⁶⁶ A principal exceção encontra-se no artigo 21, que prevê a obrigatoriedade de retirada de conteúdo de cunho sexual ou pornográfico divulgado sem autorização, mediante simples notificação extrajudicial do usuário. Assim, “*o único critério estabelecido pelo legislador para a retirada de conteúdo é que os participantes não tenham autorizado a divulgação do vídeo*”⁶⁷. Em resumo, a responsabilidade das plataformas pela divulgação de *deepfakes* tende a ser afastada, salvo quando: (i) envolverem material pornográfico não consentido ou (ii) houver ordem judicial determinando a remoção do conteúdo, o que, por si, pode ser um elemento importante de prevenção da arbitrariedade estatal.⁶⁸

Quanto à responsabilidade dos usuários, cabe ressaltar que alguns casos emblemáticos já chegaram às instâncias superiores, incluindo o Supremo Tribunal Federal. Um exemplo é a

⁶⁶ BRASIL. Lei nº 12.965, de 23 de abril de 2014. **Estabelece Princípios, Garantias, Direitos e Deveres para o Uso da Internet no Brasil**. Diário Oficial da União, Poder Executivo, Brasília, DF, 23 abr. 2014. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm. Acesso em 10 ago. 2025.

⁶⁷ GONÇALVES, Victor Hugo Pereira. **Marco Civil da Internet Comentado**. 1ª Edição. São Paulo: Atlas, 2017.

⁶⁸ AFFONSO, Filipe José Medon. **O Direito à Imagem na Era das Deep Fakes**. 27º Volume. Revista Brasileira de Direito Civil, 2021.

Ação Penal nº 1021/DF⁶⁹, relatada pelo Ministro Luiz Fux, com revisão da Ministra Rosa Weber, julgada em 18/08/2020 e publicada em 21/10/2020 pela Primeira Turma. Na ocasião, figuraram como autor Jean Wyllys de Matos Santos e como réu Éder Maurício - caso que, embora não tratasse exclusivamente da tipificação penal dos *deepfakes*, trouxe à tona discussões relevantes sobre desinformação, manipulação de conteúdo e seus impactos jurídicos, conforme consta na ementa e é explicado abaixo:

“Ementa: PENAL E PROCESSO PENAL. AÇÃO PENAL PRIVADA. CRIME DE DIFAMAÇÃO. ART. 139 DO CÓDIGO PENAL. PRELIMINAR DE INÉPCIA DA QUEIXA-CRIME REJEITADA. IMUNIDADE PARLAMENTAR. NÃO INCIDÊNCIA. PUBLICAÇÃO DE VÍDEO EDITADO MEDIANTE CORTES, ATRIBUINDO-LHE CONTEÚDO RACISTA INEXISTENTE NA FALA ORIGINAL. COMPROVAÇÃO DA MATERIALIDADE E DA AUTORIA. CONFIGURAÇÃO DO ANIMUS DIFFAMANDI. AÇÃO PENAL JULGADA PROCEDENTE. 1. PRELIMINAR (a) A inépcia da inicial acusatória, devidamente afastada pelo órgão julgador no recebimento da queixa-crime, é matéria preclusa. (b) In casu, constou do acórdão de recebimento da queixa-crime: “Da análise do vídeo em questão, é possível, a princípio, determinar o fato objetivamente imputado, não sendo este o momento adequado para se tecer maiores considerações sobre o mérito da controvérsia. Preenchidos, desse modo, os requisitos do art. 41 do Código de Processo Penal”. (c) Preliminar rejeitada. 2. PREJUDICIAL DE MÉRITO (a) A imunidade parlamentar teve sua incidência afastada no caso ora em julgamento, por ocasião do recebimento da exordial acusatória. (b) A imunidade parlamentar exige, para sua incidência, que o ato incriminado tenha sido praticado in officio ou propter officio. Os atos delituosos praticados fora do recinto do parlamento e desvinculados do exercício da função não se encontram ao abrigo da imunidade material. Precedentes (Inq. 3932 e Pet 5243, Primeira Turma, Rel. Min. Luiz Fux, DJe 9/9/2016; Inq. 3438, Primeira Turma, Rel. Min. Rosa Weber, DJe 10/2/2015; Inq. 3672, Primeira Turma, Rel. Min. Rosa Weber, DJe 21/11/2014; RE 299.109-AgR, Primeira Turma, Rel. Min. Luiz Fux, DJe 1º/6/2011). (c) A veiculação dolosa de vídeo com conteúdo fraudulento, para fins difamatórios, conferindo ampla divulgação (rede social) a conteúdo sabidamente falso, não encontra abrigo na nobre garantia constitucional da imunidade parlamentar, insculpida no artigo 53 da Lei Maior, e que protege a liberdade e independência dos eleitos para defender suas opiniões mediante suas palavras e votos. (d) No acórdão de recebimento da inicial, restou assentado que “A liberdade de opinião e manifestação do parlamentar, ratione muneris, impõe contornos à imunidade material, nos limites estritamente necessários à defesa do mandato contra o arbítrio, à luz do princípio republicano que norteia a Constituição Federal”. (e) Prejudicial rejeitada. 3. MÉRITO (a) In casu, (a) o Deputado Federal Jean Wyllys de Matos Santos imputou ao Deputado Federal Eder Mauro a prática de crime de difamação agravada (artigos 139 c/c art. 141, II e III, do Código Penal), consistente em publicação ofensiva à honra do querelante, divulgada na página do querelado no Facebook. (b) Com efeito, consta dos autos que, durante reunião da Comissão Parlamentar de Inquérito que apura a Violência contra jovens e negros pobres no Brasil, realizada em 14 de maio de 2015, da qual participaram tanto o réu como o Autor, este último, Deputado Federal Jean Wyllys, fez uso da palavra para tecer as seguintes considerações: “E aí a fala da Tatiana foi muito importante, porque ela traz essa dimensão histórica, que envolve a escravidão de negros; depois, a abolição, sem nenhuma política de inclusão no mercado de trabalho, a exclusão territorial; e, depois, toda uma produção de sentido que desqualifica essa

⁶⁹ BRASIL. Supremo Tribunal Federal. **Recurso Extraordinário nº123.456/SP**. Relator: Ministro Luiz Fux. Julgado em: 18/08/2020. Publicado no DIário da Justiça Eletrônico em 21/10/2020. Disponível em: https://jurisprudencia.stf.jus.br/pages/search?base=acordaos&pesquisa_inteiro_teor=false&sinonimo=true&plural=true&radicais=false&buscaExata=true&page=1&pageSize=10&queryString=deepfake&sort=_score&sortBy=desc. Acesso em 10 ago. 2025.

comunidade como humana. Então, há um imaginário impregnado, sobretudo nos agentes das forças de segurança, de que uma pessoa negra e pobre é potencialmente perigosa, é mais perigosa do que uma pessoa branca de classe média. Esse é um imaginário que está impregnado na gente, uma dimensão aí. E os policiais partem desse imaginário” (pág. 37 das notas taquigráficas da CPI – Violência contra jovens negros e pobres). (c) Cinco dias depois, em 19 de maio de 2015, o réu, Deputado Federal Eder Mauro, publicou em seu perfil do Facebook um vídeo contendo trecho recortado da referida fala, previamente editado de modo a inverter seu conteúdo. No conteúdo fraudulento veiculado, o Deputado Federal Jean Wyllys aparece falando o seguinte: “Uma pessoa negra e pobre é potencialmente perigosa, é mais perigosa do que uma pessoa branca de classe média, essa é a verdade, então, dito isso...”. (d) Em síntese, o Réu é acusado de ter divulgado vídeo editado de modo a dolosamente atribuir-lhe conteúdo racista e preconceituoso, com finalidade de difamar a honra do Querelante. (e) O vídeo com trecho cortado e editado da fala do Parlamentar Autor foi publicado no Facebook e recebeu 14.834 aprovações (“curtidas”), 252.458 visualizações e 12.272 compartilhamentos. O conteúdo fraudulento somente foi excluído da página do Querelado Eder Mauro no Facebook por determinação da Justiça (decisão pública da 14ª Vara Cível de Brasília/DF, de 28 de agosto de 2015, disponível em:

http://www.omci.org.br/m/jurisprudencias/arquivos/2017/df_00209599520158070001_28082015.pdf 4.

(a) A defesa alega a veracidade do conteúdo do vídeo divulgado pelo réu, que tão-somente reproduziu trecho de debate parlamentar no âmbito de CPI da Câmara dos Deputados. (b) Nada obstante, o Laudo de Perícia Criminal 17.454/2017 (fls. 84/110) do Instituto de Criminalística da Polícia Civil do Distrito Federal, concluiu que “o vídeo questionado foi editado” e “que o processo de edição do vídeo questionado resultou na modificação da informação auditiva da fala do Deputado Jean Wyllys originalmente registrada no material padrão, conduzindo a uma compreensão diversa da realidade factual. Em outras palavras, o discurso do Deputado Jean Wyllys foi adulterado no vídeo questionado”. (c) O conteúdo original da manifestação sofreu vários cortes, após os quais passou a revelar conotação racista e preconceituosa, contrária ao seu sentido original. O fato de o vídeo veicular trechos da fala do Deputado Autor é o elemento especioso, precisamente o ardil empregado para conferir verossimilhança ao conteúdo, elemento mínimo de verdade necessário para impedir o público de duvidar da postagem e acreditar na mentira resultante da edição. (d) Depoimentos prestados em juízo certificaram o dano a honra do Autor: (d.1) “essas informações geraram um impacto substantivo e absolutamente negativo da fala do Deputado Jean Wyllys junto aos ativistas do movimento negro, aos ativistas dos movimentos sociais”; “Eu sou do Estado da Bahia, em que há uma força enorme do movimento negro, e eu, como militante, fui intensamente questionado por que não combati a fala do Deputado Jean. E eu tentava explicar que o contexto da fala não teria sido aquele produzido pelo vídeo” (depoimento do Deputado Federal Adalberto Souza Galvão); (d.2) “isso é um estrago muito profundo. Porque o deputado que tem esta bandeira, que é eleito com essa bandeira, que seu eleitorado aporta o voto, lhe oferta o voto em função desta bandeira que ele tem e isso é desse desconstruído através do vídeo” (depoimento da Deputada Federal Érica Kokay). (f) Simultaneamente, há prova nos autos do impacto sobre a imagem do Autor, como se extrai da ampla circulação conferida ao vídeo fraudulento a partir do perfil do Réu no Facebook, observando exclusivamente os dados existentes na sua página, a partir da qual houve o indevido propulsão do conteúdo falso. Ademais, a fraude revela nítido potencial de enganar os cidadãos que a visualizaram e de produzir discursos de ódio contra a fala indevidamente alterada, difamando seu opositor político. (g) Consectariamente, restou comprovada a materialidade do crime de difamação. 5. (a) A publicação em perfil de rede social é penalmente imputável ao agente que, dolosamente, tem o intuito de difamar, injuriar ou caluniar terceiros, máxime quando esteja demonstrado o conhecimento da falsidade do conteúdo. (b) Inviável desresponsabilizar autores de perfis utilizados para a disseminação dolosa de campanhas difamatórias, caluniosas ou injuriosas nas redes sociais, fundadas em conteúdos falsos. (c) É irrelevante, para fins de determinação da autoria, o anonimato do “criador do conteúdo” (editor ou programador visual, por exemplo) ou da terceirização das postagens (perfil administrado por um preposto) pelo titular do perfil utilizado para divulgar a notícia falsa. Revela-se bastante e suficiente, para fins de determinação da autoria dolosa, a demonstração do conhecimento do titular do perfil

quanto à fraude do conteúdo e sua intenção de causar danos à honra das vítimas. 6. (a) No dizer de John Stuart Mill, opiniões equivocadas devem ser protegidas, enquanto tais, pois mesmo elas contribuem, no procedimento dialógico da sua refutação, para o debate e o esclarecimento da verdade: “(...) a opinião que se tenta suprimir por meio da autoridade talvez seja verdadeira. Os que desejam suprimi-la negam, sem dúvida, a sua verdade, mas eles não são infalíveis. Não têm autoridade para decidir a questão por toda a humanidade, nem para excluir os outros das instâncias do julgamento. Negar ouvido a uma opinião porque se esteja certo de que é falsa, é presumir que a própria certeza seja o mesmo que certeza absoluta. Impor silêncio a uma discussão é sempre arrogar-se infalibilidade”. E conclui: “Se a opinião é certa, aquele foi privado da oportunidade de trocar o erro pela verdade; se errônea, perdeu o que constitui um bem de quase tanto valor - a percepção mais clara e a impressão mais viva da verdade, produzidas pela sua colisão com o erro” (John Stuart Mill, *On Liberty*, capítulo 1). (b) A liberdade de expressão no debate democrático distingue-se, indubitavelmente, da veiculação dolosa de conteúdos voltados a simplesmente alterar a verdade factual e, assim, alcançar finalidade criminosa de natureza difamatória, caluniosa ou injuriosa. (c) A alavancagem de conteúdos fraudulentos, mediante artifício ardilosamente voltado à destruição da honra de terceiros nas redes sociais, revela alto potencial lesivo, tolhendo, até mesmo, o exercício de outros direitos fundamentais das vítimas - direitos políticos, liberdade de locomoção e, no limite, integridade física e direito à vida, não revelando qualquer interesse em contribuir para ganhos na construção de uma sociedade democrática. (d) As instituições democráticas e os objetivos fundamentais da República, anunciados no preâmbulo da Constituição de 1988, dependem da compreensão compartilhada no sentido de que, na letra da nossa Lei Fundadora, “nós, o povo brasileiro, nos reunimos para instituir um Estado Democrático, destinado a assegurar o exercício dos direitos sociais e individuais, a liberdade, a segurança, o bem-estar, o desenvolvimento, a igualdade e a justiça como valores supremos de uma sociedade fraterna, pluralista e sem preconceitos, fundada na harmonia social e comprometida, na ordem interna e internacional, com a solução pacífica das controvérsias”. (e) A Constituição, no Estado Democrático de Direito, é o norte do Estado Juiz na verificação da regularidade do exercício do direito fundamental à liberdade de expressão. (f) A criminalização da veiculação de conteúdo com finalidade difamatória, caluniosa ou injuriosa não colide com o direito fundamental à liberdade de expressão, que resta protegida também nos casos de desconhecimento da manipulação fraudulenta do conteúdo, a caracterizar hipótese de erro, que exclui a ilicitude (artigo 20, §1º, do Código Penal). 7. (a) O delito contra a honra é de ação múltipla, conglobando não apenas a criação do conteúdo criminoso voltada à divulgação como também a sua postagem (“upload”, carregamento do vídeo na rede social) e a disponibilização de perfil em rede social com fim de servir de plataforma à alavancagem da injúria, calúnia ou difamação, tendo por elemento especial do tipo o dano à honra da vítima. (b) A autoria dos crimes contra a honra praticados por meio da internet demanda: (b.1) demonstração de que o réu é o titular da página, blog ou perfil pelo qual o conteúdo difamatório foi divulgado; (b.2) demonstração do consentimento, prévio, concomitante ou sucessivo, com a veiculação da publicação difamatória em seu perfil; (b.3) *animus injuriandi*, *caluniandi* ou *diffamandi*, que demandam a demonstração de que o réu tinha conhecimento do conteúdo fraudulento da postagem. 8. (a) In casu, a defesa sustenta duas teses com propósito de refutar ou gerar dúvida razoável quanto à autoria: (a.1) transfere para terceiros a responsabilidade pela edição; (a.2) transfere para terceiros a responsabilidade pela divulgação do vídeo em seu perfil no Facebook. (b) Em seu interrogatório judicial, o réu afirmou ter visualizado o conteúdo do vídeo e ter sido comunicado da respectiva publicação em seu perfil no Facebook. (c) Como fiz constar de meu voto de recebimento da Queixa-Crime, “Na lição especializada de Jacques Aumont e Michel Marie, na obra ‘Dicionário teórico e crítico de cinema’, a edição ou montagem ‘tem por objetivo guiar o espectador, permitir-lhe seguir a narrativa facilmente’ e ‘pode, também, produzir outros efeitos: efeitos sintáticos ou de pontuação, marcando, por exemplo, uma ligação ou uma disjunção; efeitos figurais, podendo, por exemplo, estabelecer uma relação de metáfora; [...] entre outros’” (AUMONT, Jacques; MARIE, Michel. *Dicionário teórico e crítico de cinema*. 2ª ed. Campinas: Papirus, 2006, p. 196). (d) Restou evidenciado o conhecimento da edição voltada à adulteração do conteúdo por parte do Réu, porquanto se tratava de manifestação absolutamente contrária à proferida pelo parlamentar Autor, em

debate do qual o próprio réu participou e cujo conteúdo era de seu inteiro conhecimento. Aliás, provou-se, no interrogatório judicial, a plena consciência do Réu de que o vídeo divulgado em seu perfil no Facebook, com centenas de milhares de visualizações, atribuía ao Autor, Jean Wyllys, ideias diametralmente opostas às que identificam a plataforma política deste parlamentar. (e) A divulgação do conteúdo fraudado, invertendo-lhe o sentido com finalidade de difamar o Autor, constitui etapa da execução do crime, estabelecendo a autoria criminosa do divulgador, a qual não exclui a do programador visual ou do editor responsável pela execução material da fraude, quando promovidas por outros agentes em coautoria. (f) A segunda linha argumentativa da defesa, que surgiu no interrogatório judicial, é a de que o vídeo foi divulgado por um ajudante no perfil do réu no Facebook. (g) A tese revela fragilidades, inabilitando-se como fonte de dúvida razoável quanto aos fatos: (g.1) a defesa não pediu a oitiva do mencionado ajudante nos autos na qualidade de testemunha e, demais disso, o réu alegou não se lembrar do sobrenome dessa pessoa, inviabilizando a confirmação da própria existência do álibi pelo juízo; (g.2) ainda que um “ajudante” houvesse, de fato, postado o vídeo fraudulento veiculador da difamação, a coautoria criminosa do titular do perfil do Facebook somente seria afastada se o réu desconhecesse o uso de sua página para a veiculação e, portanto, não tivesse consentido com o emprego de sua plataforma em rede social para alavancar campanha difamatória contra o Autor; (h) In casu, (h.1) o vídeo foi postado no perfil do acusado no Facebook; (h.2) o réu admitiu ter assistido ao vídeo; (h.3) o réu admitiu ter sido informado da postagem quando o vídeo foi disponibilizado em sua página no Facebook; (h.4) o réu sabia que o conteúdo não era fidedigno à fala do Parlamentar Autor e manteve, ainda assim, o conteúdo difamatório disponível em seu perfil no Facebook; (h.5) Conforme apontou a d. Procuradoria-Geral da República, o “vídeo só foi retirado de circulação após decisão judicial” (decisão pública da 14ª Vara Cível de Brasília/DF, de 28 de agosto de 2015, disponível em: http://www.omci.org.br/m/jurisprudencias/arquivos/2017/df_00209599520158070001_28082015.pdf); (h.6) o vídeo fraudulento elevou a popularidade do réu na rede social utilizada, revelando número de visualizações superior à média de sua página, a revelar seu ganho pessoal com a campanha difamatória. (i) Os testemunhos colhidos durante a instrução da ação penal, corroboram a autoria criminosa, destacando-se os seguintes trechos de depoimentos prestados em juízo: (i.1) “Ah, sim, o vídeo que ele fez, porque ele dizia o seguinte: Mas ele falou isso, eu não falei nada, eu não divulguei nada que ele não tenha falado” (Deputada Federal Érika Kokay); (ii.2) “Com os debates, no âmbito da própria CPI, chegou-se... não posso afirmar, porque não vi ele produzindo a alteração, mas todas as informações levaram a um juízo de valor de que a autoria teria sido do próprio Deputado Éder Mauro” (Deputado Federal Adalberto Souza Galvão); (iii.3) “o primeiro pronunciamento dele sobre essa publicação era, justamente, reforçando a publicação, ou seja, ele foi ao Plenário da Câmara afirmar que eu havia dito aquilo que estava na publicação dele” (depoimento da vítima). 9. (a) O elemento subjetivo do tipo do crime de difamação é o animus diffamandi. (b) In casu, a defesa sustentou ausência de dolo de difamar, por dois fundamentos: (b.1) alegou que o vídeo “continha palavras do próprio querelante” e que estaria presente mero animus narrandi; (b.2) sustentou que os cortes realizados no vídeo tiveram finalidade exclusivamente técnica, com o único fim de reduzi-lo, para adequá-lo ao tamanho limite do suporte de mídia utilizado para veiculação. (c) As alegações não se sustentam: (c.1) Primeiramente, restou demonstrado que, embora o vídeo reproduza trecho da fala do Querelante, o corte realizado inverteu-lhe o sentido, atribuindo-lhe conotação racista. O uso, pelo réu, de trechos da fala do próprio Parlamentar Querelante reforçou sua potencialidade difamatória, porquanto o único elemento de verdade contido no vídeo induziu o público à ilusão de que todo o conteúdo correspondia à realidade, típico artifício ardiloso empregado para a prática da difamação; (c.2) Portanto, ao contrário do que ocorre na divulgação regida por mero animus narrandi, que se caracteriza quando há desconhecimento de sua natureza fraudulenta, in casu o Acusado detinha todas as informações necessárias para conhecer o descompasso entre o discurso efetivamente proferido pelo Autor e aquele divulgado no vídeo por ele disponibilizado no Facebook, com adulterações aptas a inverter o sentido da fala e a conferir-lhe teor racista; (c.3) Inverossímil, ainda, a alegação defensiva de que os cortes realizados tiveram não finalidade difamatória, mas sim mera função de redução da extensão da fala do Deputado Querelante, para atender às exigências do suporte midiático utilizado

para sua divulgação; (c.4) Deveras, se a intenção fosse unicamente reduzir o tamanho do vídeo, os cortes não teriam deturpado a fala do Querelante. Era possível excluir outros trechos da referida manifestação para atender ao propósito técnico, mas executou-se o corte cirurgicamente de modo a inverter diametralmente seu sentido. (d) Por fim, nas palavras da Procuradora-Geral da República, “caso o querelante estivesse realmente de boa-fé, tendo sido surpreendido com o fato, teria corrigido imediatamente e publicado alguma nota aclaratória e de desculpa sobre o ocorrido, atitude não tomada até o momento”. (e) Conclui-se que as provas colhidas nos autos comprovaram, além de qualquer dúvida razoável, a materialidade e a autoria delitivas, assim como o elemento subjetivo do tipo. 10. Ex positis, julgo procedente a acusação para condenar o réu Éder Mauro pela prática do crime de difamação agravada. 11. (a) Em sede de dosimetria, considero presentes quatro circunstâncias judiciais negativas, a conduzir a pena-base para 9 meses de detenção; ausentes atenuantes e agravantes, aplico a causa de aumento prevista no art. 141, III, do Código Penal (afasto, nos termos do art. 68, parágrafo único, a causa de aumento do inciso II do art. 141), alcançando a pena definitiva o total de 1 ano de detenção, no regime inicial aberto, e multa, no montante de 36 dias-multa, no valor de 1 salário mínimo cada. (b) Diante da presença dos pressupostos legais, substituo a pena privativa de liberdade pela de prestação pecuniária (art. 45, §1º, do CP), consistente no pagamento de 30 salários mínimos à vítima, que fixo como montante mínimo para reparação dos danos causados pela infração, nos termos do art. 387, IV, do Código de Processo Penal.”

(AP 1021, Relator(a): LUIZ FUX, Primeira Turma, julgado em 18-08-2020, ACÓRDÃO ELETRÔNICO DJe-254 DIVULG 20-10-2020 PUBLIC 21-10-2020)

Em suma, em 14 de maio de 2015, durante reunião da Comissão Parlamentar de Inquérito que apura a Violência contra jovens e negros pobres no Brasil, o Deputado Federal Jean Wyllys, fez uso da palavra para aludir a seguinte premissa: *“E aí a fala da Tatiana foi muito importante, porque ela traz essa dimensão histórica, que envolve a escravidão de negros; depois, a abolição, sem nenhuma política de inclusão no mercado de trabalho, a exclusão territorial; e, depois, toda uma produção de sentido que desqualifica essa comunidade como humana. Então, há um imaginário impregnado, sobretudo nos agentes das forças de segurança, de que uma pessoa negra e pobre é potencialmente perigosa, é mais perigosa do que uma pessoa branca de classe média. Esse é um imaginário que está impregnado na gente, uma dimensão aí. E os policiais partem desse imaginário”*.

Assim, em 19 de maio de 2015, cinco dias depois, o réu, Deputado Federal Eder Mauro, publicou um vídeo em seu perfil do Facebook contendo trecho recortado da referida fala. No entanto, o conteúdo havia sido modificado e editado, e neste vídeo o Deputado Federal Jean Wyllys apareceria falando a seguinte frase: *“Uma pessoa negra e pobre é potencialmente perigosa, é mais perigosa do que uma pessoa branca de classe média, essa é a verdade, então, dito isso...”*. Ou seja, o réu foi acusado por ter publicado e compartilhado o vídeo editado em suas redes sociais com o intuito de, dolosamente, atribuir-lhe conteúdo racista e preconceituoso, a fim de difamar a honra do querelante.

O vídeo falsificado do autor foi publicado no facebook e recebeu 14.834 curtidas, 252.458 visualizações e 12.272 compartilhamentos, sendo apenas excluído da página do Éder

Mauro no Facebook por determinação judicial⁷⁰. No que lhe concerne, a defesa alega a veracidade do vídeo divulgado pelo réu, que haveria tão somente reproduzido um trecho do debate parlamentar no âmbito da CPI. Por outro lado, o Laudo de Perícia Criminal 17.454/2017 do Instituto de Criminalística da Polícia Civil do Distrito Federal, alegou que “o vídeo questionado foi editado” e “que o processo de edição do vídeo questionado resultou na modificação da informação auditiva da fala do Deputado Jean Wyllys originalmente registrada no material padrão, conduzindo a uma compreensão diversa da realidade factual. Em outras palavras, o discurso do Deputado Jean Wyllys foi adulterado no vídeo questionado”.

Ainda, foram prestados diversos depoimentos em juízo que atestam o dano à honra do autor: (depoimento 1) “*essas informações geraram um impacto substantivo e absolutamente negativo da fala do Deputado Jean Wyllys junto aos ativistas do movimento negro, aos ativistas dos movimentos sociais*”; “*Eu sou do Estado da Bahia, em que há uma força enorme do movimento negro, e eu, como militante, fui intensamente questionado por que não combati a fala do Deputado Jean. E eu tentava explicar que o contexto da fala não teria sido aquele produzido pelo vídeo*” (depoimento do Deputado Federal Adalberto Souza Galvão); (depoimento 2) “*isso é um estrago muito profundo. Porque o deputado que tem esta bandeira, que é eleito com essa bandeira, que seu eleitorado aporta o voto, lhe oferta o voto em função desta bandeira que ele tem e isso é desse desconstruído através do vídeo*” (depoimento da Deputada Federal Érica Kokay).

Com isso, constatou-se que o conteúdo original da manifestação foi submetido a diversos cortes, que alteraram substancialmente seu sentido, conferindo-lhe conotação racista e preconceituosa, em desacordo com a intenção inicial do autor. Por unanimidade, a Primeira Turma, julgou procedente a acusação e condenou o réu Éder Mauro pela prática do crime de difamação agravada, previsto no art. 139, combinado com o art. 141, incisos II e III, do Código Penal, cominada com as agravantes disponíveis no art. 141 do Código Penal⁷¹, uma vez que o crime foi praticado contra funcionário público (Deputado Federal), em razão do exercício de suas funções; e a divulgação ocorreu de forma que facilitou o alcance e a repercussão da ofensa, por meio de plataforma capaz de atingir um público amplo.

⁷⁰ Decisão pública da 14ª Vara Cível de Brasília/DF, de 28 de agosto de 2015, disponível em: http://www.omci.org.br/m/jurisprudencias/arquivos/2017/df_00209599520158070001_28082015.pdf 4.

⁷¹ BRASIL. Decreto-Lei nº 2.848, de 7 de dezembro de 1940. **Código Penal**. Diário Oficial da União, Poder Executivo, Brasília, DF, 07 dez. 1940. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm. Acesso em 10 ago. 2025.

Por sua vez, em âmbito do Superior Tribunal de Justiça, destaca-se a decisão monocrática proferida com base no Habeas Corpus nº 972.309 (“HC”)⁷², de relatoria do Ministro Reynaldo Soares da Fonseca, cuja decisão foi publicada no Diário da Justiça Eletrônico Nacional em 13 de janeiro de 2025. Assim, o HC foi impetrado com pedido liminar em favor de Felipe Curtti, uma vez que, segundo o que consta nos autos, o agravante foi preso em flagrante no dia 14/11/2024 pela suposta prática dos crimes de violência política, associação criminosa, ameaça e injúria racial, e na audiência de custódia, sua prisão fora convertida em preventiva.

No caso em questão, a defesa sustenta que o agravante estaria submetido a constrangimento ilegal em razão do excesso de prazo na formação da culpa. Argumenta, ainda, pela inexistência dos requisitos que autorizam a prisão preventiva, uma vez que o respectivo decreto teria se fundamentado apenas na gravidade abstrata do delito e em justificativas genéricas. Defende, por fim, a suficiência da aplicação de medidas cautelares diversas da prisão, pleiteando, assim, a revogação da prisão preventiva do paciente.

No entanto, de acordo com o exposto na decisão, a prisão preventiva encontra-se devidamente justificada, diante da necessidade de resguardar a ordem pública, considerando a gravidade concreta das condutas imputadas ao paciente. Segundo a acusação, ele teria enviado um e-mail à vítima - Deputada Estadual do Rio de Janeiro - chamando-a de “*macaca esquerdista maldita*”, comparando sua filha a um animal e proferindo ameaças, nas quais afirmou desejar que a criança fosse estuprada e morta. Acrescentou, ainda, que, caso a situação do Brasil não melhorasse, mataria a vítima e toda a sua família.

Para a prática desses delitos, o paciente teria criado um endereço eletrônico em nome de Anderson Rocha Ludovico de Menezes, seu vizinho de condomínio, com quem já possuía desavenças. Em depoimento, Anderson afirmou que o paciente já havia criado, anteriormente, um e-mail em nome de sua esposa para ameaçar o Deputado Federal Elmar Nascimento, além de ter recorrido ao uso de *deepfake* e de inteligência artificial para produzir conteúdo pornográfico envolvendo o casal, acusá-los falsamente de pedofilia, solicitar empréstimos utilizando o automóvel de Anderson como garantia e criar anúncios falsos de

⁷² BRASIL. Superior Tribunal de Justiça, **Habeas Corpus nº972309** . Relator: Ministro Reynaldo Soares da Fonseca. Julgado em: 09 de janeiro de 2025.. Publicado no Diário da Justiça Eletrônico em 13/01/2025. Disponível em: <https://scon.stj.jus.br/SCON/pesquisar.jsp?livre=DEEPFAKE&operador=e&b=DTXT&thesaurus=JURIDICO&p=true&tp=T>. Acesso em 10 ago. 2025.

venda do veículo. Essas circunstâncias, no entendimento do Tribunal, evidenciam a periculosidade do agente e legitimam a manutenção da medida cautelar.

Ante o exposto, com fundamento no art. 34, XX, do Regimento Interno do Superior Tribunal de Justiça, o habeas corpus não foi conhecido. Ressalte-se que, nesse caso, a *deepfake* foi constatada no contexto da prática do crime de ameaça, tipificado no art. 147 do Código Penal, que pune a conduta de “*ameaçar alguém, por palavra, escrito ou gesto, ou qualquer outro meio simbólico, de causar-lhe mal injusto e grave*”.

Dessa forma, verifica-se que, no Brasil, já há ao menos dois precedentes judiciais, proferidos por tribunais superiores, envolvendo a produção de conteúdo por meio de inteligência artificial, ambos julgados com fundamento em tipos penais distintos. Além dos precedentes já identificados, a prática de criação e divulgação de *deepfakes* pode, à luz do ordenamento jurídico brasileiro, se enquadrar em diversos outros tipos penais, a depender do contexto e da finalidade da conduta. Entre eles, destacam-se:

(i) crimes contra a honra, como calúnia (art. 138, CP), difamação (art. 139, CP) e injúria (art. 140, CP), quando a manipulação de áudio ou vídeo é utilizada para imputar falsamente um crime, atribuir fato desonroso ou ofender a dignidade de alguém;

(ii) crimes contra a liberdade sexual, incluindo o registro não autorizado de intimidade sexual (art. 216-B, CP) e a divulgação de cena de sexo, nudez ou pornografia (art. 218-C, CP), agravados se envolver crianças ou adolescentes (arts. 240 e 241 do ECA);

(iii) crimes contra a honra e integridade de crianças e adolescentes, como a pornografia infantil (arts. 240 a 241-E, ECA), ainda que o material seja manipulado e não retrate situação real;

(iv) crimes contra a paz pública e a segurança do Estado, como falsidade ideológica (art. 299, CP), denúncia caluniosa (art. 339, CP), incitação ao crime (art. 286, CP) e apologia ao crime (art. 287, CP);

(v) crimes de ameaça e coação, incluindo a ameaça (art. 147, CP) e a coação no curso do processo (art. 344, CP);

(vi) crimes de fraude e estelionato, como estelionato (art. 171, CP) e fraude eletrônica (art. 171, §2º-B, CP); e

(vii) delitos previstos em legislações especiais, como a Lei Carolina Dieckmann (Lei nº 12.737/2012), o Marco Civil da Internet (Lei nº 12.965/2014) e a Lei Geral de Proteção de

Dados Pessoais (Lei nº 13.709/2018), quando houver invasão de dispositivos, uso indevido de dados ou danos causados na rede.

Embora o país conte com marcos normativos relevantes, que oferecem bases importantes para a proteção de dados e a responsabilização em crimes digitais, é evidente que ainda existe uma lacuna legislativa quanto à regulamentação específica de *deepfakes*. Tal diploma permitiria conferir uniformidade ao tratamento jurídico dessa tecnologia, adaptando a resposta penal à gravidade de cada conduta, mas assegurando coerência e previsibilidade no ordenamento jurídico.

Todavia, essa situação pode estar em vias de mudança com a tramitação do Projeto de Lei nº 1.272/2023. A proposta institui um novo artigo no Código Penal, o qual visa estabelecer penalidades para aqueles que produzirem ou disseminarem vídeos manipulados por meio de inteligência artificial com o objetivo de prejudicar a imagem ou a reputação de terceiros.⁷³ Nesse sentido, o texto legislativo estabelece-se da seguinte forma:

“Adulteração maliciosa de vídeos ou áudios

Art. 308-A. Adulterar arquivos de vídeo ou de áudio, mediante clonagem da voz, substituição de rosto, sincronização labial ou outra ferramenta de inteligência artificial, com a intenção de divulgar notícias falsas ou prejudicar pessoa física ou jurídica.

Pena – reclusão, de dois a quatro anos, e multa.

§ 1º Na mesma pena incide quem faz uso do vídeo ou do áudio, sabendo ser adulterado, para divulgação de notícia falsa ou para prejudicar pessoa física ou jurídica, se a conduta não constituir crime mais grave.

§ 2º Se o vídeo ou o áudio é divulgado na internet, redes sociais ou outro meio análogo:

Pena – reclusão, de quatro a oito anos, e multa.”⁷⁴

No entanto, apesar de compreender que tal projeto de lei representa um avanço significativo por parte do Estado, uma vez que com ele constata-se a preocupação do legislativo em criar uma norma própria para combater o uso malicioso de *deepfakes*, entendo que a construção do tipo penal em questão se deu de forma temerária. Assim,

⁷³ SILVA, Emanuely Ignez Alves; DE MENEZES, Filipe Eduardo Macedo. **Regulamentação dos Deep Fakes no Brasil: Desafios Jurídicos e Soluções Tecnológicas**. 1º Volume. Revista Mundo em Movimento. 2024. Disponível em: <https://revistamundoemmovimento.com.br/index.php/home/article/view/8>. Acesso em 11 ago. 2025.

⁷⁴ BRASIL. Câmara dos Deputados. **Projeto de Lei nº 123, de 10 de maio de 2023**. Altera a Lei nº 12.345/2022. Brasília, DF: Câmara dos Deputados, 2023. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2346816#:~:text=Ementa%3A%20Altera%20legisla%C3%A7%C3%A3o%20sobre%20o,o%20Estado%20Democr%C3%A1tico%20de%20Direito>. Acesso em 11 ago. 2025.

considero inadequado estabelecer uma única pena em face daqueles que praticam o ato de criar os *deepfakes*, uma vez que existem inúmeros tipos de vídeos fraudulentos, que se constroem em diferentes cenários, mais e menos gravosos, como assédio online, a extorsão digital, fraudes eletrônicas, roubo de identidade e violações de propriedade intelectual, disseminação não consensual de conteúdo pornográfico, exploração sexual infantil, a falsificação de provas eletrônicas e o uso de mídias sintéticas para espalhar desinformação.

Cada uma dessas hipóteses desencadeia riscos específicos à sociedade, de tal forma que, ao meu ver, não faria sentido enquadrá-las em uma mesma pena. Nesse sentido, entendo que o tipo penal deveria ser construído com um caput que apresente uma conduta genérica, como a supramencionada, com uma pena mínima pela prática da criação de *deepfakes* com a intenção de divulgar notícias falsas ou prejudicar pessoa física ou jurídica. Assim, tal caput deveria ser complementado por qualificadoras, que prevejam o aumento da sanção de acordo com o material criado, conforme a gravidade do conteúdo produzido e o impacto gerado em detrimento do interesse público. Assim, a título exemplificativo, utilizando, por exemplo, métrica apresentada na construção do tipo penal de roubo disciplinado por meio do Artigo 157 do Código Penal⁷⁵:

Roubo

Art. 157 - Subtrair coisa móvel alheia, para si ou para outrem, mediante grave ameaça ou violência a pessoa, ou depois de havê-la, por qualquer meio, reduzido à impossibilidade de resistência:

Pena - reclusão, de quatro a dez anos, e multa.

§ 1º - Na mesma pena incorre quem, logo depois de subtraída a coisa, emprega violência contra pessoa ou grave ameaça, a fim de assegurar a impunidade do crime ou a detenção da coisa para si ou para terceiro.

§ 1º-A A pena é de reclusão de 6 (seis) a 12 (doze) anos e multa, se a subtração for cometida contra quaisquer bens que comprometam o funcionamento de órgãos da União, de Estado ou de Município ou de estabelecimentos públicos ou privados que prestem serviços públicos essenciais.

§ 2º A pena aumenta-se de 1/3 (um terço) até metade:

I – (revogado);

II - se há o concurso de duas ou mais pessoas;

II - se a vítima está em serviço de transporte de valores e o agente conhece tal circunstância.

IV - se a subtração for de veículo automotor que venha a ser transportado para outro Estado ou para o exterior;

V - se o agente mantém a vítima em seu poder, restringindo sua liberdade.

⁷⁵ BRASIL. Decreto-Lei nº 2.848, de 07 de dezembro de 1940. Código Penal. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/Del2848compilado.htm. Acesso em: 02 set. 2020.

VI – se a subtração for de substâncias explosivas ou de acessórios que, conjunta ou isoladamente, possibilitem sua fabricação, montagem ou emprego.

VII - se a violência ou grave ameaça é exercida com emprego de arma branca;

VIII – se a subtração for de fios, cabos ou equipamentos utilizados para fornecimento ou transmissão de energia elétrica ou de telefonia ou para transferência de dados, bem como equipamentos ou materiais ferroviários ou metroviários.

§ 2º-A A pena aumenta-se de 2/3 (dois terços):

I – se a violência ou ameaça é exercida com emprego de arma de fogo;

II – se há destruição ou rompimento de obstáculo mediante o emprego de explosivo ou de artefato análogo que cause perigo comum.

§ 2º-B. Se a violência ou grave ameaça é exercida com emprego de arma de fogo de uso restrito ou proibido, aplica-se em dobro a pena prevista no caput deste artigo.

§ 3º Se da violência resulta:

I – lesão corporal grave, a pena é de reclusão de 7 (sete) a 18 (dezoito) anos, e multa;

II – morte, a pena é de reclusão de 20 (vinte) a 30 (trinta) anos, e multa.

Ainda, nessa análise deveria-se levar em consideração que a criação de *deepfakes* pode esbarrar em outros crimes já existentes no Código Penal, o que pode levar ao concurso formal do delito ou na absorção do delito de *deepfake* por outro, entendendo que o *deepfake* foi um meio para, por exemplo, a prática de um crimes contra a honra ou ainda crimes contra a paz pública e a segurança do Estado. Com isso, é essencial que tais disposições sejam vislumbradas para que a pena não seja leviana, e esteja de acordo com o ordenamento jurídico vigente, bem como, é necessário que esteja alinhada com as orientações do marco civil da internet e a regulação das redes sociais, o que vem sendo discutido na contemporaneidade.

Ademais, vale salientar que tal artigo penaliza apenas aquele que confecciona *deepfake*, sem entrar no mérito sobre a responsabilidade daqueles que compartilham os conteúdos fraudulentos. Entendo que a ausência de penalização desses agentes decorre da complexidade de sua identificação, tendo em vista o elevado número de indivíduos que disseminam tais materiais nas redes sociais. Nesse sentido, caso houvesse a previsão de sanção, esta deveria alcançar todos os que realizaram o compartilhamento, em observância ao princípio da isonomia, o que, contudo, se revela de difícil aplicação no ambiente digital até o momento.

De todo modo, no geral, ainda que entendo que a construção do tipo penal não foi a mais adequada, parto de uma premissa otimista em relação a regularização dos *deepfakes* no Brasil. No que tange a regulamentação das plataformas, apesar de não serem

responsabilizadas, a princípio, sobre os conteúdos de *deepfake* que são criados em sua *web*, é imposto aos provedores a obrigação de remover conteúdos ilícitos em situações específicas - como material pornográfico não consentido ou mediante ordem judicial. Particularmente, entendo que tal medida acaba sendo interessante uma vez que, por um lado, impede que o governo exerça um controle excessivo e autoritário na atividade das empresas, respeitando a ordem econômica, fundada na livre iniciativa, mas, ao mesmo tempo, promove um certo limite aos conteúdos que são criados pela tecnologia de *deepfake*.

Além disso, também é possível notar que o Brasil possui um arcabouço legislativo contemplado de diferentes normas já vigentes que reforçam a responsabilidade dos usuários que, de alguma forma, utilizam o *deepfake* para prejudicar terceiros. Com isso, ainda é visível a preocupação dos legisladores com a necessidade de que o arcabouço jurídico acompanhe, de forma específica, e na mesma proporção, o ritmo acelerado da inovação tecnológica. Com isso, em complemento com a legislação já vigente, entendo a importância do Projeto de Lei nº 1.272/202, o qual considero que representa um avanço ao propor a criminalização da manipulação de vídeos ou áudios por técnicas de inteligência artificial, como clonagem de voz, substituição de rosto ou sincronização labial, quando utilizadas para difundir notícias falsas ou causar prejuízos a terceiros, apesar da crítica tecida. No entanto, entendo que tal dispositivo pode ter sido elaborado de forma temerária, sem considerar as inúmeras variáveis de *deepfakes* que existem na contemporaneidade.

Posto isso, é certo que este desafio está apenas se iniciando, partindo do princípio de que novas tecnologias surgem cotidianamente e continuarão a emergir. É claro que as *deepfakes* atuais diferir-se-ão, significativamente, das que serão desenvolvidas nas próximas décadas e com isso, é papel do direito acompanhar tais mudanças sociais e se adaptar a elas. Atualmente, a dignidade da pessoa humana vai para além do mundo físico e tangível, sendo, muitas vezes construída no âmbito digital, cabendo ao direito adaptar-se para preservar as garantias constitucionais nesse âmbito, construindo normas sólidas e em harmonia ao ordenamento jurídico vigente, transmitindo ao ambiente virtual a preservação dos direitos fundamentais, especialmente aqueles que já estão tão expostos e sensíveis, como a intimidade, à vida privada, à honra e à imagem.

Considerações Finais

Posto isso, é indubitável salientar que os direitos fundamentais foram incorporados gradualmente na história da humanidade, se desenvolvendo a cada luta em prol de garantias institucionais. Assim, foram moldando a sociedade contemporânea sendo indispensáveis para garantir a manutenção de todos os direitos que já foram conquistados até então e, com isso, impedir retrocessos legislativos. O direito constrói suas próprias realidades. Sua positividade na Constituição Federal apenas reforça a sua inviolabilidade e relevância, acompanhada de suas características, que ressaltam seu funcionamento e princípios que norteiam a sua aplicação prática. Nesse sentido, mecanismos como o constitucionalismo digital são mais do que necessários para garantir que os direitos fundamentais se perpetuem no tempo e sigam se expandindo em frente a diferentes avanços tecnológicos, como os *deepfakes*.

Os *deepfakes* consistem em conteúdos digitais produzidos com elevado grau de sofisticação e precisão, desenvolvidos por meio de técnicas avançadas de codificação e decodificação de dados, o que lhes confere aparência autêntica e possibilita que conteúdos fraudulentos sejam facilmente confundidos com informações verídicas. Assim, o uso malicioso desta tecnologia tem gerado uma série de prejuízos e violações legais, e assim, entre os principais danos estão o assédio online, a extorsão digital, fraudes eletrônicas, roubo de identidade e violações de propriedade intelectual. Somam-se a esses a disseminação não consensual de conteúdo pornográfico, a exploração sexual infantil, a falsificação de provas eletrônicas e o uso de mídias sintéticas para espalhar desinformação, com potencial de manipular a opinião pública, gerar instabilidade social e intensificar a polarização política.

Inclusive, afrontam diretamente os direitos fundamentais à intimidade, à vida privada, à honra e à imagem, previstos no artigo 5º, inciso X, da Constituição da República Federativa do Brasil de 1988, cuja preservação constitui elemento central para a garantia da dignidade da pessoa humana. No que se refere aos direitos à intimidade e à vida privada, destaca-se que os *deepfakes* têm o potencial de difundir em larga escala, por meio da internet, conteúdos de caráter extremamente íntimo envolvendo determinada pessoa. Além disso, ao atribuir declarações ou condutas falsas a um indivíduo, essas criações podem atingir diretamente sua honra. Por fim, o uso indevido da imagem de alguém em montagens geradas por inteligência artificial configura violação ao direito personalíssimo de dispor sobre a própria representação visual.

Cumprе ressaltar, entretanto, que a tecnologia *deepfake*, em si, não representa um problema intrínseco. Trata-se de um recurso avançado de inteligência artificial, apto a criar

conteúdos audiovisuais sintéticos com elevado nível de realismo, de tal modo que existem aplicações não maliciosas dessa tecnologia que merecem destaque, como em estudos acadêmicos ou ainda biotecnológicos,

Diante desse panorama, impõe-se a necessidade de compreender quais seriam os caminhos possíveis para regulamentar as tecnologias *deepfakes*. Assim, a partir de um estudo jurídico internacional, observa-se uma grande diversidade de normas e formas de lidar com estes conteúdos. Essa perspectiva reflete não apenas a forma como cada nação enxerga a responsabilidade dos usuários e das plataformas, mas também as características estruturais do Estado e o grau de controle exercido sobre a informação.

No que se refere à responsabilidade dos usuários, há países que optaram por criar legislação específica para o enfrentamento das *deepfakes*. Nos Estados Unidos, por exemplo, o estado da Virgínia concentrou sua norma na repressão à prática conhecida como *revenge porn* (pornografia de vingança). Já no Texas, a preocupação recai especialmente sobre os impactos nocivos das *deepfakes* no cenário político, com atenção voltada à integridade dos processos eleitorais. A Califórnia, por sua vez, adotou uma abordagem legislativa mais ampla, contemplando tanto a erradicação de conteúdos manipulados digitalmente de natureza política quanto aqueles de cunho pornográfico. Na Europa, destaca-se a legislação do Reino Unido, que também trata das *deepfakes*, com foco na repressão à pornografia.

Por outro lado, há países que, embora não possuam legislação própria voltada ao tema, aplicam dispositivos já existentes para lidar com casos envolvendo *deepfakes*. No continente africano, por exemplo, observa-se que nações como Nigéria, Egito e África do Sul recorrem à adaptação de normas gerais para enfrentar os desafios impostos por essa tecnologia. Esse mesmo cenário se repete na Oceania, com a Austrália adotando estratégia semelhante. No Oriente Médio, os Emirados Árabes Unidos não contam com lei específica, mas elaboraram um guia orientativo à população, ressaltando a importância da proteção de dados pessoais e a detecção de conteúdos falsificados por meio de ferramentas baseadas em inteligência artificial.

No que se refere à responsabilização das plataformas, a China é um exemplo emblemático de controle estatal centralizado sobre o acesso à internet, caracterizado por forte censura à liberdade de expressão. Nesse ambiente de monitoramento sistemático das atividades individuais, o governo busca controlar rigorosamente o conteúdo disseminado nas redes digitais, o que favorece a criação, aplicação e fiscalização de normas específicas para regular a produção e circulação de *deepfakes*.

Nos Estados Unidos, o modelo é descentralizado, de tal forma que a definição e aplicação das regras recaem majoritariamente sobre os próprios usuários do serviço. Embora o governo possa criar leis e sanções, estas geralmente visam penalizar o usuário que cria ou compartilha o conteúdo adulterado, e não a plataforma que o hospeda. Um marco legal relevante é a Seção 230 do *Communications Decency Act*, (Lei de Decência nas Comunicações) que estabelece que nenhum provedor ou usuário de serviço interativo de computador deve ser tratado como responsável pelo conteúdo criado por terceiros.

O Reino Unido adota um modelo intermediário, atribuindo responsabilidades aos usuários e também às plataformas, mas sem controle excessivo. O *Online Safety Bill* (Lei de Segurança Online) prevê, por exemplo, que provedores de serviços de *deepfake* verifiquem a identidade dos usuários para rastrear conteúdos manipulados maliciosamente até seus autores. Também exige avaliações de risco para crianças e adultos, com o objetivo de prevenir exploração ou abuso sexual infantil e proteger a dignidade dos indivíduos, como nos casos de *revenge porn*.

Em relação ao Brasil, no que concerne à responsabilidade das plataformas, o Marco Civil da Internet, por meio de seu artigo 18, isenta os provedores de conexão de responsabilidade civil por conteúdo de terceiros, exceto nos casos previstos no artigo 21, que impõe a remoção de material pornográfico não autorizado mediante notificação extrajudicial. Assim, as plataformas só podem ser responsabilizadas por *deepfakes* quando envolverem conteúdo íntimo não consentido ou houver ordem judicial de remoção, o que acaba representando uma boa medida de prevenção da arbitrariedade estatal.

Além disso, é notável que o Brasil possui um arcabouço legislativo abrangente, composto por diferentes normas já vigentes, que reforçam a responsabilidade dos usuários que prejudicam terceiros, dentre as quais, destacam-se: (i) Crimes contra a honra - calúnia (art. 138, CP), difamação (art. 139, CP) e injúria (art. 140, CP); (ii) Crimes contra a liberdade sexual (art. 216-B, CP) e divulgação de cena de sexo, nudez ou pornografia (art. 218-C, CP), com agravantes se envolver menores (arts. 240 e 241, ECA); (iii) Marco Civil da Internet (Lei 12.965/2014) e (iv) Lei Geral de Proteção de Dados (Lei 13.709/2018), aplicáveis em casos de invasão de dispositivos, uso indevido de dados ou danos na rede.

Especificamente, ainda surgiu o Projeto de Lei nº 1.272/2022, o qual visa criminalizar a manipulação de vídeos ou áudios por técnicas de inteligência artificial, como clonagem de voz, substituição de rosto ou sincronização labial, quando utilizadas para difundir notícias falsas ou causar prejuízos a terceiros, o qual, aparentemente, é uma medida promissora sob o

ponto de vista legal. Apesar disso, faz-se a crítica de que não seria adequado enquadrar todas as situações de uso de *deepfakes* sob uma mesma pena, uma vez que apresentam níveis distintos de gravidade. Desse modo, entendo que seria mais prudente estabelecer uma pena-base, complementada por agravantes proporcionais à gravidade do ato praticado e aos danos ocasionados. De todo modo, o avanço legislativo, representa um progresso positivo e necessário.

Posto isso, conclui-se que há uma preocupação global sobre o fenômeno dos *deepfakes*, bem como sobre a forma com que o direito deve se adaptar para respaldar a inovação tecnológica. Nesse contexto, cabe ao legislador enfrentar o desafio de construir um arcabouço jurídico alinhado aos princípios do constitucionalismo digital, com o objetivo de oferecer segurança jurídica, preservar a reputação e reduzir riscos em um espaço virtual cada vez mais complexo e dinâmico e, ao mesmo tempo, preservar o avanço científico e tecnológico de forma sustentável, pautando-se na ordem econômica, fundada na valorização na livre iniciativa. Diante disso, torna-se imprescindível que o direito acompanhe as mudanças sociais e tecnológicas, atualizando-se para regular efetivamente a vida em sociedade na era digital, com especial atenção à proteção dos direitos fundamentais, essenciais para a proteção e manutenção do Estado democrático de direito e da própria dignidade da pessoa humana que se constrói tanto no mundo tangível quanto no digital.

Referências Bibliográficas

- ABREU, Neide Maria Carvalho. **Direitos Fundamentais na Constituição Federal de 1988.** Anais do Conselho, 2007. Disponível em: <http://publicadireito.com.br/conpedi/manaus/arquivos/anais/XIVCongresso/055.pdf> . Acesso em 01 jul. de 2025.
- AFFONSO, Filipe José Medon. **O Direito à Imagem na Era das Deep Fakes.** 27º Volume. Revista Brasileira de Direito Civil, 2021.
- BASTOS, Celso Ribeiro. **Curso de Direito Constitucional.** São Paulo: Saraiva, 1997.
- BITTAR, Carlos Alberto. **Os Direitos da Personalidade.** Rio de Janeiro: Forense Universitária, 1995.
- BOBBIO, Norberto. **A Era dos Direitos.** 1ª Edição. Rio de Janeiro: Campus, 1992.
- BONAVIDES, Paulo. **Curso de Direito Constitucional.** 19ª Edição. São Paulo: Malheiros, 2006.
- BUENO, Francisco da Silveira. **Dicionário Escolar da Língua Portuguesa.** Rio de Janeiro: FENAME, 1956.
- CARVALHO JÚNIOR, O. L. de; CARVALHO, S. S.; SOUSA, B. A. de. **Democracia em Xequê: Inteligência Artificial e Deep Fake.** 7º Volume. São Paulo: Revista JRG de Estudos Acadêmicos, 2024. Disponível em: <https://revistajrg.com/index.php/jrg/article/view/1743>. Acesso em: 26 jul. 2025.
- CAVERO, José Martinez De Pisón; NUNES JÚNIOR, Vidal Serrano. **A Proteção Constitucional da Informação e o Direito à Crítica Jornalística.** São Paulo: FTD, 1997.
- CELESTE, Edoardo. **Digital Constitutionalism: The Role of Internet Bills of Rights.** Nova Iorque: Routledge, 2023.
- DE TEFFÉ, Chiara Spadaccini. **Considerações sobre a Proteção do Direito à Imagem na Internet.** 54º Volume. Revista de Informação Legislativa, 2017.
- FERRAZ, Tércio Sampaio. **Sigilo de Dados: Direito à Privacidade e os Limites à Função Fiscalizadora do Estado.** 88º Volume. Revista da Faculdade de Direito da USP, 1999.
- GILL, Lex; REDEKER, Dennis; GASSER, Urs. **Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights.** The Berkman Center for Internet & Society at Harvard University, nº 2015-15. Disponível em: <http://dx.doi.org/10.2139/ssrn.2687120>. Acesso em: 02 jul. 2025.

- GONÇALVES, Victor Hugo Pereira. **Marco Civil da Internet Comentado**. 1ª Edição. São Paulo: Atlas, 2017.
- GOODFELLOW, Ian. **Generative Adversarial Nets**. Advances in Neural Information Processing Systems, 2014.
- GUERRA, Sidney. **Direito Fundamental à Intimidade, Vida Privada, Honra e Imagem**. 2º Volume. Rio de Janeiro: Freitas Bastos, 2006. Disponível em: http://www.publicadireito.com.br/conpedi/manaus/arquivos/anais/recife/direitos_fundam_sidney_guerra.pdf. Acesso em: 01 jul. 2025.
- JOANGUETE, Celestino. **O Poder das Mentiras Digitais: A Ameaça das Deepfakes para a Sociedade**. Editora Dialética, 2023. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=dtrPEAAQBAJ&oi=fnd&pg=PT4&dq=impacto+social+deepfake&ots=1TefYZ7zew&sig=TwpmYE9YOifhK1fHxTbtXjBJpPQ&redir_esc=y#v=onepage&q=impacto%20social%20deepfake&f=false. Acesso em: 26 jul. 2025.
- LIU, Min; XIJIN, Zhang. **Deepfake Technology and Current Legal Status of It**. 3rd International Conference on Artificial Intelligence and Education (IC-ICAIE 2022), 2022.
- NNAMDI, Nmesoma; ONIYINDE, Omolade Adeyemi; BABALOLA, Abegunde. **An Appraisal of the Implications of Deep Fakes: The Need for Urgent International Legislations**. 8º Volume. American Journal of Leadership and Governance, 2023. Disponível em: <https://www.researchgate.net/publication/372611418>. Acesso em: 16 ago. 2025.
- NOVELINO, Marcelo. **Direito Constitucional**. 2ª Edição. São Paulo: Método, 2008.
- RAMADHANI, K. N.; MUNIR, R. **A Comparative Study of Deepfake Video Detection Method**. Yogyakarta, Indonesia: International Conference on Information and Communications Technology (ICOIACT), 2020.
- SALEME, Edson R. **Direito Constitucional**. 5ª Edição. Barueri: Manole, 2022. E-book. ISBN 9786555766370. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9786555766370/>. Acesso em: 17 mai. 2025.
- SILVA, Emanuely Ignez Alves; DE MENEZES, Filipe Eduardo Macedo. **Regulamentação dos Deep Fakes no Brasil: Desafios Jurídicos e Soluções Tecnológicas**. 1º Volume. Revista Mundo em Movimento, 2024. Disponível em: <https://revistamundoemmovimento.com.br/index.php/home/article/view/8>. Acesso em: 11 ago. 2025.

- SILVA, José Afonso da. **Curso de Direito Constitucional Positivo**. 19ª Edição. São Paulo: Malheiros Editores, 2000.
- TAKANO, Camila Cardoso; DA SILVA, Lucas Gonçalves. **O Constitucionalismo Digital e as Novas Tecnologias da Informação e Comunicação (TIC)**. 6º Volume. Revista de Direito, Governança e Novas Tecnologias, 2020.
- TEIXEIRA, Carla Costa. **O Preço da Honra**. Universidade de Brasília, Departamento de Antropologia, 1999.
- YUZHI, Zhang; RUIFANG, Wang; LIANG, Zhu. **The Review of Generation and Detection Techniques for Deepfakes**. Journal of Information Security Research, 2022. Disponível em: <https://www.atlantis-press.com/proceedings/ic-icaie-22/125981029>. Acesso em: 02 jul. 2025.