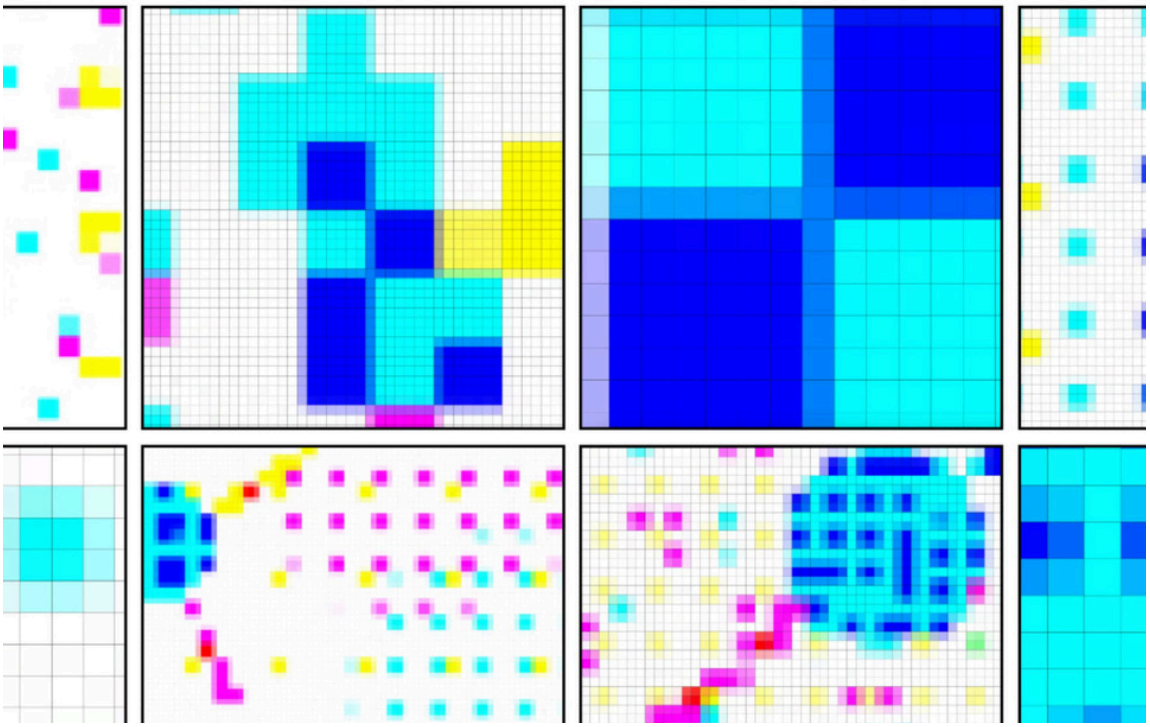
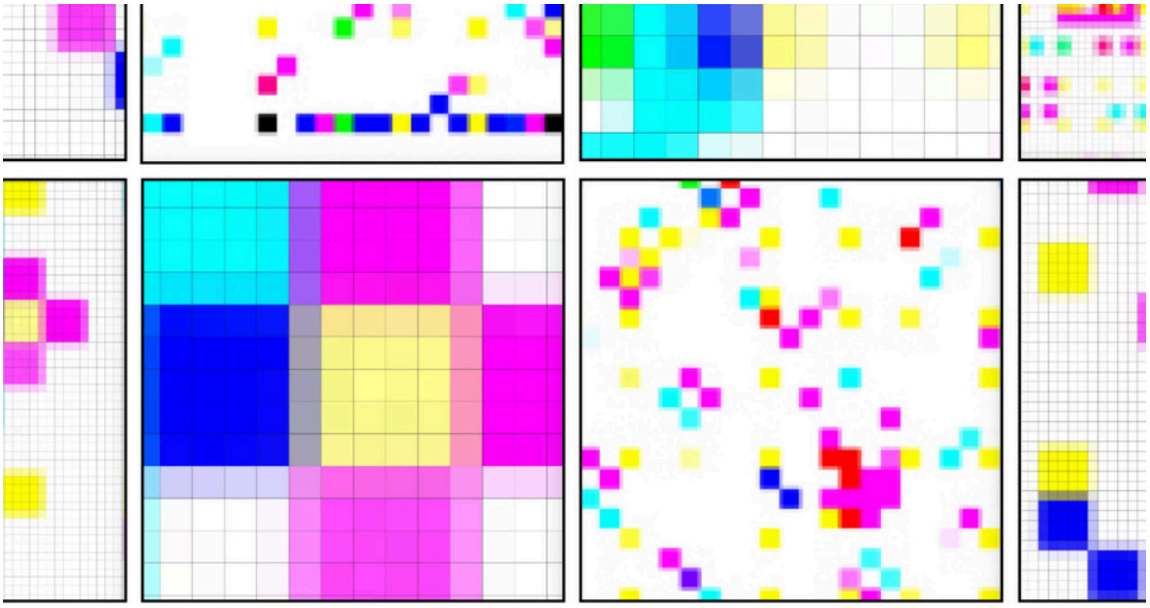


Enfrentando
DEEPFAKES

Gustavo Souza e Tarcízio Silva (org.)



desvelar

A imagem da capa em CC BY 4.0. Game of Pixels x Toy Models, de Elise Racine.

Gustavo Souza e Tarcízio Silva (org.)

Enfrentando DEEPFAKES

desvelar.org

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Enfrentando deepfakes [livro eletrônico] /
organizadores Gustavo Souza, Tarcízio Silva. -- 1. ed. -- Brasília, DF : Desvelar, 2025. PDF

Vários autores.
Bibliografia.
ISBN 978-65-985781-1-4

1. Comunicação digital 2. Cultura digital 3. Gênero e sexualidade 4. Inteligência artificial - Inovações tecnológicas 5. Notícias falsas 6. Racismo 7. Tecnologia I. Souza, Gustavo. II. Silva, Tarcízio.

25-274300 CDD-660.02

Índices para catálogo sistemático:

1. Tecnologia 660.02

Aline Graziele Benitez - Bibliotecária - CRB-1/3129

Sumário

1. Deepfake como questão de gênero, por Ana Gabriela Ferreira.

2. Infância, juventude e deepfakes: o caminho do PL 3821/2024 para uma proteção digital efetiva, por Diego Abreu.

3. Deepfakes como ameaça global e multidimensional, por Lauro Accioly Filho.

4. Racismo recreativo em tempos de deepfake, por Glenda Dantas

5. O corpo preto, por Gustavo Souza.

6. Deepfakes e a fabricação do real: a distorção da realidade como forma de controle político social, por Bianca Marques Galvão.

7. Deepfake, desinformação e racismo: ciber-resistência como estratégia de defesa para populações negras e periféricas nas redes sociais, Gabriela de Almeida Pereira.

8. Deepfakes e desordem informacional: a banalização do sintético, por Laura Pereira.

9. Deepfakes no cenário brasileiro: efeitos e estratégias de mitigação, Júlia Caldeira.

10. É tudo verdade? A Moça do Táxi e as deepfakes, por Thiane Neves Barros.

11. Casos, sistematizados por Lauro Accioly Filho e Gustavo Souza

DEEPFAKE COMO QUESTÃO DE GÊNERO

Ana Gabriela Ferreira

Como citar este artigo?

FERREIRA, Ana Gabriela. Deepfake como questão de gênero. In: SOUZA, Gustavo; SILVA, Tarcizio (org.). **Enfrentando Deepfakes.** Desvelar: Brasília, 2025. Disponível em: desvelar.org/enfrentando-deepfakes.

Não é uma novidade que as ferramentas de inteligência artificial (IA) reproduzem vieses sociais e podem, se não reguladas devidamente, apressar uma intensificação das opressões vivenciadas. No contexto de violências de gênero, os usos de IA para criação de *deepfakes* estão diretamente associados aos danos ao feminino. De imediato, pela geração de conteúdos sexuais explícitos falsos através da imagem de meninas e mulheres vítimas. Além disso, pela reificação de um ideal de feminino inexistente e simulado no contexto da pornografia, que afeta a visão masculina sobre sua relação com o feminino, reforçando contextos de poder em lugar de interdependência.

Deepfakes são conteúdos falsos criados através de *deep learning*, em que ferramentas são treinadas com imagens e características como vozes reais e, a partir disso, são capazes de inserir tais características em cenários existentes, criando quadros de vídeo que simulam uma realidade falsa, mas verossímil[1]. As características apreendidas das pessoas são utilizadas em vídeos totalmente falsos, mas com alto grau de realismo[2].

Uma pesquisa realizada pela Deeptrace[3] demonstrou que cerca de 96% dos vídeos de *deepfakes* são de conteúdo pornográfico ou de relações sexuais não consentidas, **exclusivamente** produzidos sobre mulheres. Nos vídeos de conteúdo não pornográfico, isso se modifica – 61% dos representados são do sexo masculino. A diferença denota um alvo específico nos usos de *deepfakes* como meio de exposição sexual indevida – o feminino. Não é exagero, portanto, dizer que as deepfakes têm funcionado como verdadeiras armas de gênero.

---

[1] SPIVAK, Russel. *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*. Georgetown Technology Review, [s.l.], Vol. 3.2, p. 339–400, 2019. p. 344. Disponível em: <https://georgetownlawtechreview.org/deepfakes-the-newest-way-to-commit-one-of-the-oldest-crimes/GLTR-05-2019/ >. Acesso em 10/05/2025.

[2] VENEMA, Agnes E. Deepfakes como uma questão de segurança: por que o gênero importa. Disponível em: < https://wiisglobal.org/deepfakes-as-a-security-issue-why-gender-matters/> Acesso em 11/05/2025.

[3] AJDER, *et al*. The State of Deepfakes: Landscape, Threats, and Impact. Henry Ajder, Giorgio Patrini, Francesco Cavalli, and Laurence Cullen, p. 8. September 2019. Disponível em: <https://regmedia.co.uk/2019/10/08/deepfake_report.pdf > Acesso em 10/05/2025.

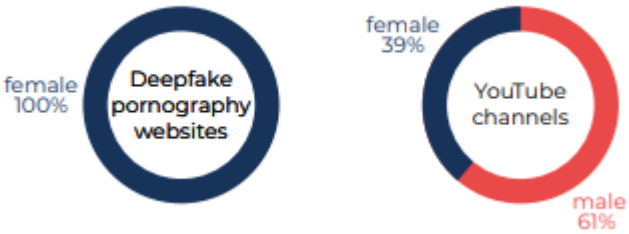


Figura 1: Gráfico demonstra o percentual de representações femininas e masculinas em *deepfakes.*[4] 100% das representações em pornografia são de mulheres. Fora das *deepfakes* pornográficas, 61% das representações são masculinas.

No mesmo sentido, a pesquisa State of Deepfakes[5], desenvolvida pela Security Hero, mostra que 98% dos conteúdos de deepfake em 2023 eram pornográficos e houve um crescimento de 464% na quantidade pornografia deepfake criada. O público alvo é notório – 99% das atingidas são mulheres.

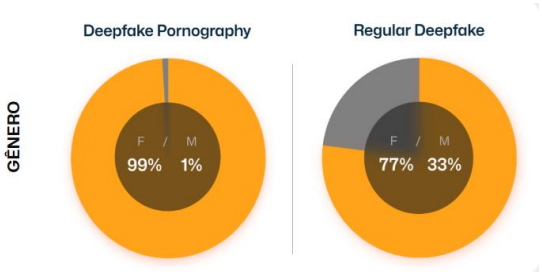


Figura 2: percentual de deepfake pornográfica por gênero[6]

Figura 3: número de pornografia deepfake[7]

[4] AJDER, *et al*. *Op. Cit*., p.2.
[5] HOME SECURITY HEROES. 2023 State of Deepfakes: realities, threats, and impact. 2023. Disponível em: https://www.homesecurityheroes.com/state-of-deepfakes/. Acesso em 10/05/2025.
[6] Op Cit.
[7] Op. Cit.

O desenvolvimento das redes operou uma modificação nas possibilidades de controle sobre o feminino, num momento de *backlash*[8] como poucos vistos na história quanto aos avanços na equidade de gênero. De um lado, cresce o fenômeno de grupos ditos "masculinistas" que encontraram nas redes acesso a uma cultura que reifica o ideal de opressão de gênero e de desqualificação do feminino como maneira de reorganizar o ideal do "ser homem", aparentemente abalado com a crise neoliberal. De outro, novas ferramentas como as *deepfakes* permitiram a criação de uma esfera de submissão *pretensamente* absoluta – aqui, se o domínio real do feminino é negado ao masculino, a ferramenta permite criar vídeos verossímeis de qualquer violência que se imagine, através da atribuição de uma realidade simulada à imagem e características da vítima.

Sabemos que o controle de corpos e da sexualidade feminina é um modo de operar o poder em todas as estruturas sociais. Conforme Tilio et al[9]:

> "De acordo com Foucault (**1999**), o corpo resulta de um conjunto de forças que não se limitam aos seus aspectos orgânicos, pois sobre ele operam diferentes dispositivos de poder. Por ser resultado de uma série de investimentos de forças externas, o corpo/sexo não é a causa das relações de poder, mas sim efeito e resultado delas. Todavia, o sistema binário (masculino versus feminino) típico de nossa sociedade produziu e reproduziu a ideia de que o gênero refletiria o sexo e, assim, as características psicossociais se dariam de acordo com as supostas disposições naturais dos corpos (**Bento, 2012**). Trata-se, portanto, de um processo de naturalização.
>
> No espaço público o corpo das mulheres que performa aspectos da feminilidade assume duas possíveis interpretações (**Matos & Soihet, 2003**): corpo privado que deve permanecer oculto, e corpo público apropriado e dominado pelos homens para ser exibido como troféu. É nesse cenário que a violência contra as mulheres nas relações entre gêneros adquire consistência."

Há, portanto, um mecanismo que permite realizar o que parece ser o desejo advindo do backlash de gênero – recriar um modelo de submissão total do feminino ao masculino, desumanizando e objetificando a vítima através da imagem e, ao mesmo tempo, promover uma "punição" à mulher, ambos via uso e manipulação de

---

[8] Cf.: https://www.hrw.org/news/2023/03/07/global-backlash-against-womens-rights

[9] TILIO, Rafael *et al.* Corpo Feminino E Violência De Gênero: Uma Análise Do Documentário "Chega De Fiu Fiu". In: Psicol. Soc. 33 • 2021. Disponível em: https://www.scielo.br/j/psoc/a/wWtLhjQP3hRQC5hDt6Pz7qq Acesso em 11/05/2025

deepfakes. Segundo dados da pesquisa State of the deepfakes, um em cada três aplicativos de *deepfake* permite a criação de pornografia. Simbolicamente, a percepção de que as *deepfakes* são 96 a 98% formadas por pornografia e imagens de sexo não consentido traduz muito da experiência coletiva de ser mulher neste momento cultural e na era das redes. São, portanto, um sintoma de uma onda de violência e opressão facilitada e abrigada pelas tecnologias[10].

De um lado, a criação de imagens falsas de meninas e mulheres em contextos sexuais explícitos, veiculados pelas redes, tem gerado danos agigantados à sua psique e segurança. E, com a dificuldade de demonstração de que as imagens são falsas, até que a *deepfake* seja removida, se alastram versões que afetam as relações sociais da vítima, expondo-a a um processo de violento descrédito. Através das deepfakes, instituem-se duas esferas de controle. O controle da imagem feminina que passa a ser descredibilizada porque atacada na esfera sexual e o pseudo-controle do corpo, através da simulação de realidade, gerando efeitos nefastos.

> "estas manipulaciones tienen repercusiones graves en la vida personal y profesional de las mujeres afectadas, dañando su reputación, seguridad y bienestar emocional. Las víctimas pueden experimentar desde acoso en línea hasta daños en sus relaciones personales y profesionales, lo que subraya la urgencia de abordar este fenómeno. La proliferación de este tipo de contenido falso o manipulado plantea desafíos significativos para la detección y mitigación de la ciberviolencia contra las mujeres.
>
> Dada la naturaleza altamente realista de las imágenes y videos generados por IA, puede resultar difícil para los usuarios discernir entre contenido auténtico y manipulado, lo que aumenta el riesgo de que la desinformación y la difamación se propaguen sin control. Esto también puede llevar a una erosión de la confianza en los medios de comunicación y en las plataformas digitales, donde la veracidad de la información es cada vez más cuestionada"

Atentos às possibilidades de danos diversos, muitos países[11] têm buscado regulamentar os usos de ferramentas tecnológicas ou criminalizar a criação de

---

[10] CHAMPION, AR, OSWALD, F., KHERA, D. *et al.* Examinando os impactos de gênero da violência sexual facilitada pela tecnologia: uma abordagem de métodos mistos. *Arch Sex Behav* **51** , 1607–1624 (2022). https://doi.org/10.1007/s10508-021-02226-y

[11] Cf.: https://www.techpolicy.press/the-uks-online-safety-act-is-not-enough-to-address-nonconsensual-deepfake-pornography/

conteúdo íntimo ou sexual não autorizado via *deepfake*. É o que aconteceu no Brasil[12], com a previsão de aumento de pena para violência psicológica praticada por meio de tecnologias, o que engloba as *deepfakes*. A criminalização, no entanto, não é suficiente, já que os danos causados se alastram rapidamente e a reparação à imagem e ao abalo psicológico da vítima serão sempre insuficientes diante de sua exposição agigantada nas redes. Além disso, sabemos, mesmo a eventual sanção do ofensor dependerá dos filtros inerentes ao sistema criminal[13] .

Mecanismos eficazes de contenção parecem envolver dois fatores básicos – o impedimento de que as plataformas permitam a criação de deepfakes sexuais ou pornográficas e a obrigação de que incluam mandatoriamente a identificação de conteúdo gerado por IA em qualquer deepfake e, em conjunto, a estruturação de uma revisão da cultura de gênero que permita questionar o modelo de masculinidade naturalizador dos usos de diversos meios para perpetrar violência de gênero.

Não há solução para os usos de tecnologias como armas de opressão sem que haja abordagem conjunta do controle e regulação das tecnologias e revisão dos modelos que permitem socialmente a opressão. Meios que apenas atuam na esfera simbólica da punição, ao invés da evitação por via regulamentadora e cultural, abusam de um direito de emergência, utilizado para estancar problemas que precisam de modificação das estruturas que os engendraram. Vale dizer – não se resolve apenas na IA o que é problema socialmente estruturado.

[12] https://www.camara.leg.br/noticias/1153179-sancionada-lei-que-agrava-pena-em-crime-de-violencia-contra-a-mulher-com-uso-de-ia/

[13] MONTEIRO BARRETO, B. L., & de SOUZA PREUSSLER, G. (2023). Sistema penal e violência racial: uma discussão teórica decolonial. *Revista De Direitos E Garantias Fundamentais*, *24*(3), 189–219.

INFÂNCIA, JUVENTUDE E DEEPFAKES: O CAMINHO DO PL 3821/2024 PARA UMA PROTEÇÃO DIGITAL EFETIVA

Diego Abreu

Como citar este artigo?

ABREU, Diego. Infância, juventude e deepfakes: o caminho do PL 3821/2024 para uma proteção digital efetiva (org.). In: SOUZA, Gustavo; SILVA, Tarcizio. **Enfrentando Deepfakes.** Desvelar: Brasília, 2025. Disponível em: desvelar.org/enfrentando-deepfakes.

A expansão acelerada das tecnologias digitais têm reconfigurado não apenas os modos de interação social, mas também os contornos da proteção jurídica de direitos fundamentais, com impactos particularmente intensos sobre crianças e adolescentes. Inserida nesse cenário, a disseminação das chamadas deepfakes — conteúdos audiovisuais manipulados por inteligência artificial que simulam, com alto grau de fidelidade, imagens, áudios ou vídeos de pessoas — constitui uma ameaça concreta à dignidade infantojuvenil, sobretudo quando utilizada para fabricar material de nudez ou conotação sexual, expondo sujeitos em desenvolvimento a violências simbólicas e psicológicas de proporções devastadoras.

O PL 3821/2024 propõe alterações no Código Penal e no Código Eleitoral para criminalizar a produção, manipulação ou divulgação de conteúdo falso de nudez ou ato sexual gerado por inteligência artificial ou outros meios tecnológicos. A pena prevista é de reclusão de dois a seis anos e multa, podendo ser aumentada em diversos casos, como, por exemplo, de um terço até a metade, se a vítima for mulher, criança, idoso ou pessoa com deficiência; de um terço até o dobro, se houver disseminação em massa em redes sociais ou plataformas digitais; e, especificamente no contexto eleitoral, pode variar de dois a oito anos de reclusão, além de multa, com a possibilidade de cassação do registro de candidatura ou do diploma caso o crime seja cometido por um candidato [1].

Apesar de o projeto avançar ao reconhecer a especificidade da violência digital impulsionada por novas tecnologias, ele também revela tensões jurídicas relevantes, que demandam análise crítica, especialmente diante dos limites da atuação penal frente a condutas que se inserem em dinâmicas digitais cada vez mais complexas. A esse respeito, o relatório *Construindo Consensos*, publicado em 2023 por uma coalizão de organizações da sociedade civil, evidencia que as eleições municipais brasileiras apresentaram dificuldades significativas na coleta de provas relacionadas

[1] GAZETA DO POVO. Câmara aprova punição para quem divulgar imagem de nudez gerada por IA. *Gazeta do Povo*, 24 abr. 2024. Disponível em: https://www.gazetadopovo.com.br/republica/camara-aprova-punicao-para-quem-divulgar-imagem-de-nudez-gerada-por-ia/. Acesso em: 10 de maio de 2025.

a crimes digitais, bem como na responsabilização eficaz dos agentes envolvidos e na atuação tempestiva das plataformas digitais na remoção de conteúdos prejudiciais [2].

A manipulação de imagens envolvendo crianças e adolescentes, em particular, suscita implicações ainda mais sensíveis, pois ultrapassa a proteção da honra ou da imagem para atingir diretamente a integridade psíquica e o desenvolvimento saudável de sujeitos em formação, conforme previsto pelo Estatuto da Criança e do Adolescente (ECA), que garante prioridade absoluta aos direitos da população infantojuvenil [3]. Nessa esfera, quando conteúdos falsos sexualizados envolvendo crianças ou adolescentes são produzidos ou disseminados, mesmo sem base em imagens reais, o impacto psicológico e social pode ser devastador, sobretudo porque a percepção coletiva tende a não distinguir o que é verdadeiro do que é manipulado.

Diante disso, torna-se necessário interpretar a legislação vigente — que já tipifica crimes como a pornografia infantil — à luz dos novos contextos tecnológicos, uma vez que ainda há lacunas quanto à responsabilização por conteúdos sintéticos que simulam situações sem necessariamente utilizar material real. O PL 3821/2024 é louvável ao incluir essas situações no campo da criminalidade penal e eleitoral, pois reconhece que a aparência de veracidade das imagens criadas por inteligência artificial pode gerar efeitos devastadores na vida das vítimas.

De forma coerente, o projeto também reconhece que determinados grupos — como mulheres, crianças, idosos e pessoas com deficiência — são especialmente suscetíveis a esse tipo de violência digital, o que justifica a previsão de agravantes nas penas aplicadas. A deputada Amanda Gentil (PP-MA), autora da proposta, destacou que os *deepnudes* – imagens falsas de nudez geradas por IA – contribuem para afastar mulheres da política e reforçam sua sub-representação nos espaços de poder. Esse mesmo argumento pode ser estendido às crianças e adolescentes, que,

---

[2] **CONSTRUINDO CONSENSOS: Recomendações para plataformas digitais nas eleições brasileiras.** São Paulo: Coalizão de Organizações da Sociedade Civil, 2023. Disponível em: conectas.org/noticias/organizacoes-cobram-mudancas-das-big-techs-contra-desinformacao-e-violencia-politica-nas-eleicoes-brasileiras/. Acesso em: 10 de maio de 2025.

[3] BRASIL. **Lei nº 8.069, de 13 de julho de 1990**. Dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências. Diário Oficial da União: seção 1, Brasília, DF, 16 jul. 1990. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8069.htm; Acessado em: 10 de maio 2025.

ao sofrerem esse tipo de exposição, têm sua autoestima, confiança e relações sociais comprometidas, sendo privadas de um desenvolvimento pleno e saudável.

Conquanto, apesar da proposta legislativa representar um passo importante, é preciso reconhecer que a criminalização isolada não tem se mostrado eficaz para prevenir práticas danosas no ambiente digital, sobretudo considerando a velocidade com que essas ações se propagam nas redes sociais. Como aponta a literatura especializada, o Direito Penal não deve ser visto como solução única para problemas complexos como os crimes digitais, sob pena de produzir uma ilusão de controle jurídico sobre fenômenos que exigem ação coordenada e multidisciplinar [4].

Outra questão que merece atenção diz respeito aos riscos de instrumentalização da norma para fins de censura ou repressão à liberdade de expressão, especialmente no contexto eleitoral. Parlamentares da oposição alertaram para a possibilidade de uso político da tipificação penal, como no caso da deputada Bia Kicis (PL-DF), que argumentou que o projeto poderia servir como "lobo em pele de cordeiro", ao transferir poderes excessivos a autoridades eleitorais para perseguir adversários políticos. Tais preocupações não devem ser ignoradas, uma vez que a regulação de tecnologias emergentes deve sempre buscar um equilíbrio entre a proteção de direitos e as garantias fundamentais de liberdade, transparência e devido processo legal.

Dentro dessa lógica, a responsabilização de agressores que utilizam *deepfakes* para atingir crianças e adolescentes não pode se basear unicamente na capacidade individual das vítimas de acionar o sistema de justiça, pois essa dinâmica afronta o princípio da proteção integral consagrado pela Constituição Federal de 1988. Conforme já advertia Ferdinand Lassalle, uma constituição — e, por extensão, qualquer norma jurídica — é apenas uma "folha de papel" senão refletir e for sustentada pelas forças sociais e institucionais reais que atuam na sociedade [5]. Assim, é dever do Estado, da sociedade e da família assegurar o pleno exercício dos direitos infantojuvenis, voltado ao enfrentamento estrutural das novas formas de violência digital.

---

[4] GRECO, Rogério. *Direito Penal: parte geral*. 21. ed. Rio de Janeiro: Impetus, 2022.
[5] LASSALLE, Ferdinand. *A Essência da Constituição*. 2. ed. São Paulo: Editora Revista dos Tribunais, 2001.

O PL 3821/2024, portanto, representa um avanço ao tipificar condutas específicas envolvendo o uso de inteligência artificial para manipulação de imagens com conteúdo sexual. Contudo, sua eficácia dependerá da existência de um ecossistema jurídico e institucional robusto, capaz de assegurar a aplicação prática da norma e acompanhar as constantes transformações tecnológicas, pois, na ausência dessas condições, há o risco de se produzir uma legislação meramente simbólica, isolada da realidade das vítimas.

Para além do campo jurídico e político, é imprescindível adotar uma perspectiva sociotécnica que leve em consideração o papel das plataformas digitais na contenção da disseminação de conteúdos manipulados. Nessa linha, destaca-se a recente iniciativa da deputada federal Erika Hilton, que acionou o Ministério Público Federal para apurar a atuação do TikTok por promover sistematicamente, por meio de seu algoritmo, conteúdos misóginos e discriminatórios a adolescentes[6]. Tal iniciativa evidencia não apenas a urgência de responsabilizar as plataformas digitais pela ausência de critérios éticos em seus mecanismos de moderação, mas também revela uma afronta direta aos direitos assegurados pelo Estatuto da Criança e do Adolescente, comprometendo a formação psíquica e social de sujeitos em condição peculiar de desenvolvimento.

A partir desse cenário, torna-se evidente que a lógica operacional das plataformas digitais — orientada prioritariamente por métricas de engajamento e maximização de lucros — não pode permanecer dissociada de deveres normativos claros e eficazes, sendo imprescindível que a regulamentação incorpore obrigações tangíveis de transparência algorítmica, protocolos céleres de responsabilização e sanções compatíveis com a gravidade das violações, sob o risco de se consolidar um ambiente digital permissivo à exploração reiterada da vulnerabilidade infantojuvenil.

Assim sendo, torna-se fundamental o envolvimento de órgãos como o Ministério Público, os Conselhos Tutelares e as Defensorias Públicas, que atuam diretamente na defesa dos direitos de crianças e adolescentes. Por consectário lógico, essas instituições devem ser fortalecidas e capacitadas para enfrentar as

[6]CARTACAPITAL. Erika Hilton vai ao MPF contra TikTok por sugerir a adolescentes posts de ódio contra mulheres. 15 maio 2024. Disponível em: https://www.cartacapital.com.br/politica/erika-hilton-vai-ao-mpf-contra-tiktok-por-sugerir-a-adolescentes-posts-de-odio-contra-mulheres/. Acesso em: 10 de maio de 2025.

especificidades dos crimes digitais, inclusive por meio de formação técnica continuada sobre inteligência artificial. Além disso, é urgente que escolas e famílias sejam incorporadas a esse processo de conscientização, de modo que possam identificar e intervir diante de situações de exposição indevida de jovens no ambiente virtual.

Ao tratarmos do impacto das deepfakes sobre crianças e adolescentes, não se pode reduzir a discussão apenas ao campo criminal, trata-se de uma problemática complexa e de natureza multidimensional, que exige ser compreendida a partir de uma perspectiva de direitos humanos, capaz de articular prevenção, responsabilização, reparação e promoção de um ambiente digital que seja verdadeiramente seguro, inclusivo e respeitoso das especificidades do desenvolvimento infantojuvenil.

Sob essa ótica, longe de pretender esgotar a complexidade do tema, o presente artigo propõe-se a analisar o Projeto de Lei 3821/2024 como ponto de partida para refletir sobre os desafios normativos, institucionais e sociais que se impõem diante das novas formas de violência digital — sem, no entanto, ignorar a necessidade de estratégias mais amplas, articuladas e contínuas de proteção integral.

DEEPFAKES COMO AMEAÇA GLOBAL E MULTIDIMENSIONAL

Lauro Accioly Filho

Como citar este artigo?

ACCIOLY FILHO, Lauro. Deepfakes como ameaça global e multidimensional. In: SOUZA, Gustavo; SILVA, Tarcizio (org.). **Enfrentando Deepfakes.** Desvelar: Brasília, 2025. Disponível em: desvelar.org/enfrentando-deepfakes.

As deepfakes são falsificações realistas de fotos, áudios e vídeos geradas por aplicações de Inteligência Artificial (IA). Elas representam um risco global pela facilidade e qualidade com que podem ser produzidas. Além de serem criadas de forma rápida e realista, preocupa ainda mais o fato de poderem ser geradas a baixo custo, com softwares amplamente disponíveis e, muitas vezes, gratuitos.[1]

No cenário global, a preocupação vai muito além de casos de difamação ou desinformação direcionados a indivíduos ou instituições; trata-se de uma ameaça potencial à segurança internacional e aos conflitos militares. Embora o engano e a manipulação de informações sempre tenham feito parte das estratégias de guerra, jamais se imaginou que qualquer pessoa pudesse criar áudios, vídeos ou textos altamente realistas simulando líderes políticos ou militares do lado adversário. Isso abre caminho para a disseminação instantânea de informações falsas, capazes de desestabilizar sociedades, desorientar populações civis e até influenciar decisões estratégicas em conflitos armados.[2] Um exemplo marcante[3]foi o caso do presidente ucraniano Volodymyr Zelensky, alvo de um vídeo deepfake divulgado nas redes sociais no qual ele parecia instruir os soldados ucranianos a se renderem às forças russas.[4]

Essa preocupação é destacada em documento[5] do Departamento de Segurança Interna dos EUA, que classifica as deepfakes como parte da chamada mídia sintética — conteúdos criados ou modificados por meios digitais ou artificiais, como pessoas geradas por computador, ou manipulados com tecnologia, seja digital ou analógica. Já os "*cheapfakes*" (falsificações baratas) usam técnicas digitais simples para distorcer a percepção de eventos. Além disso, há um relatório[6] do *Congressional*

---

[1] ACCIOLY FILHO, Lauro; NOBRE, Fábio; COSTA, Saulo. Fakes News realística: os desafios sociopolíticos na era da inteligência artificial. In: Osorio, Aline et al. Desinformação, o mal do século: o futuro da democracia, inteligência artificial e direitos fundamentais. Brasília: Supremo Tribunal Federal, 2025. Disponível em: https://www.stf.jus.br/arquivo/cms/campanha/anexo/combate/ebookdesinformacao_VOL2.pdf

[2] AJDER, Henry *et al*. The state of deepfakes: Landscape, threats, and impact. Amsterdam: Deeptrace, v. 27, 2019. Disponível em: https://regmedia.co.uk/2019/10/08/deepfake_report.pdf.

[3]Discussão apresentada no site da Reuters: https://www.reuters.com/world/europe/deepfake-footage-purports-show-ukrainian-president-capitulating-2022-03-16/.

[4] BUSCH, Ella; WARE, Jacob; BRIEF, ICCT Policy. The Weaponisation of Deepfakes. Hague, Netherlands: International Center for Counter Terrorism, 2023. Disponível em: https://icct.nl/sites/default/files/2023-12/The%20Weaponisation%20of%20Deepfakes.pdf.

[5]O material encontra-se disponível em: https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf.

[6]O material encontra-se disponível em: https://www.congress.gov/crs-product/IF11333.

Research Service dos EUA que traz as deepfakes como um problema de segurança nacional[7].

Outro documento[8] relevante, o *Cybersecurity Information Sheet*, produzido pela *National Security Agency* (NSA), pelo *Federal Bureau of Investigation* (FBI) e pela *Cybersecurity and Infrastructure Security Agency* (CISA), analisa os riscos associados ao uso de deepfakes, com destaque para fraudes corporativas, chantagem, campanhas de desinformação e manipulação política. Esse relatório apresenta casos reais em que deepfakes foram usados para enganar executivos e empresas, incluindo fraudes com uso de áudio sintético para desviar recursos financeiros. Além disso, identifica os desafios de detecção devido à rápida evolução tecnológica e recomenda medidas como o desenvolvimento de ferramentas de autenticação, capacitação de equipes, maior transparência nas plataformas digitais e cooperação entre governos e setor privado para mitigar os riscos.

Neste contexto, destaca-se o papel central da **Guerra de Informação** — um cenário marcado por disputas estratégicas que buscam obter vantagem competitiva por meio de esforços ofensivos e defensivos relacionados ao uso da informação. Termos como guerra híbrida e guerra de zona cinzenta são frequentemente utilizados como conceitos correlatos.[9] Em especial, porque o impacto da informação em tais disputas não envolve necessariamente conflito armado ou operações militares diretas, essa dinâmica se insere na esfera da guerra política, entendida como o uso de meios políticos para compelir o oponente a atender à própria vontade. Diferentemente da guerra cibernética, a guerra de informação, embora amplamente disseminada na internet, não se limita ao domínio digital, abrangendo também dimensões sociais, políticas e cognitivas.[10]

---

[7] O material pode ser acessado em: https://www.congress.gov/crs-product/R45178.

[8] O material encontra-se disponível em: https://media.defense.gov/2023/Sep/12/2003298925/-1/-1/0/CSI-DEEPFAKE-THREATS.PDF.

[9] WHYTE, Christopher; THRALL, A. Trevor; MAZANEC, Brian M. (Ed.). Information warfare in the age of cyber conflict. London & New York: Routledge, 2021.

[10] JENSEN, Benjamin; RAMJEE, Divya. Beyond Bullets and Bombs: The Rising Tide of Information War in International Affairs. Washington DC: Center for Strategic & International Studies (CSIS), 2023. Disponível em: https://www.csis.org/analysis/beyond-bullets-and-bombs-rising-tide-information-war-international-affairs.

O governo canadense também vem concentrando esforços para lidar com essa questão. Recentemente, lançou um relatório[11] sobre a evolução da desinformação e os desafios trazidos pela iminência das deepfakes, elaborado pelo Serviço Canadense de Inteligência de Segurança. O documento destaca exemplos de ferramentas de deepfakes comumente utilizadas, como a troca de rosto, sincronização labial, clonagem de voz e, principalmente, os deepfakes baseados em GAN — os mais avançados e preocupantes. Esses últimos utilizam redes adversariais generativas (GANs) para criar imagens, vídeos ou áudios extremamente realistas, exigindo grande poder computacional e conhecimento técnico para sua produção. No mesmo relatório, são apontadas medidas essenciais, como o investimento em tecnologias de ponta para a detecção de deepfakes, capazes de desmascarar impostores digitais e expor intenções maliciosas, além do fortalecimento de marcos legais e da promoção da alfabetização midiática.

Essas deepfakes baseadas em GANs (redes adversariais generativas) funcionam a partir da interação entre dois algoritmos: o gerador e o discriminador. O gerador cria imagens ou áudios falsos a partir de dados reais (por exemplo, rostos ou vozes), enquanto o discriminador tenta identificar se aquele conteúdo é falso ou verdadeiro. À medida que competem entre si, ambos vão se aperfeiçoando: o gerador aprende a produzir conteúdos cada vez mais realistas, e o discriminador melhora na detecção. Esse processo contínuo resulta na criação de deepfakes altamente convincentes e difíceis de distinguir de conteúdos autênticos.[12]

Nos últimos anos, as deepfakes deixaram de ser curiosidades tecnológicas para se tornarem um tema central no entretenimento e no debate jurídico, trazendo avanços criativos, mas também grandes desafios em propriedade intelectual e direitos de imagem. Em Nova Iorque, desde 2020,[13]é proibido usar réplicas digitais de artistas falecidos por 40 anos após sua morte. Neste cenário, os desafios contratuais que emergem incluem definir até onde vai o direito sobre voz e imagem digitalizadas e como garantir consentimento válido. Além disso, surgem dilemas morais, como o direito de artistas vetarem usos com os quais não concordam. Essas questões afetam

[11]O material encontra-se disponível em: https://www.canada.ca/en/security-intelligence-service/corporate/publications/the-evolution-of-disinformation-a-deepfake-future.html.
[12] CHESNEY, Robert; CITRON, Danielle. Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. Foreign Affairs, v. 98, p. 147, 2019.
[13] A lei pode ser lida no site oficial: https://www.nysenate.gov/legislation/bills/2019/S5959.

não só celebridades, mas também atores, influenciadores e *gamers*, em um cenário que pode reduzir empregos e transformar imagem e voz em mercadorias digitais. O futuro exigirá não apenas inovação tecnológica, mas regras claras para proteger direitos e manter a confiança do público.[14]

Na Indonésia, profissionais de comunicação veem a IA generativa como uma ferramenta promissora para aumentar produtividade e criar conteúdo, mas enfrentam desafios ligados à infraestrutura tecnológica, falta de regulamentação e necessidade de capacitação. Há preocupação com o risco de vieses culturais, já que muitas ferramentas foram treinadas com dados ocidentais, o que pode gerar conteúdos inadequados ao contexto local.[15]

Outra ramificação das externalidades negativas das deepfakes é a criação de materiais sexuais explícitos sem consentimento, os chamados deepnudes, que vêm se tornando alarmantemente comuns. Sites especializados nesse tipo de conteúdo recebem cerca de 17 milhões de acessos por mês, assim como aplicativos que "nudificam" fotos comuns de mulheres e meninas também se espalham rapidamente. Além de reforçar desigualdades de gênero, o fenômeno envolve poder, controle e humilhação, espalhando-se em comunidades online que incentivam a prática. Países como o Reino Unido criminalizam apenas a distribuição, mas cresce a pressão para incluir também a produção como crime, o que ajudaria a conter a prática na origem. Isso ampliaria não só a resposta penal, mas também a possibilidade de ações civis para remoção de conteúdos e indenizações, além de pressionar plataformas, buscadores e aplicativos a interromper o suporte a esse ecossistema nocivo.[16]

Diante desse cenário, em 28 de abril, a Câmara dos Representantes aprovou a Lei Take It Down[17], a primeira legislação relevante nos EUA para combater danos causados pela IA. O projeto, já aprovado pelo Senado e aguardando a assinatura do

---

[14] CURREN, Alexandra. Digital Replicas: Harm Caused by Actor's Digital Twins and Hope Provided by the Right of Publicity. Texas Lar Review, v. 102, 2025. Disponível em: https://texaslawreview.org/digital-replicas-harm-caused-by-actors-digital-twins-and-hope-provided-by-the-right-of-publicity/.

[15] HENDRAWAN, Daniel *et al*. Juridical Review of Copyright Infringement in the Use of Deepfakes in the Creative Industry in Indonesia. International Journal of Cyber Criminology, v. 17, n. 1, p. 211-222, 2023.

[16] MCGLYNN, Clare. Deepfake porn: why we need to make it a crime to create it, not just share it. The Conversation. Disponível em: https://theconversation.com/deepfake-porn-why-we-need-to-make-it-a-crime-to-create-it-not-just-share-it-227177.

[17] A lei encontra-se disponível no site oficial do Congresso dos EUA: https://www.congress.gov/bill/119th-congress/senate-bill/146.

então presidente Trump, criminaliza a pornografia deepfake não consensual e obriga as plataformas a removerem esse conteúdo em até 48 horas após serem notificadas. A medida busca conter o crescimento alarmante de imagens ilícitas criadas por IA, impulsionado pelo avanço das ferramentas tecnológicas nos últimos anos. Apesar do amplo apoio bipartidário e da aprovação tranquila nas duas casas, a lei enfrenta críticas: especialistas apontam falhas na redação e alertam que ela pode ser usada de forma abusiva, permitindo que quase qualquer conteúdo seja denunciado como ilícito e removido da internet.[18]

Em vista disso, as deepfakes representam um desafio que vai muito além da desinformação em períodos eleitorais: elas têm impactos amplos e profundos em áreas como propriedade intelectual, privacidade, segurança e direitos individuais. No campo da propriedade intelectual, surgem questões complexas sobre o uso não autorizado de imagem, voz e identidade, especialmente quando contratos tradicionais não preveem o uso dessas características em conteúdos gerados por IA — colocando em risco oportunidades de trabalho e abrindo espaço para um mercado desregulado de “identidades digitais”.

Portanto, o cenário atual, marcado pelo mau uso da IA, reforça a urgência de medidas mais eficazes e de uma participação multissetorial — envolvendo governos, empresas de tecnologia, setor jurídico, organizações de direitos civis e sociedade — para compreender os desafios e criar marcos legais robustos. As iniciativas recentes, como a lei Take It Down nos EUA e os debates no Reino Unido, evidenciam avanços, mas também mostram limitações, já que muitas legislações punem apenas quem compartilha conteúdos ilícitos, ignorando a raiz do problema: a criação. Na imagem abaixo, é possível observar como a deepfake desempenha um problema transversal e preocupante para diversos temas.

[18] CHOW, Andrew R. Inside the First Major U.S. Bill Tackling AI Harms - and Deepfake Abuse. Time Magazine. Disponível em: https://time.com/7277746/ai-deepfakes-take-it-down-act-2025/.

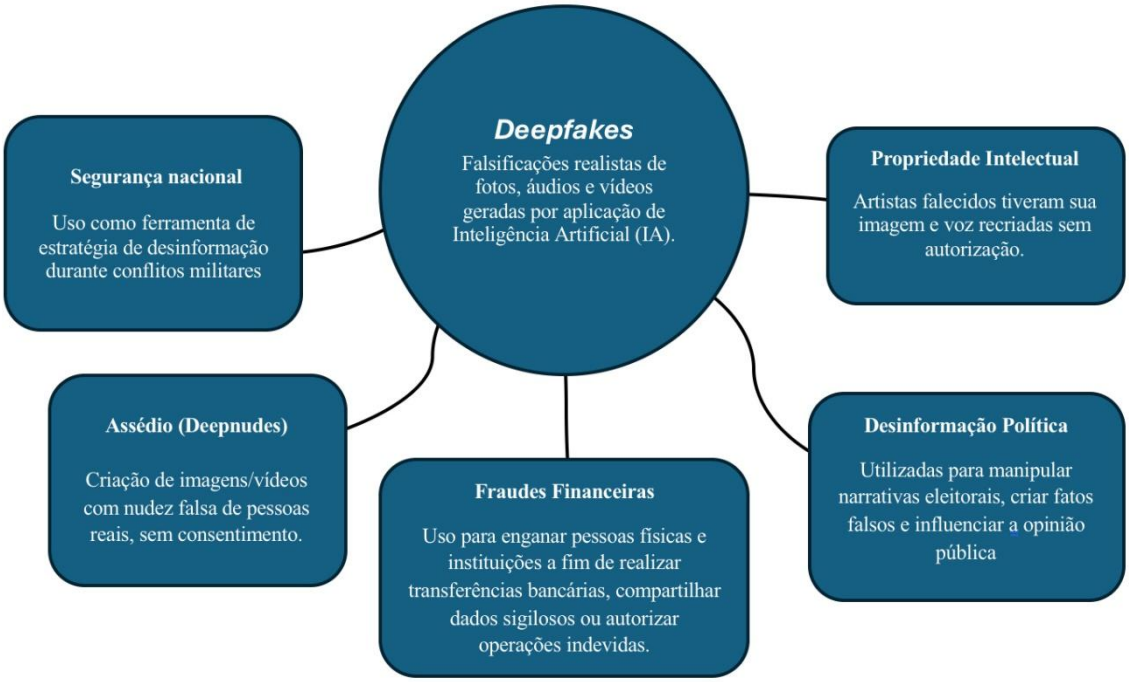


Figura 01. Principais usos e riscos associados às deepfakes. Elaborado pelo próprio autor

No cenário brasileiro, é observável uma crescente articulação entre o Supremo Tribunal Federal e organizações da sociedade civil em torno dos desafios impostos pelas tecnologias emergentes. Iniciativas como o Guia Ilustrado contra Deepfakes,[19]desenvolvido em parceria com o Data Privacy Brasil, exemplificam esforços de conscientização e alfabetização digital. Nesse mesmo sentido, o Programa de Combate a Desinformação - em operação desde 2021 - tem se destacado pela produção de conhecimento crítico, com a publicação de dois volumes de livros acadêmicos, sendo o mais recente, lançado em 2025, dedicado à relação entre inteligência artificial e desinformação. O programa também tem produzido materiais didáticos, como a cartilha Desinformação: Um das Dimensões da Violência de Gênero,[20]que evidencia os impactos diferenciados da desinformação sobre grupos vulnerabilizados.

Além disso, é fundamental entender que o fenômeno das deepfakes também constitui uma questão de segurança nacional. O uso dessas tecnologias para golpes,

[19] O guia está disponível no site do Supremo Tribunal Federal (STF): https://noticias.stf.jus.br/postsnoticias/supremo-lanca-guia-ilustrado-contra-as-deepfakes/.
[20] A cartilha está disponível no site do Supremo Tribunal Federal (STF): https://noticias.stf.jus.br/postsnoticias/stf-lanca-cartilha-sobre-desinformacao-e-violencia-de-genero-na-internet/.

fraudes financeiras, chantagens e campanhas de desinformação ameaça não apenas indivíduos e empresas, mas também a estabilidade das instituições democráticas e a confiança social. Enfrentar as deepfakes, portanto, exige uma abordagem coordenada, que vá além das soluções tecnológicas e legais pontuais, e que reconheça a profundidade do problema no centro das discussões sobre inovação, ética e segurança global.

RACISMO RECREATIVO EM TEMPOS DE DEEPFAKE

Glenda Dantas

Como citar este artigo?

DANTAS, Glenda. Racismo recreativo em tempos de deepfake. SOUZA, Gustavo; SILVA, Tarcizio (org.). Deepfakes. Desvelar: Brasília, 2025. Disponível em: desvelar.org/enfrentando-deepfakes.

O racismo sempre soube se reinventar. Se antes encontrava abrigo nas novelas e nos programas de humor, hoje ocupa com desenvoltura o feed das redes sociais. As mesmas imagens que durante décadas sustentaram o imaginário coletivo sobre o lugar do povo negro continuam circulando, mas agora em formatos de memes, piadas, emojis “engraçadinhos” e memes compartilháveis. O racismo recreativo, termo cunhado por Adilson Moreira[1] para descrever a prática de desumanizar pessoas negras sob o pretexto de fazer humor, não é uma novidade. Ele apenas ganhou novos cenários e ferramentas.

Tião Macalé, o feio; Vera Verão, a bicha preta; Adelaide, a desvairada; Mussum, o bêbado,[2] esses personagens, projetados pela televisão brasileira por anos, foram moldados para entreter às custas da redução da negritude a estereótipos cômicos, grotescos, hipersexualizados ou subalternos. Não é por acaso que continuam sendo figuras populares na cultura dos memes. A diferença é que, agora, esse tipo de humor está a um clique de distância, viraliza em minutos e ganha camadas tecnológicas cada vez mais sofisticadas.

Um estudo recente[3] revelou que, mesmo quando disfarçado de piada, o racismo segue operando com a mesma brutalidade simbólica nas plataformas digitais. A linguagem visual das redes, que explora os GIFs, stickers, montagens, memes, emojis e filtros, muitas vezes camufla ataques, tornando-os menos detectáveis, mas igualmente danosos.

É nesse contexto que surge o deepfake: uma tecnologia que, longe de ser apenas uma inovação curiosa, tem potencial para amplificar e tornar ainda mais complexas as manifestações de racismo recreativo. Ao permitir a criação de vídeos falsos com aparência realista, o deepfake inaugura um novo ciclo de exploração do humor como linguagem política e estética na cultura digital.

O meme marca a cultura digital

[1] O autor lançou em 2019 o livro “Racismo recreativo”. Disponível em: https://bibliotecadigital.mdh.gov.br/jspui/handle/192/10392
[2] No livro “Racismo recreativo” o autor utiliza essas figuras caricatas como exemplos do racismo recreativo na televisão brasileira.
[3] Do Meme ao Ódio: Como o Racismo se Manifesta nas Plataformas Digitais. De autoria do Aláfia Lab e LABHDUFBA. 2025

O meme é uma das expressões mais intrínsecas do digital. Sua capacidade de viralização está diretamente ligada ao fato de que é um fenômeno que se alimenta de acontecimentos cotidianos, traduzindo-os de forma rápida, criativa e acessível. Ao tocar em vivências compartilhadas e emoções comuns, como indignação, riso, surpresa ou vergonha, o meme se espalha como um reflexo coletivo da realidade.

Os deepfakes, nesse contexto, adicionam aos memes um potencial ampliado de verossimilhança e impacto simbólico, diferentemente das montagens tradicionais. Se por um lado esses vídeos podem funcionar como sátiras sofisticadas e críticas sociais, sua aparência realista pode gerar confusão, reforçar estereótipos ou até ser instrumentalizada para desinformação.

Um episódio emblemático evidencia a força simbólica e comunicacional dos memes, assim como os seus riscos. Donald Trump não apenas compartilhou[4], como também elogiou e sugeriu alterações a um vídeo-meme da página C3PMeme, especializada em conteúdos que exaltam sua figura pública por meio de sátiras ofensivas, discursos extremistas e que se auto-intitula a "máquina de guerra online de Trump".

Ao navegar pela página do C3PMeme no YouTube, é possível encontrar com facilidade vídeos de Kamala Harris e Michelle Obama, ambas mulheres negras e figuras centrais do Partido Democrata dos Estados Unidos, editados com tecnologias como deepfake para reforçar estereótipos racistas, misóginos e xenofóbicos. Com vídeos que ultrapassam milhões de visualizações, a página instrumentaliza o humor para atacar adversários e reforçar o culto à personalidade de Trump, construindo uma estética digital que camufla a violência simbólica sob a roupagem da piada.

Um dos exemplos mais explícitos e recentes é o vídeo "Is that Big Mike?"[5], que utiliza técnicas de deepfake para representar Michelle Obama como um homem negro, com cabelo black power, vestindo roupas masculinas, em um suposto cenário de noticiário da MSNBC. A legenda na tela diz "BIG MIKE", apelido derivado de uma

[4] Donald Trump encontrou nos *memes* uma máquina de propaganda *online* https://www.publico.pt/2023/12/30/mundo/noticia/donald-trump-encontrou-memes-maquina-propaganda-online-2074349

[5] https://www.youtube.com/watch?v=eIWhbUy0LgA

teoria da conspiração transfóbica e racista que alega que a advogada e ex-primeira-dama seria, na verdade, um homem trans. O vídeo apresenta ainda gráficos e design inspirados em telejornais reais, como o selo da MSNBC e o uso de sobreposições gráficas.

Ao saudar abertamente esse tipo de produção, Trump não só endossa essas narrativas conspiratórias e ofensivas, como também valida a mobilização da desinformação como instrumento político, utilizando o riso como escudo para mensagens de ódio e intolerância.

Figura 1: Captura de tela do canal no YouTube do C3PMeme, em 12/05/2025.

A página do M3PMeme evidencia como os memes podem ser mobilizados como armas de propaganda afetiva e ideológica, funcionando como instrumentos de polarização, manipulação emocional e disseminação de ódio, sobretudo quando aliados a tecnologias como deepfakes. Os usos supostamente humorísticos do deepfake se inserem na lógica do racismo recreativo e das dinâmicas de opressão simbólica, na medida em que podem reafirmar hierarquias raciais e de gênero sob o disfarce da “brincadeira”.

No Brasil, é comum a circulação de memes racistas envolvendo jogadores de futebol. Com o aumento da visibilidade de Vinicius Jr. tanto como atleta de destaque quanto como símbolo na luta contra o racismo, especialmente após os inúmeros episódios de violência racial sofridos durante partidas na Europa, sua imagem passou a ser amplamente explorada nas redes sociais, inclusive de forma ofensiva.

Um caso emblemático ocorreu em 2024, quando um vídeo manipulado por inteligência artificial mostrava Vinicius Jr. supostamente dizendo que “roubaram na cara dura” o resultado da Bola de Ouro daquele ano. A adulteração consistia na substituição do áudio original, no qual o jogador comentava sua atuação em campo, por uma dublagem falsa. Uma agência de checagem verificou[6] que não há qualquer registro de que ele tenha feito declaração semelhante à apresentada no vídeo.

Figura 2: Montagem feita por Aos Fatos mostra de um lado uma captura de tela da entrevista que o jogador concedeu em maio de 2021 e que foi utilizada como plano de fundo para a criação do deepfake (Reprodução/AosFatos).

Na época, a peça enganosa acumulava mais de 1,1 milhão de visualizações no TikTok e, ainda hoje, uma busca rápida em plataformas como Instagram e TikTok revela sua circulação em diversas páginas de memes brasileiras.

[6] Vídeo em que Vini Jr. diz que ‘roubaram’ Bola de Ouro foi criado por IA
https://www.aosfatos.org/noticias/video-inteligencia-artificial-vini-jr-roubaram-bola-de-ouro/

As redes sociais de Vinicius Jr. vêm sendo objeto constante de monitoramento por parte do Observatório do Racismo nas Redes[7], dado o volume de ataques que recebe. O relatório mais recente[8] apontou que, nos comentários de suas postagens, é recorrente o uso de termos como “mono” e “macaco”, além de emojis e GIFs de macacos (🐵🙈🐵🐒🦍🤦), empregados de forma pejorativa e desumanizante.

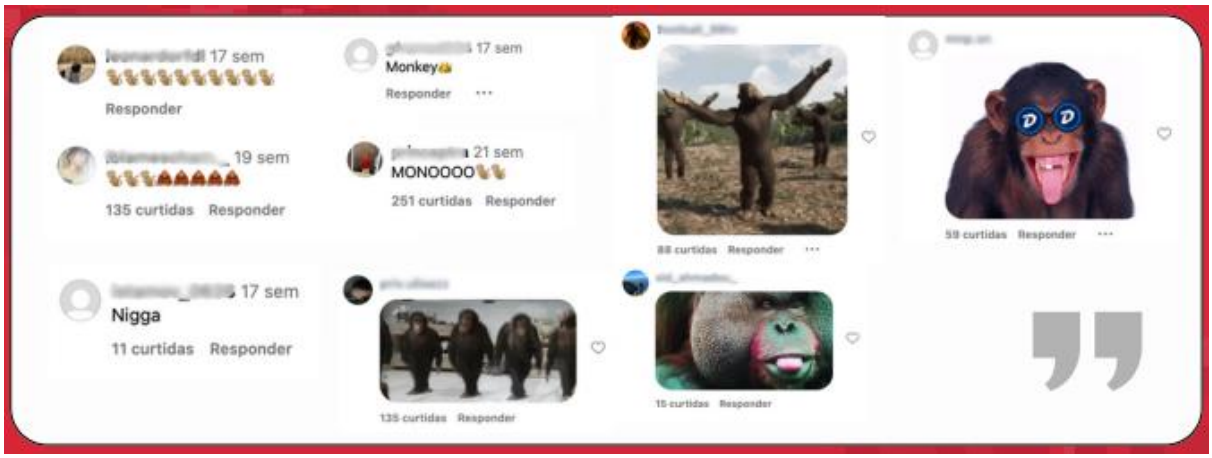


Figura 3: Captura de tela do relatório “Do meme ao ódio: como o racismo se manifesta nas plataformas digitais”, com exemplos de comentários racistas no Instagram

O conceito de racismo recreativo nos convida a compreender que o humor, longe de ser neutro ou inocente, pode funcionar como instrumento de dominação racial. Para Moreira[9], essa forma de racismo deve ser entendida como um projeto político e cultural, que busca perpetuar relações assimétricas de poder entre pessoas brancas e negras por meio da naturalização da violência simbólica.

Ao transformar estereótipos em piadas e agressões em "brincadeiras", o racismo recreativo oculta sua hostilidade sob a aparência do riso, tornando socialmente aceitável e até desejável a humilhação de grupos racializados. Trata-se, portanto, de uma estratégia de controle sofisticada, pois promove a exclusão sob a aparência de leveza e espontaneidade, dificultando a crítica e a responsabilização de seus agentes.

Embora o conceito tenha sido cunhado a partir do contexto sócio-histórico brasileiro, e um dos exemplos analisados neste texto não se refira ao Brasil, não é

[7] O observatório de racismo nas redes é um projeto do Aláfia Lab que busca ampliar a capacidade de monitoramento e combate a conteúdos racistas publicados em redes sociais.
[8] https://alafialab.org/do-meme-ao-odio-como-o-racismo-se-manifesta-nas-plataformas-digitais/
[9] MOREIRA, Adilson. *Racismo recreativo*. São Paulo: Sueli Carneiro; Pólen, 2019. P. 95

difícil traçar paralelos entre nossa realidade e a de outros países da diáspora africana. Apesar das especificidades locais, esses contextos compartilham experiências comuns de violências imagéticas e discursivas produzidas pelo racismo, que operam de forma semelhante na manutenção de hierarquias e exclusões sociorraciais.

A camuflagem dos discursos racistas nas redes sociais

Os principais achados do estudo *"Do meme ao ódio: como o racismo se manifesta nas plataformas digitais"* revelam o uso recorrente e aparentemente banal de memes, emojis e GIFs como veículos de propagação de discursos racistas. A análise também destaca como essas manifestações variam de acordo com a plataforma em que circulam, propondo uma distinção entre redes sociais de superfície (como Facebook, TikTok e Instagram), amplamente acessíveis e visíveis e redes sociais subterrâneas (como grupos fechados no Telegram e no Discord), menos reguladas e mais opacas.

Nesses espaços subterrâneos, os discursos tendem a ser abertamente racistas, misóginos e xenofóbicos, favorecendo a criação de comunidades de ódio com pouca ou nenhuma moderação. Já nas redes de superfície, o racismo se expressa de forma mais sutil, evitando termos explícitos e recorrendo a eufemismos, emojis e comentários opinativos de cunho discriminatório que frequentemente escapam dos filtros das plataformas.

É uma dinâmica que revela não apenas a percepção generalizada de que certos discursos racistas, quando veiculados por meio de humor, memes e símbolos aparentemente inofensivos, são mais socialmente aceitáveis, mas também a inércia institucional diante dessas manifestações. Tanto as plataformas digitais quanto o sistema jurídico demonstram dificuldade ou falta de interesse em reconhecer e tipificar esse tipo de violência simbólica como crime. No Brasil, por exemplo, foi apenas recentemente que a justiça passou a tipificar o racismo recreativo como crime, com a promulgação da Lei nº 14.532/2023.

Essa negligência está diretamente relacionada à dimensão institucional do racismo recreativo[10], onde piadas e "brincadeiras" discriminatórias contra minorias

[10] MOREIRA, Adilson. *Racismo recreativo*. São Paulo: Sueli Carneiro; Pólen, 2019. P. 98 e 99.

raciais são frequentemente desconsideradas como ofensivas ou criminosas porque as instituições que deveriam responsabilizar têm uma natureza estruturalmente racista.

Enquanto isso, os impactos sobre os grupos racializados são profundos e muitas vezes imensuráveis, afetando diretamente sua confiança social, suas relações de convivência, sua saúde mental e emocional, entre outras dimensões fundamentais para uma vida digna e segura.

As plataformas digitais, nesse contexto, desempenham um papel central na forma como discursos racistas são produzidos, disseminados e normalizados no ambiente online. Embora afirmem investir em ferramentas de moderação de conteúdo, essas ações se mostram frequentemente insuficientes, sobretudo quando se trata de conteúdos disfarçados de humor, opinião ou crítica social. A lógica algorítmica dessas plataformas, centrada no engajamento e na viralização, acaba favorecendo justamente os conteúdos polêmicos e sensacionalistas, mesmo que carreguem discursos de ódio.

É um cenário em que a ausência de critérios claros e efetivos de moderação, somada à negligência das empresas em ouvir vozes racializadas e especialistas em direitos humanos, contribui para a manutenção de um ecossistema extremamente permissivo à violência simbólica.

Com o avanço das tecnologias de manipulação de imagem, como os deepfakes, a sofisticação das violências raciais no ambiente digital atinge um novo patamar, tornando-se ainda mais difícil de detectar, moderar e punir. A velocidade com que essas violências se adaptam às novas ferramentas digitais supera em muito o ritmo das respostas institucionais, seja na moderação das plataformas, seja na atuação legislativa e jurídica, escancarando a urgência da regulação das tecnologias emergentes e a proteção dos direitos de grupos historicamente marginalizados.

O CORPO PRETO

Gustavo Souza

Como citar este artigo?

SOUZA, Gustavo. O corpo preto. In: SOUZA, Gustavo; SILVA, Tarcizio (org.). **Enfrentando Deepfakes.** Desvelar: Brasília, 2025. Disponível em: desvelar.org/enfrentando-deepfakes.

Em uma aula de Geografia, um professor aponta para um corpo de terra, pintado na cor preta, onde todos deveriam se concentrar. Com a introdução da Inteligência Artificial Generativa nas diversas instituições, inclusive de ensino, no que deveríamos estar prestando a devida atenção?

Figura 1: The Geography Lesson or "The Black Spot", pintura de Albert Bettannier, feita em 1887.

Convenções e preconceitos desconhecidos na educação

Existem convenções para a construção de mapas. Uma delas é usar azul apenas para água. Nesta pintura acima, de Albert Bettannier em 1887, um professor aponta para a França no mapa, certificando-se de que toda a sala de aula, composta por crianças brancas do sexo masculino, se concentre na lição por trás do corpo preto. Como as convenções vêm e vão, uma extensão de terra cercada por azul é uma ilha, e o preto geralmente é aplicado para linhas em mapas. Optar por representar uma área inteira em preto é uma abordagem incomum. Mas a atenção dada aos corpos pretos, além dos mapas, ainda vale a pena ser analisada.

Pesquisas documentam há muito tempo o escrutínio desproporcional que corpos pretos enfrentam na segurança pública, onde são mais frequentemente submetidos a abordagens e vigilância policial. No século passado, esforços para

naturalizar o racismo estavam presentes em charges[1] e a mídia em geral, mas agora isso pode ser feito de forma muito realista por meio de deepfakes.

Infelizmente, esse mesmo padrão de escrutínio surge nas escolas, onde estudantes negros enfrentam preconceitos de educadores que moldam suas experiências acadêmicas desde cedo. Estudantes negros são monitorados desproporcionalmente por problemas comportamentais em comparação com seus colegas, mesmo quando não há mau comportamento[2]. Esse escrutínio mais rigoroso pode resultar em mais suspensões e expulsões, reforçando desvantagens sistêmicas que afetam os resultados educacionais.

Com o rápido choque entre Inteligência Artificial Generativa (IAG) e educação, o preconceito racial se manifesta de novas maneiras. Ferramentas de detecção projetadas para identificar conteúdo gerado por inteligência artificial adicionaram outra camada de escrutínio, impactando até mesmo jornalistas[3], e não tão surpreendentemente aos estudantes negros. Um estudo de 2024 revelou que os adolescentes negros têm duas vezes mais probabilidade do que os alunos brancos de relatar que seus professores sinalizaram incorretamente suas tarefas escolares como geradas por IA quando, na verdade, não foram[4]. Essa discrepância sugere que os mesmos preconceitos implícitos que influenciam as expectativas dos professores em relação ao comportamento dos alunos agora estão se estendendo às avaliações do desempenho acadêmico.

---

[1]Tintim no Congo.https://en.wikipedia.org/wiki/Tintin_no_Congo
[2]Preconceito contra estudantes negros na pré-escola.https://www.theguardian.com/world/2016/oct/04/black-students-teachers-implicit-racial-bias-preschool-study
[3]Ferramentas para identificar conteúdo de IA.https://gizmodo.com/ai-detectors-inaccurate-freelance-writers-fired-1851529820
[4]O alvorecer da era da IA.https://www.commonsensemedia.org/sites/default/files/research/report/2024-the-dawn-of-the-ai-era_final-release-for-web.pdf

Inteligência Artificial Generativa nas escolas

Em um conto, o escritor argentino Borges descreve um império tão obcecado por cartografia e detalhes, que acabou criando um mapa perfeito. Tão perfeito que cada ponto do mapa coincidia com o império, abrangendo as cidades ponto por ponto. Mapas exigem alfabetização para que os cidadãos possam interpretar informações, bem como compreender conceitos como escala, fronteiras, colonialidade e seus limites. Caso contrário, pode ser muito difícil entender como o Brasil, o país mais populoso da América do Sul, tem a maior fronteira terrestre com a França, como corretamente apontado pelo Copilot, um chatbot desenvolvido pelo chatGPT para Windows.

Figura 2: Captura de tela feita pelo autor.

Há possibilidades de uso da IAG nas escolas e, como novas ferramentas, elas exigirão alfabetização por meio de treinamento e compreensão crítica. A implementação dessas tecnologias exige conscientização sobre os potenciais impactos, especialmente o aprofundamento da desigualdade, como consequência da indisponibilidade da Internet.

Depois de tantas manchetes preocupantes nos jornais, parte da população se preocupa com a consequências negativas da difusão da inteligência artificial, como vigilância, desinformação, propaganda, criação de deepfakes e outros usos perigosos que levantam muitas preocupações diferentes, desde a proteção de dados e a

segurança das crianças até os esforços de regulamentação[5]. Quando fotos pessoais de crianças estão sendo usadas para treinar modelos de IA sem consentimento[6], é fundamental entender que esses debates estão presentes nas salas de aula, reuniões e corredores das escolas. Em alguns casos, a IA desconectada pode ser a melhor solução[7].

Professores estão preocupados com a forma como os alunos utilizam a IA. Em geral, educadores expressam preocupações sobre a integridade das informações oferecidas pela IAG e as estratégias possivelmente adotadas pelos alunos para evitar o esforço de fazer a lição de casa. As rápidas mudanças nos debates públicos sobre tecnologia levaram os governos a tomar medidas.

No Brasil, o uso de celulares por alunos nas escolas foi recentemente proibido[8]. O Governo do estado de São Paulo enfrentou reação negativa por tentar substituir professores no planejamento de aulas[9]. No entanto, há preocupações em torno da IAG que vão muito além das tarefas de casa e da dinâmica da sala de aula. A interseção da tecnologia com estereótipos de raça e gênero tornou-se um ponto focal nos debates sobre as implicações éticas da IAG.

O debate público se voltou para casos em que alunos utilizaram IAG para criar imagens fabricadas de seus colegas e professores. Em julho de 2024, sete alunos da cidade de Maceió geraram imagens deepfakes de colegas de classe[10], ato que levou a uma ação disciplinar. Poucos meses depois, em setembro de 2024, alunos da cidade

---

[5]Inteligência Artificial: Alfabetização Midiática e Informacional, Direitos Humanos e Liberdade de Expressão.https://iite.unesco.org/publications/artificial-intelligence-media-and-information-literacy-rights-human-and-freedom-of-expression/

[6]Brasil: Fotos pessoais de crianças são usadas indevidamente para alimentar ferramentas de IAhttps://www.hrw.org/news/2024/06/10/fotos-pessoais-de-criancas-brasileiras-utilizam-poder-ferramentas-de-ia

[7] “Hora de rever como se avalia o ensino”, diz presidente de associação de IA na Educação https://lupa.uol.com.br/educacao/2024/07/22/hora-de-rever-como-se-avalia-o-ensino-diz-presidente-de-associacao-de-ia-na-educacao

[8]Brasil restringe uso de celular nas escolas.https://agenciabrasil.ebc.com.br/educacao/noticia/2025-01/brasil-restringe-uso-de-celular-em-escolas-publicas-e-particulares

[9] Governo Tarcísio anuncia uso de IA na produção de aulas no ensino médio e fundamental https://oglobo.globo.com/politica/noticia/2024/04/17/governo-tarcisio-anuncia-uso-de-ia-na-producao-de-aulas-no-ensino-medio-e-fundamental.ghtml

[10] Deepfake: Sete adolescentes são responsabilizados por manipulação de imagens em Maceió https://g1.globo.com/al/alagoas/noticia/2024/07/23/operacao-deepfake-sete-adolescentes-sao-responsabilizados-por-manipulacao-de-imagens-em-maceio.ghtml

de Cuiabá foram expulsos da escola[11] por ativadades semelhantes. A manipulação de imagens através de IAG está sendo utilizada de forma prejudicial, especialmente contra mulheres e grupos marginalizados. Mais recentemente, em fevereiro de 2025, dois estudantes estão sendo investigados por compartilharem deepnudes, que usavam a imagem de uma professora em conteúdo erótico, por meio de IA[12].

A prevalência impressionante de pornografia deepfake amplifica ainda mais essas preocupações. Em 2023, uma estimativa indicou que o conteúdo explícito representava 98% de todos os vídeos deepfakes online[13].Essa realidade inquietante destaca a necessidade urgente de aprimorar a educação midiática. Equipar os alunos com as habilidades necessárias para avaliar criticamente o conteúdo digital e reconhecer a manipulação impulsionada pela IA é essencial para combater a disseminação de materiais nocivos e enganosos.

[11] Alunos são expulsos após usar inteligência artificial para criar nudes falsos de professora e colegas em escola particular de Cuiabá https://g1.globo.com/mt/mato-grosso/noticia/2024/09/25/alunos-sao-expulsos-apos-usar-inteligencia-artificial-para-criar-nudes-falsos-de-professora-e-colegas-na-escola-particular-de-cuiaba.ghtml

[12] Adolescentes são apreendidos suspeitos de usar inteligência artificial para criar nudes de professora. https://g1.globo.com/pb/paraiba/noticia/2025/02/11/adolescentes-sao-apreendidos-suspeitos-de-usar-inteligencia-artificial-para-criar-nudes-de-professora.ghtml

[13]Estados dos Deepfakes em 2023https://www.securityhero.io/state-of-deepfakes/

Um imaginário alternativo

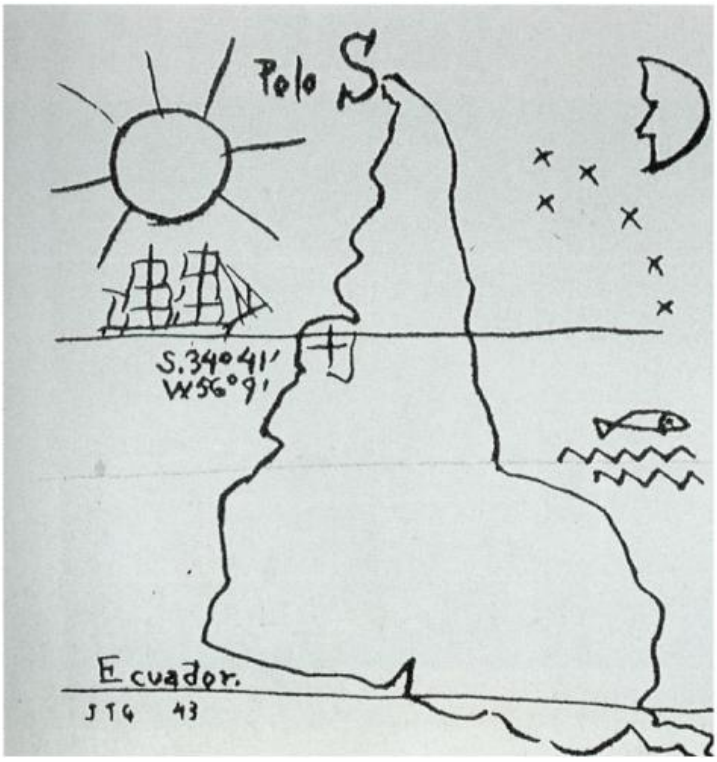


Figura 3: Obra “América Invertida”, que exibe o continente sul-américano a partir de outra ótica, do autor uruguaio Joaquín Torres García.

Enquanto o Sul Global enfrenta a erosão da confiança e da integridade da informação, os desdobramentos da IAG irão demandar novas abordagens locais para a educação midiática. É importante questionar se as escolas oferecerão oportunidades para desenvolvimento do pensamento crítico, permitindo que os alunos diferenciem entre conteúdo legítimo e deepfakes.

Esse processo só será possível se levarmos em conta a diversidade, a equidade e a inclusão. No Brasil, as mulheres negras compõem o maior grupo demográfico. Esse cenário significa que grande parte da população é submetida à violência de gênero e microagressões raciais desde a pré-escola até a vida adulta.

Em abril de 2025, o parlamento brasileiro aprovou e o presidente Lula sancionou a lei 15123[14], que altera o Código Penal e tipifica o crime de violência psicológica contra a mulher sempre que ferramentas de Inteligência Artificial forem utilizadas para produzir deepfakes, manipulando a imagem ou a voz da vítima. A

[14]Lei 15123, sobre deep fakeshttps://www.planalto.gov.br/ccivil_03/_Ato2023-2026/2025/Lei/L15123.htm#art2

legislação penal, somada a A proibição de celulares nas escolas, pode ter um efeito de desaceleração destes impactos. É importante notar que uma abordagem criminal pode ser parte de uma solução para lidar com o uso indevido da Inteligência Artificial Generativa, mas a educação e as escolas devem ser parte essencial dela. Isso deve significar mais do que a proibição de celulares nas escolas.

Como os argumentos para implementação de IA são frequentemente baseados na redução de custos e em uma suposta eficiência, existe a exigência de uma compreensão mais profunda das desigualdades e de como a educação é oferecida. Nesse sentido, iniciativas globais como a Semana da Alfabetização Midiática e Informacional[15], promovida pela Unesco, e a Semana Brasileira de MIL[16]são importantes.

Se corpos negros outrora marcavam territórios disputados em mapas coloniais, hoje eles revelam a necessidade urgente de tomada de ação para conceber imaginários do nosso futuro. Essa transformação não pode ser ditada apenas pelos centros tradicionais de poder. O Sul Global, com seus diversos sistemas de conhecimento e tecnologias ancestrais, deve liderar a criação de novos imaginários — onde a tecnologia sirva à justiça, à inclusão e à emancipação coletiva.

[15]Semana MIL da Unescohttps://www.unesco.org/en/weeks/media-information-literacy-2024/around-world
[16]Semana das Mães no Brasilhttps://www.gov.br/secom/pt-br/assuntos/educacao-midiatica/2asbem/2aSemana

DEEPFAKES E A FABRICAÇÃO DO REAL: A DISTORÇÃO DA REALIDADE COMO FORMA DE CONTROLE POLÍTICO SOCIAL

Bianca Galvão Marques

Como citar este artigo?

MARQUES, Bianca Galvão. Deepfakes e a fabricação do real: a distorção da realidade como forma de controle político social. In: SOUZA, Gustavo; SILVA, Tarcizio (org.). **Enfrentando Deepfakes.** Desvelar: Brasília, 2025. Disponível em: desvelar.org/enfrentando-deepfakes.

Saber o que é realidade e o que é inventado nos dias atuais tem se tornado uma atividade cada vez mais complexa. O fortalecimento de instrumentos capazes de manipular a verdade tem possibilitado a ocupação de espaços que as ferramentas de enfrentamento das mentiras não conseguem alcançar. O avanço da Inteligência Artificial Generativa (IAG) tem auxiliado na corrosão do que é verdadeiro a partir do uso de *deepfakes*.

Apesar de terem sido fortalecidas e aprimoradas pela difusão de IAG, as *deepfakes* não são novas. Em 2020, Giselle Beilguelman já fazia análises sobre as problemáticas envolvendo a criação de imagens através do *machine learning*[1]. Na época, a manipulação de imagens era direcionada, primordialmente, para produção de vídeos pornográficos e troca de rostos de atrizes de vídeos íntimos para de artistas.

É difícil distinguir entre o que é verdadeiro e o que é falso. A facilidade com que hoje se consegue alterar imagens e vozes permite a criação de conteúdos enganosos, capazes de convencer o público de algo que nunca foi dito, mas que foi artificialmente manipulado para parecer real[2].

Durante muito tempo, as *deepfakes* foram usadas sobretudo como ferramenta da misoginia - e ainda são, considerando que, nas eleições de 2024, a Deputada Federal e candidata à prefeitura de São Paulo, Tabata Amaral, teve sua imagem exposta em posições de cunho sexual[3]. No entanto, esse desdobramento da IAG passou a ocupar um novo espaço no cenário social tão preocupante quanto a violência de gênero: a democracia.

---

[1] Cada vez mais sofisticados, “deepfakes” vieram para ficar. Disponível em: https://jornal.usp.br/cultura/cada-vez-mais-sofisticados-deepfakes-vieram-para-ficar/

[2] AUGÉ, Andréa Barros; JACOB, Raphael Rios Chaia. Do uso de deepfakes e outros modelos de inteligência artificial para manipulação de conteúdos em larga escala. In: OSORIO, Aline Rezende Peres; MARANHÃO, Ana Carolina Kalume; LEITE, Ana Gabriela Guerreiro Viola da Silveira; ALVIM, Frederico Franco; ALMEIDA, Paulo; PANIAGO, Paulo; RUSSI, Pedro; JORGE, Thias de Mendonça; DURIGAN, Victor Carnevalli (org.). Desinformação o mal do século: O Futuro da Democracia: Inteligência Artificial e Direitos Fundamentais. Brasília: : Faculdade de Comunicação da Universidade de Brasília, 2024. v. 2, ISBN 978-65-6141-050-2, p. 243. Disponível em: https://www.stf.jus.br/arquivo/cms/campanha/anexo/combate/ebookdesinformacao_VOL2.pdf.

[3] Tabata Amaral aciona a Justiça por crime de injúria eleitoral em deepfakes de cunho sexual. Disponível em: https://www.estadao.com.br/politica/eleicoes-2024-tabata-amaral-psb-justica-inteligencia-artificial-deepfake-cunho-sexual-criadora-conteudo-adulto-nprp/?srsltid=AfmBOoqHC2jL2eDK6YHVs3Ii_-aiLWgeWop62V0CAjy9Otusyftp_wg8

Na tentativa de manter a existência do processo democrático eleitoral brasileiro, o Tribunal Superior Eleitoral tem se esforçado para combater a desinformação e, principalmente, a manipulação de imagens e sons por meio da Inteligência Artificial, de modo a alterar as disposições sobre a propaganda eleitoral realizada por meio da Resolução nº 23.732/2024, que passou a determinar a indicação dos conteúdos que foram gerados e manipulados com uso da Inteligência Artificial.

A disputa eleitoral para presidência do Brasil, ocorrida em 2022, ficou marcada como um momento de fragilidade democrática no país, em razão dos diversos discursos voltados ao questionamento do processo eleitoral e à confiabilidade das urnas eleitorais. Todavia, o discurso voltado a construir um cenário de fragilidade democrático já acontecia desde as eleições de 2018, quando eleitores do ex-Presidente Jair Bolsonaro afirmavam que este havia ganhado as eleições ainda no primeiro turno[4]. Utilizando, principalmente, redes sociais, o discurso fortaleceu movimentos extremistas no país, dando voz e credibilidade para teorias conspiratórias presentes no cotidiano brasileiro, como explicam Augé e Jacob[5], "ao criar um conteúdo convincente, mas totalmente fabricado, os atores mal-intencionados podem enganar o público criando toda a sorte de prejuízos".

Assim, o que era verdadeiro, passou a ser questionado. E o que era inventado passou a ser validado.

O cenário atual, todavia, não usa apenas a manipulação de dados e montagem de discursos fantasiosos como forma de influenciar as massas. O controle político-social ganha nas novas tecnologias e na Inteligência Artificial novas ferramentas de alteração do real, na expectativa de confirmar a imagética de eleitores e de seguidores de ideologias carentes de uma única realidade. Como explica Han[6], "o regime de informação revela traços totalitários", de modo que, a partir disso, também se afasta do que é entendido como realidade. Afinal, se a realidade em que se vive não é aquela

---

[4] É #FAKE que denúncia de fraude em eleição de 2018 comprovou vitória de Bolsonaro no 1º turno. Disponível em: https://g1.globo.com/fato-ou-fake/noticia/2020/10/06/e-fake-que-denuncia-de-fraude-em-eleicao-de-2018-comprovou-vitoria-de-bolsonaro-no-1o-turno.ghtml

[5] AUGÉ, Andréa Barros; JACOB, Raphael Rios Chaia. Ob. Cit., p. 244.

[6] HAN, Byung-Chul. **Infocracia: digitalização e a crise da democracia**. Editora Vozes, 2022, p. 20.

que se espera viver, basta que se crie uma sociedade que acredite na possibilidade de viver em uma realidade inventada.

Em Rondônia, estado onde o ex-Presidente Jair Bolsonaro venceu em 1º e 2º turno nas eleições de 2022[7], divulgou-se uma *deepfake* do candidato e atual Governador, Marcos Rocha, cumprimentando o então candidato e atual presidente da República como forma de afastar o seu eleitorado. Na imagem verdadeira, Marcos Rocha cumprimentava o, à época, presidente da Assembleia Legislativa de Rondônia, Deputado Alex Redano[8].

Figura 1: A captura de tela foi recuperada através da petição inicial do processo 0600417-14.2022.6.22.0000, cuja tramitação se deu no Tribunal Regional Eleitoral de Rondônia.

Nas eleições à Prefeitura de Porto Velho de 2024, com o intuito de afetar a campanha do candidato Léo Moraes - vencedor do pleito -, foi compartilhado um vídeo onde Coronel Chrisóstomo, líder do Partido Liberal de Porto Velho e Deputado Federal, demonstrava um suposto apoio à Léo[9]. No vídeo original, entretanto, Chrisóstomo estava criticando as alianças políticas da candidata do União Brasil, Mariana Carvalho.

A nível nacional, tentou-se fortalecer a narrativa fraude eleitoral caso o atual Presidente fosse eleito desde as eleições, onde houve manipulação da verdade a

[7] Apuração por Estado. Disponível em: http://especiaisg1.globo/ro/rondonia/eleicoes/2022/mapas/apuracao-estado-presidente/2-turno/
[8] TRE-RO - Rp: 0600417-14 .2022.6.22.0000 PORTO VELHO - RO 060041714, Relator.: Carlos Augusto Teles De Negreiros, Data de Julgamento: 15/12/2022, Data de Publicação: PSESS-246, data 14/12/2022
[9] COMBATE ÀS 'FAKE NEWS': Justiça determina novo bloqueio de vídeo montado por militante de Mariana contra Léo. Disponível em: https://expressaorondonia.com.br/combate-as-fake-news-justica-determina-novo-bloqueio-de-video-montado-por-militante-de-mariana-contra-leo/

partir da alteração da voz e imagem da jornalista Renata Vasconcellos, para que os eleitores acreditassem que o candidato Jair Bolsonaro estava à frente nas pesquisas da corrida presidencial. Em 2025, a jornalista novamente foi vítima de *deepfake* para a divulgação de um vídeo que noticiava, de forma falsa, que os Estados Unidos aplicariam sanções ao Brasil caso Bolsonaro fosse preso. Nestes casos, não é possível esperar que uma pessoa seguidora da ideologia da extrema-direita haja de forma racional ao ser atingida por esses vídeos manipulados, haja vista que ela será afetada e estimulada por informações que lhe convém[10].

No entanto, em um cenário digital onde tudo pode ser questionado e entendido como *deepfake,* o que é verdade se torna mentira e o contrário também acontece. Explicam Augé e Jacob que "o maior dano, porém, não vem da manipulação que pode ser promovida pelo material manipulado, mas da própria ressignificação da verdade", isto é, assim como uma imagem real é alterada para caber na realidade conivente através do *deepfake*, também se pode acusar imagens reais de serem versões manipuladas de algo real, de modo a criar uma realidade invertida, assim, nas palavras de Han[11], "notícias se tornam similares a uma narrativa. A distinção entre realidade e ficção desaparece".

Quando se manipula a realidade, o impacto não se dá apenas na imagem e na voz, mas também na confirmação de um discurso criado a partir daquilo que se quer fazer acreditar dentro do cenário político e, consequentemente, no social. A divulgação de informações falsas através de áudios e imagens alteradas objetivando a manipulação das massas se mostra como um meio, aparentemente, eficaz para fazer com que a verdade de um se transforme também na verdade de todos. Afinal, a comunicação política tem se baseado menos na integridade das informações e mais na distorção da realidade. Assim, o controle político-social não se baseia mais na integridade informacional, mas na distorção da realidade.

Na tentativa de combater manipulação da realidade e a desinformação, é importante ter em mente que a geração e disseminação de *deepfakes* acontecem para além do período de campanha eleitoral, da mesma forma que sua criação independe de candidatos a cargos políticos, como a criação do vídeo do Presidente Lula beijando

[10] HAN, Byung-Chul. **Infocracia: digitalização e a crise da democracia**. Editora Vozes, 2022, p. 37.
[11] Ibidem, p. 29.

o Ministro do Supremo Tribunal Federal, Alexandre de Moraes[12]. Apesar de parecer cômico, a *deepfake* fortalece o discurso de perseguição sofrida pelo ex-Presidente Jair Bolsonaro ante a proximidade de Lula e Moraes. Assim, esperar que a distorção da realidade aconteça apenas em anos eleitores se mostra, no mínimo, ingênuo.

Com o intuito de manter campanhas de conscientização acerca dos modos de desinformação, o Tribunal Regional Eleitoral de Rondônia tem intensificado o debate para além do período eleitoral através de postagem nas redes sociais[13] tem mantido campanhas de conscientização dos modos de desinformação. O uso de plataformas sociais pelo órgão para campanhas de veracidade informacional aconteceram também em 2022, quando lançou a série “Desinformação: 7 tipos de fake news. Sátira ou Paródia”[14].

Apesar das limitações e desafios enfrentados, os Instrumentos de checagem de fatos como Lupa[15], Aos Fatos[16] e TrueMedia[17], por exemplo, precisam ser fortalecidos e difundidos aos eleitores, de forma a viabilizar a pesquisa do que é verdade e mentira - inclusive, para além de questões políticas. Campanhas de conscientização acerca da existência, uso e reconhecimento de informações geradas por *deepfakes*, como o Programa de Combate à Desinformação do Supremo Tribunal Federal[18], também devem ser intensificadas de modo que chegue de forma acessível e entendível a todos os eleitores.

O melhor cenário que se encontra frente ao avanço desenfreado da atualização tecnológica, é a educação midiática e crítica como forma emancipadora de perceber o mundo a sua volta e, com isso, ser capaz de notar o que é de verdade e o que é distorção da realidade. A construção de políticas públicas de literacia digital para o uso de novas – e antigas – tecnologias se mostra essencial no contexto atual e futuro.

---

[12] Vídeo que mostra beijo entre Lula e Alexandre de Moraes foi gerado por IA. Disponível em: https://www.aosfatos.org/noticias/deepfake-beijo-lula-alexandre-de-moraes/
[13] Autor @tre_rondonia. Voto sem fake. Instagram, Data de publicação: 8 de maio de 2025. Disponível em: https://www.instagram.com/p/DJUI1ouxbH_/?img_index=1
[14] Fake news: a desinformação produz consequências graves. Disponível em: https://www.tre-ro.jus.br/comunicacao/noticias/2022/Julho/fake-news-a-desinformacao-produz-consequencia-graves
[15] Disponível em: https://lupa.uol.com.br/
[16] Disponível em: https://www.aosfatos.org/
[17] Disponível em: https://www.truemedia.org/
[18] Programa de Combate à Desinformação: Sociedade Informada, Democracia Forte. Disponível em: https://portal.stf.jus.br/desinformacao/

Ainda assim, atribuir toda responsabilidade às pessoas comuns e às instituições da sociedade civil não é o ideal, especialmente considerando que as redes sociais são o principal canal de disseminação das *deepfakes*. Diante desta perspectiva, torna-se essencial não apenas uma legislação capaz de regular a Inteligência Artificial, mas também estabelecer normas acerca do funcionamento das redes sociais, de modo a proteger direitos humanos e garantir segurança informacional, em razão do inquestionável uso desse tipo de tecnologia na manipulação da realidade político-social.

DEEPFAKE, DESINFORMAÇÃO E RACISMO: CIBER-RESISTÊNCIA COMO ESTRATÉGIA DE DEFESA PARA POPULAÇÕES NEGRAS E PERIFÉRICAS NAS REDES SOCIAIS

Gabriela de Almeida Pereira

Como citar este artigo?

PEREIRA, Gabriela de Almeida. Deepfake, desinformação e racismo: ciber-resistência como estratégia de defesa para populações negras e periféricas nas redes sociais. In: SOUZA, Gustavo; SILVA, Tarcizio (org.). **Enfrentando Deepfakes.** Desvelar: Brasília, 2025. Disponível em: desvelar.org/enfrentando-deepfakes.

Em entrevista à revista Veja, em 1992, a Ialorixá Gildásia dos Santos e Santos, conhecida como Mãe Gilda, expressou sua preocupação com os rumos da economia brasileira ao se juntar a manifestações populares pelo impeachment do então presidente Fernando Collor de Mello. Na foto que acompanhava a matéria, a mãe de santo vestia seus trajes religiosos habituais e apresentava sua contribuição ao país: uma oferenda preparada como súplica aos orixás. Sete anos depois, a mesma imagem foi retirada de contexto e usada na capa da Folha Universal, veículo da Igreja Universal do Reino de Deus, que a acusava de ser uma “macumbeira charlatã”. Na época, o jornal voltado ao público evangélico tinha uma tiragem de 1.372.000 exemplares e distribuiu de forma ampla e gratuita uma reportagem que colocava Mãe Gilda como exemplo de “macumbeiros charlatões que lesam o bolso e a vida de clientes”.

A publicação rapidamente espalhou a informação manipulada, causando pânico, medo e raiva entre o público-alvo. Como resposta imediata à mentira, o terreiro Ilê Axé Abassá de Ogum foi invadido, e Mãe Gilda tornou-se alvo de ataques de ódio[1], sofrendo uma série de agressões verbais e morais por parte de fiéis da Assembleia de Deus, dentro de seu próprio espaço religioso, em Salvador. A indignação motivada pela reportagem levou a repetidos episódios de agressão verbal contra a mãe de santo, figura central na articulação política por justiça social para a população de terreiro em sua época. Meses depois, a violência sofrida agravou seu estado de saúde e Mãe Gilda teve um infarto fulminante. O dia de sua morte, 21 de janeiro de 2000, passou a ser lembrado oficialmente como o Dia Nacional de Combate à Intolerância Religiosa, após a sanção da Lei nº 11.635, em 2007, pelo então presidente Luiz Inácio Lula da Silva.

Mais de 20 anos se passaram e a violência daquela foto posta em contexto mentiroso ainda respinga na memória de Mãe Gilda. O busto criado em sua homenagem, em Salvador, já foi vandalizado diversas vezes desde então. Um dos casos, datado de julho de 2020, foi praticado por um homem munido do que chama de fé, quebrou a estátua criada em homenagem à mãe de santo alegando que o ato

[1] Vozes Populares | Intolerância religiosa ou racismo religioso? https://www.brasildefato.com.br/podcast/vozes-populares/2022/01/20/vozes-populares-intolerancia-religiosa-ou-racismo-religioso/

teria sido motivado “à mando Deus”[2]. O estrago feito por uma única publicação teve consequências físicas, financeiras, psíquicas e memoriais para Mãe Gilda e seus familiares. Trago a história de Mãe Gilda para este trabalho para dar uma dimensão de como o ódio ocasionado pela disseminação de conteúdos falsos pode acarretar violência on e offline, justamente porque não há dissociação de nossa existência “dentro e “fora” da internet, visto que vivemos uma experiência híbrida, nomeada como “onlife”, pelo pesquisador italiano Luciano Floridi[3].

Criar mentiras e falsas acusações sobre pessoas negras não é e nem nunca foi uma novidade, tampouco surgiu com as plataformas e redes sociais. No entanto, o ambiente virtual acentuou processos acusatórios e difamatórios que têm como objetivo normalizar práticas racistas, aumentando a visibilidade e o alcance de conteúdos odiosos.

Mãe Gilda, infelizmente, não é exceção. Casos como os de Marielle Franco[4], João Pedro[5] e Marcos Vinícius[6] escancaram um padrão: a associação entre desinformação, discurso de ódio e violência racial que atravessa gerações — agora intensificada pelas novas tecnologias, como os deepfakes. O termo deepfake designa vídeos, áudios ou imagens manipuladas com inteligência artificial para simular de forma realista ações, falas ou rostos de pessoas conhecidas ou não. Embora existam usos legítimos e criativos – como o recente álbum que “devolveu” a voz ao sambista negro Cleber Augusto[7], ex-Fundo de Quintal, que não fala mais há 20 anos, desde a descoberta de um câncer nas cordas vocais - a tecnologia tem sido utilizada de forma cada vez mais comum em contextos de desinformação política, pornografia de

---

[2] Busto de Mãe Gilda é alvo de vandalismo em Salvador; suspeito foi levado para delegacia https://g1.globo.com/ba/bahia/noticia/2020/07/15/busto-de-mae-gilda-e-alvo-de-vandalismo-em-salvador-suspeito-foi-levado-para-delegacia.ghtml

[3] FLORIDI, Luciano. The Onlife Manifesto: Being Human in a Hyperconnected Era. London: Informática; Rio de Janeiro: Editora 34, 2015.

[4] Mentiras sobre Marielle Franco continuam a se espalhar três anos após sua execução. https://g1.globo.com/rj/rio-de-janeiro/noticia/2021/03/13/mentiras-sobre-marielle-franco-continuam-a-se-espalhar-tres-anos-apos-sua-execucao.ghtml

[5] Testemunha de caso João Pedro foi conduzida para depor em ‘caveirão’ da polícia. https://www.cnnbrasil.com.br/nacional/testemunha-de-caso-joao-pedro-foi-conduzida-para-depor-em-caveirao-da-policia/

[6] Justiça determina que Facebook remova fake news contra estudante morto na Maré. https://noticias.uol.com.br/cotidiano/ultimas-noticias/2018/06/29/justica-determina-que-facebook-apague-fake-news-contra-menino-marcos-vinicius.htm

[7] O show tem que continuar: IA devolve voz para sambista do Fundo de Quintal.https://www.uol.com.br/tilt/noticias/redacao/2025/05/12/como-ia-fez-sambista-gravar-album-mesmo-sem-poder-falar-ha-20-anos.htm

vingança, golpes e campanhas de difamação. E, como de costume, são as populações negras e periféricas — sobretudo mulheres negras, pessoas trans e ativistas de direitos humanos — que enfrentam os impactos mais violentos.

No Brasil, já há registros de uso de vídeos e áudios falsos para descredibilizar candidaturas negras, manipular falas de lideranças de favelas[8] ou mesmo simular cenas de violência que reforçam estereótipos racistas. O uso de deepfakes como arma política provoca pânico moral. Esta estratégia se combina a um ecossistema de plataformas de redes sociais com cada vez menos transparência algorítmica, bem como, a um imaginário social forjado no racismo estrutural e na violência como norma. Estes, por sua vez, são combustíveis que tornam a desinformação promovida por inteligência artificial ainda mais eficaz e danosa quando direcionada a corpos racializados.

Esse tipo de conteúdo viraliza por meio de um chamado contágio moral, onda que tem capacidade de atrair atenção para esse tipo de conteúdo, facilitando sua rápida disseminação na internet[9] justamente porque essa viralização é facilitada pelo fato de o conteúdo moral e emocional ser mais propenso a capturar a atenção do usuário. A mensagem viraliza quando o receptor se sente motivado, e a indignação, o medo e o entusiasmo motivam o indivíduo a repassar o conteúdo recebido.

E, como é de se imaginar, a criação e a viralização de boatos envolvendo pessoas negras ganham impulso por reforçar outras tantas mentiras que durante séculos foram construídas sobre a população negra para normalizar práticas racistas.

De acordo com dados do Global Digital Report 2024[10], o Brasil é o segundo país em que os usuários passam mais tempo online, atrás somente da África do Sul. As cerca de nove horas que o brasileiro dedica ao ambiente online são atravessados por uma série de riscos aos quais os usuários ficam expostos em plataformas digitais como Instagram, TikTok, Kwai, Facebook e WhatsApp.

---

[8] O ativista que é vítima em série de fake news bolsonaristas. https://www.bbc.com/portuguese/brasil-64288499

[9] BRADY, William; CROCKETT, M; VAN BAVEL, Jay. The mad model of moral contagion: the role of motivation, attention and design in the spread of moralized content online. Perspectives on Psychologial Science, 15 (4), 978-1010. https://new.safernet.org.br/content/crimes-de-odio-tem-crescimento-de-ate-650-no-primeiro-semestrede-2022.

[10] Global Digital Report 2024. https://datareportal.com/reports/digital-2024-global-overview-report

A alta exposição a uma oferta variada de conteúdos, aliado a um baixo letramento midiático, propiciam um ambiente fértil para desinformação e discursos de ódio. Neste contexto, alguns grupos ficam ainda mais vulneráveis, como idosos, pessoas de baixa renda, crianças e adolescentes, além de população negra e LGBTQIAP+.

Diante desse cenário, é cada vez mais urgente que se disseminem estratégias de ciber-resistência para uma resposta coletiva de enfrentamento aos danos que podem ser provocados por deepfakes e conteúdos desinformativos. Mais do que reagir a ataques pontuais, trata-se de fortalecer redes e práticas de defesa que combinem vigilância crítica, ação política e construção de autonomia digital — especialmente entre comunidades negras e periféricas. Isso passa por educação midiática racializada, por políticas públicas que alcancem a população para além do ciclo escolar, por marcos regulatórios com abordagem interseccional e por uma profunda mudança na forma como plataformas e governos tratam a integridade da informação.

Cabe destacar que a própria ideia de "integridade da informação" não pode ser neutra. Ela precisa ser conectada à integridade dos corpos, da memória e dos direitos de populações historicamente marginalizadas. Quando um vídeo manipulado simula que um jovem negro cometeu um crime; quando um áudio manipulado faz parecer que uma ativista mentiu; ou quando um rosto é utilizado para gerar conteúdo pornográfico falso, seja de mulheres e até mesmo de crianças[11] — o que está em jogo não é só a reputação, mas a vida. O dano é emocional, político e coletivo.

É fundamental, ainda, destacar a resistência crescente, especialmente nas redes sociais, onde as ciberativistas negras, como Jéssica Mara Raul[12] aponta, desempenham um papel crucial na luta contra a desumanização. O ciberespaço se transforma em uma arena de resistência antirracista e de produção e preservação das memórias negras, servindo como um canal de educação e mobilização. Esses espaços coletivos têm se mostrado essenciais para enfrentar a violência simbólica e

[11] IA aumenta circulação de vídeos fakws com abuso sexual de crianças https://agenciabrasil.ebc.com.br/internacional/noticia/2024-07/ia-aumenta-circulacao-de-videos-fakes-com-abuso-sexual-de-criancas

[12] RAUL, Jessica Mara. Entre silêncios e protestos: uma reflexão sobre escrita preta no ciberespaço. Revista Docência e Cibercultura, Rio de Janeiro, v. 3, n. 3, p.166-194, 31 dez. 2019. Universidade de Estado do Rio de Janeiro. http://dx.doi.org/10.12957/redoc.2019.44955.

reverter o apagamento da memória negra, criando uma contraposição significativa ao discurso de supremacia e violação de direitos.

Para uma saída que seja de fato coletiva, é preciso que nos conectemos com o amor enquanto prática política de bell hooks, assim como Audre Lorde e suas reflexões sobre cuidado e autocuidado. Muito antes do termo ter sido apropriado como sinônimo de “skin care” e de noções individuais de “bem-estar”, Lorde relacionava o cuidado à solidariedade e ao compartilhamento do afeto, à amorosidade e à importância de que esse cuidado se dê no coletivo, para que tenhamos atenção a quem nos cerca e, assim, possamos nos proteger.

Combater narrativas de ódio e fortalecer a memória coletiva de pessoas negras é algo que cabe a todos nós, enquanto sociedade. O amor enquanto prática política tem um potencial transformador e inspira novas práticas de resistência. Somente assim será possível a consolidação de uma sociedade mais justa, plural e comprometida com a dignidade e o bem-viver de todas as pessoas.

DEEPFAKES E DESORDEM INFORMACIONAL: A BANALIZAÇÃO DO SINTÉTICO

Laura Pereira

Como citar este artigo?

PEREIRA, Laura. Deepfakes e desordem informacional: a banalização do sintético. In: SOUZA, Gustavo; SILVA, Tarcizio (org.). **Enfrentando Deepfakes.** Desvelar: Brasília, 2025. Disponível em: desvelar.org/enfrentando-deepfakes.

Um retrato de momento

Em 2019, pós-eleições presidenciais, a produção de vídeos sintéticos que simulavam a imagem de atores políticos passou a integrar a gramática política brasileira, sobretudo a partir das produções humorísticas do jornalista Brunno Sarttori. Naquele ano, em um episódio do podcast Hoje Tem, Sarttori comentou sobre as barreiras para o uso indevido da técnica:

> Acho que a gente ainda tem essa limitação para que as pessoas não saiam aloucadamente fazendo deepfakes, que é a falta de um *hardware* potente, porque senão a pessoa vai ficar ali uma vida inteira esperando o processamento ser feito, e a falta de conhecimento necessário para saber o que usar ou não para se chegar a um resultado. (...) Por exemplo, tem quase 100 dias que aquele vídeo (...) do Chapolin veio à tona e virou o que virou. E até hoje não apareceu nenhum trabalho semelhante na internet. Por quê? Porque as pessoas estão tendo dificuldade para reproduzir. Não é uma coisa simples de se fazer (...).[1]

A incorporação de ferramentas de Inteligência Artificial Generativa (IAG) em aplicativos, sites e plataformas[2] de uso acessível transformou radicalmente esse cenário em um intervalo de tempo relativamente curto. O que antes era restrito a produções audiovisuais especializadas passou a se disseminar em múltiplas variações de conteúdos sintéticos, acessíveis para criação e consumo de um número crescente de usuários.

Este *paper* se propõe a sistematizar uma perspectiva analítica sobre os efeitos da IAG no contexto da desordem informacional, com base na descentralização da emissão de conteúdo e na incorporação de tecnologias disruptivas ao cotidiano social. Trata-se de um primeiro exercício teórico-reflexivo a ser posteriormente aprofundado caso os caminhos aqui ensaiados se mostrem profícuos. O enfoque recai sobre as intersecções entre tecnologia, política democrática e a dieta midiática dos cidadãos.

Na primeira seção, busca-se delimitar conceitualmente os termos desordem informacional, deepfakes e ciclo de adoção tecnológica, com ênfase em definições

[1] HOJE TEM. Podcast Hoje Tem #10: República Bolchevique e 1 ano da nova era. 2024. Disponível em: https://open.spotify.com/episode/2gH0PgresK66PTwQhW7QKi?si=9a29bb9134284415. Acesso em: 11 maio 2025.
[2] SRNICEK, Nick. **Platform Capitalism.** Cambridge: Polity Press, 2017.

operacionalizáveis, mais do que exaustivas. Na segunda seção, agrega-se uma análise inicial de casos empíricos e recentes de circulação dessa tecnologia.

Desordem informacional e tecnologias emergentes

Parte substancial da literatura sobre tecnologias e política democrática tem se debruçado sobre os efeitos da circulação de conteúdos na experiência das democracias contemporâneas. Essa preocupação é histórica no campo dos estudos sobre tecnologias da comunicação ou usadas para fins comunicativos, ganhando força a partir da modernidade com a "democracia de multidões" e a comunicação de massa[3], momento em que a comunicação moderna se torna um elemento intrínseco à própria constituição da democracia moderna[4].

Com o surgimento da Internet e, a partir da década de 2010, com a consolidação das redes sociais digitais e outras formas de comunicação mediada por tecnologia, esse processo se intensifica, incorporando novas formas de desinformação e manipulação informacional. Conceitos como *fake news*, *misinformation*, *malinformation e disinformation* passam a fazer parte do debate público e acadêmico, culminando em abordagens mais abrangentes, como desordem informacional e riscos à integridade informacional, que são hoje presentes tanto no campo acadêmico quanto no debade de elaboração de políticas (WARDLE, DERAKHSHAN, 2017[5]; SANTOS, 2024[6]).

Os *deepfakes* inserem-se de forma paradigmática nesse contexto. A técnica de produção de áudio, imagem ou vídeo, que simula características de pessoas reais a partir da substituição ou criação de elementos visuais, emergiu em 2017, quando um algoritmo foi divulgado por um usuário que publicou uma série vídeos pornográficos no Reddit, substituindo os rostos das atrizes por celebridades[7]. O uso mais conhecido na política viria em 2018, com a circulação de um vídeo em que Barack Obama

---

[3] THOMPSON, John B., **A mídia e a modernidade**: uma teoria social da mídia. 5ª ed. Petrópoles, RJ: Vozes, 1998.

[4] MANIN, Bernard. As metamorfoses do governo representativo. **Revista brasileira de ciências sociais**, v. 10, n. 29, p. 5-34, 1995.

[5] WARDLE, Claire; DERAKHSHAN, Hossein. **Information disorder: Toward an interdisciplinary framework for research and policymaking**. Strasbourg: Council of Europe, 2017.

[6] SANTOS, Nina. Por que precisamos discutir a "integridade da informação"?. *Diplomatique*. 6 fev 2024. Disponível em: <https://diplomatique.org.br/integridade-da-informacao/>. Acesso em: 11 maio 2025.

[7] MADDOCKS, Sophie. 'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political'deep fakes. Porn Studies, v. 7, n. 4, p. 415-423, 2020.

aparecia dizendo frases esdrúxulas. O material, em tom de alerta, contou com a participação de Jordan Peele e foi publicado pelo BuzzFeed (BUZZFEED, 2018)[8].

Modelos sociológicos de difusão de tecnologias ajudam a compreender essa dinâmica[9]. Inicialmente restritas a grupos de vanguarda, elas enfrentam barreiras de entrada como custo, complexidade técnica e disponibilidade de recursos. Cerca de sete anos após o início dos *deepfakes*, os usos recentes da técnica são mais facilitados e vão muito além de uma concepção restrita de desordem informacional à veracidade dos conteúdos ou na sua capacidade de enganar.

Se em 2018, no Brasil, uma das principais polêmicas eleitorais girou em torno da dúvida sobre a veracidade de um vídeo íntimo atribuído ao então candidato a governador por São Paulo, João Doria[10], os usos mais recentes da técnica evidenciam que nem tudo desemboca apenas em ambiguidade factual, assim como ocorreu com a própria noção de *fake news*. A proliferação de imagens, áudios e vídeos sintéticos tem contribuído para acostumar os cidadãos a esse tipo de ambiguidade e a potencializar, em realidade, a circulação de uma série de materiais políticos que redefinem as linhas pelas quais os cidadãos compreendem a experiência democrática e, sobretudo, a própria ação coletiva para fins políticos. O resultado é o progressivo embaralhamento das fronteiras não entre a realidade e a factualidade, mas entre a opinião ou o conteúdo atribuído de significado e os elementos que podem não traçar essa mesma linha de convencimento e confiança.

Deepfakes na atualidade

Estudos recentes, de 2024 e 2025, vêm demonstrando que o impacto da Inteligência Artificial Generativa em disputas eleitorais e contextos políticos não assumiu, ao menos até agora, a escala e as características inicialmente previstas

[8] BUZFEED. **You Won't Believe What Obama Says In This Video! 😉 - YouTube**. Disponível em: <https://www.youtube.com/watch?v=cQ54GDm1eL0>. Acesso em: 11 maio 2025.
[9] HALAWEH, Mohanad. Model of emerging technology adoption (META): virtual reality as a case study. **Journal of Information & Knowledge Management**, v. 18, n. 02, p. 1950020, 2019.
[10] Registros da época mostram que a possibilidade do material ser um *deepfake* foram rapidamente levantadas. Mais informações em: https://www.correiobraziliense.com.br/app/noticia/politica/2018/10/24/interna_politica,714810/video-de-doria-e-fake-entenda.shtml

(Desinformante, 2024[11]; FGV Comunicação, 2024a[12]). Ainda que usos variados da tecnologia já comecem a circular, não se verificou uma proliferação massiva de campanha negativa e conteúdos desinformativos. Na Índia, por exemplo, partidos e familiares de políticos já mortos autorizaram, produziram e distribuíram *deepfakes* em que essas figuras apoiavam atuais candidatos (Idem).

No Brasil, um dos poucos países a possuir uma normativa eleitoral prévia específica sobre identificação obrigatória de conteúdos produzidos com IA e proibição de deepfakes, casos variados também aconteceram (JUNQUILHO et al., 2024[13]; FGV Comunicação, 2024b[14]). Os usos mais frequentes da tecnologia, no entnanto, não se associaram diretamente à mentira ou à falsificação de fatos, mas à representação audiovisual de opiniões políticas, quase como peças de campanha, e, de forma mais chamativa e negativa, de *deepnudes*.

Segundo o Observatório de IA da organização Desinformante[15], ao menos cinco casos de deepnudes foram reportados no Brasil no 1º turno das eleições de 2024, com a produção e circulação de conteúdos sintéticos de cunho sexual envolvendo candidatas, com circulação em redes sociais e sites pornográficos. Nesses materiais, não estava em cheque o alcance do nível de qualidade técnica para gerar veracidade, embora esse convencimento sempre possa ocorrer, mas a própria atribuição da difamação que é característica da circulação não consensual de imagens íntimas de mulheres, verdadeiras ou fabricadas[16].

---

[11] CRUZ, M.; SANTOS, N.; CARREIRO, R.; NÓBREGA, L.; AMORIM, Gabriel. IA no primeiro turno: o que vimos até aqui. Salvador e São Paulo: Aláfia Lab & Data Privacy Brasil, 2024. Disponível em: https://desinformante.com.br/relatorio-ia-eleicoes/

[12] FGV Comunicação. Inteligência Artificial no mundo: mapeamento dos usos durante eleições de 2024. Relatório de pesquisa. Disponível em: https://midiademocracia.fgv.br/estudos/inteligencia-artificial-no-mundo-mapeamento-dos-usos-durante-eleicoes-de-2024. Acesso em: 14 maio 2025.

[13] JUNQUILHO, Tainá Aguiar; SILVEIRA, Marilda de Paula; FERREIRA, Lucia Maria Teixeira; MENDES, Laura Schertel; OLIVEIRA, André Gualtieri de. (org.). Construindo consensos: deep fakes nas eleições de 2024 relatório das decisões dos TREs sobre deep fakes. Brasília: Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa: Laboratório de Governança e Regulação de Inteligência Artificial, 2024. E-book. ISBN 978-65-87546-23-0

[14] FGV COMUNICAÇÃO. Imaginários sobre a Inteligência Artificial no debate digital brasileiro. Rio de Janeiro, 2024. Disponível em: https://drive.google.com/file/d/1glCfmHeuINW75-QxulS2bPnz6k8tAkM8/view

[15] CRUZ, M.; SANTOS, N.; CARREIRO, R.; NÓBREGA, L.; AMORIM, Gabriel. IA no primeiro turno: o que vimos até aqui. Salvador e São Paulo: Aláfia Lab & Data Privacy Brasil, 2024. Disponível em: https://desinformante.com.br/relatorio-ia-eleicoes/

[16] LANA, Alice de Perdigão. Nudez na internet: mulheres, corpo e direito. Curitiba: IODA, 2023

Desde então, observa-se uma tendência de aumento na proliferação desse tipo de conteúdo pela perspectiva da reiteração de imaginários políticos. Após as eleições, diversos episódios foram acompanhados da circulação de imagens criadas com IA. Agências de checagem identificaram imagens sintéticas no contexto do acidente doméstico sofrido por Lula, em dezembro de 2024 – várias delas produzidas com o Grok, modelo de IA da plataforma X e aberto aos seus usuários[17]. Outros vídeos falsos circularam com piadas sobre o ministro da Fazenda, Fernando Haddad, associando-o a taxações de produtos. No cenário político, atores como Donald Trump vem adotando prática frequente de compartilhamento de imagens alteradas em suas redes oficiais.

Imagem 1: Composição a partir de posts de Donald Trump no Truth e no X[18]

Esses casos apontam para a naturalização do convívio com imagens e vídeos alterados como uma prática cultural difundida na experiência online, que tanto já não cause espanto quanto também não é completamente distante. Isso não elimina o fato de que conteúdos com finalidade política tendem a circular de forma intencional e com financiamento estruturado. Tampouco resolve a questão central de por que, no caso das tecnologias de comunicação – diferentemente de outras –, uma das únicas

[17] Mais informações em: https://www.instagram.com/comprova/reel/Cxgl9-Wo7D6/?next=%2Fkbalaji800%2Ffeed%2F&hl=am-et e https://www.aosfatos.org/noticias/falso-haddad-disse-plano-governo-lula-taxar-tudo/ e

[18] Disponível em: https://x.com/realDonaldTrump/status/1825138139502878806 e https://truthsocial.com/@realDonaldTrump/posts/114441543826801216

barreiras efetivas ao seu uso indevido ainda está ligada à adoção massiva e aos recursos técnicos disponíveis. Ainda assim, os exemplos analisados evidenciam nuances importantes no processo de consolidação de um imaginário político e informacional.

Considerações finais

Estudos e *surveys* tem apontado para uma dinâmica complexa de receio da tecnologia e adoção utilitarista. Mais além no entanto e conforme a hipótese de fundo neste material, por que a dimensão utilitarista pode sair na frente na relação entre IA e informação política. Nesse contexto, verdade e mentira podem rapidamente ser relegadas a um segundo plano, à medida que o conteúdo mais eficaz visualmente – aquele que melhor representa uma posição política ou que gera maior entretenimento – ganha centralidade, assim como ocorreu com a simples circulação de desinformação quando criada completamente sem auxílio de ferramentas de IAG.

Mais do que isso, inúmeras referências coletivas de 2025 já são baseados nesse tipo de tecnologia, como um meme em que se simula um abraço em Jesus. Em abril, a liberação de um modelo de imagens no ChatGPT gerou uma febre de imagens no estilo do Studio Ghibli. Transformar fotos de pets em versões humanas ou criar versões de como você seria se fosse um brinquedo em uma caixa são tendências similares. Nesse contexto, os desafios para um ambiente de maior precisão e compartilhamento de noções seguem, já que poucos elementos além da desconfiança absoluta ou do viés de confirmação parecem ser aptos a assumirem um papel republicano nesse debate.

DEEPFAKES NO CENÁRIO BRASILEIRO: EFEITOS E ESTRATÉGIAS DE MITIGAÇÃO

Júlia Caldeira

Como citar este artigo?

CALDEIRA, Júlia. Deepfakes no cenário brasileiro: efeitos e estratégias de mitigação. In: SOUZA, Gustavo; SILVA, Tarcizio (org.). **Enfrentando Deepfakes.** Desvelar: Brasília, 2025. Disponível em: desvelar.org/enfrentando-deepfakes.

Conforme já apresentado, as deepfakes podem ser definidas como falsificações realistas de conteúdos midiáticos gerados por Inteligência Artificial (IA). Visto a criação de mídias a partir de recortes da realidade, são gerados muitas vezes produtos surpreendentes, capazes de manipular e enganar mesmo indivíduos com alto letramento digital. Assim, em vista à sua facilidade de criação, muitas vezes por mecanismos gratuitos, rápidos e intuitivos, tornaram-se um risco à integridade da informação, gerando um caos informacional. Em paralelo, indivíduos que já se apresentam em situações de vulnerabilidade em relação ao acesso e uso das mídias, tornam-se alvos frequentes, manifestando defasagens na educação midiática em escala micro e macro.

No cenário brasileiro, a pauta tem sido discutida sob algumas perspectivas, ganhando também espaço de discussão em âmbito regulatório. São exemplos o já comentado Projeto de Lei (PL) 3821/2024 e o PL 370/2024, aprovado pelo Senado, que amplia penas para deepfakes com fins de violência psicológica contra mulheres[1]. Ademais, insere-se no contexto a Lei de Direitos Autorais (9.610/1998), a qual costuma ser invocada para coibir usos não autorizados de conteúdos[2]. No entanto, a lei esbarra em desafios técnicos para comprovar autoria e intenção maliciosa. Paralelamente, iniciativas público-privadas buscam integrar ferramentas de detecção via IA, blockchain e campanhas de educação midiática para mitigar os impactos, refletindo um esforço emergente – ainda insuficiente – para equilibrar inovação e proteção social[3].

Voltando-se aos efeitos das deepfakes no cenário brasileiro, são exemplos de casos recentes fraudes financeiras, manipulação de imagens de celebridades e riscos à integridade informacional, materializando um aumento de 830%[4] nos incidentes

---

[1] AGÊNCIA SENADO. Senado aprova projeto que amplia penas para deepfake com fins de violência psicológica. *Agência Senado*, 2024. Disponível em: https://www12.senado.leg.br/noticias/materias/2024/03/21/senado-aprova-projeto-que-amplia-penas-para-deepfake-com-fins-de-violencia-psicologica.

[2] BRASIL. Lei nº 9.610, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. *Diário Oficial da União*, Brasília, DF, 20 fev. 1998. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9610.htm.

[3] CÂMARA DOS DEPUTADOS. Câmara aprova projeto que criminaliza deepfake com conteúdo sexual. *Câmara Notícias*, 2025. Disponível em: https://www.camara.leg.br/noticias/1065406-camara-aprova-projeto-que-criminaliza-deepfake-com-conteudo-sexual/.

[4] KASPCHAK, Rafael. Deepfakes crescem 830% no Brasil e ameaçam eleições. *Exame*, 2024. Disponível em: https://exame.com/tecnologia/deepfakes-crescem-830-no-brasil-e-ameacam-eleicoes/.

relatados até 2025. O cenário é marcado por golpes que utilizam clones digitais de figuras públicas como Anitta[5], Neymar e Paolla Oliveira[6] para disseminar desinformação ou aplicar crimes cibernéticos.

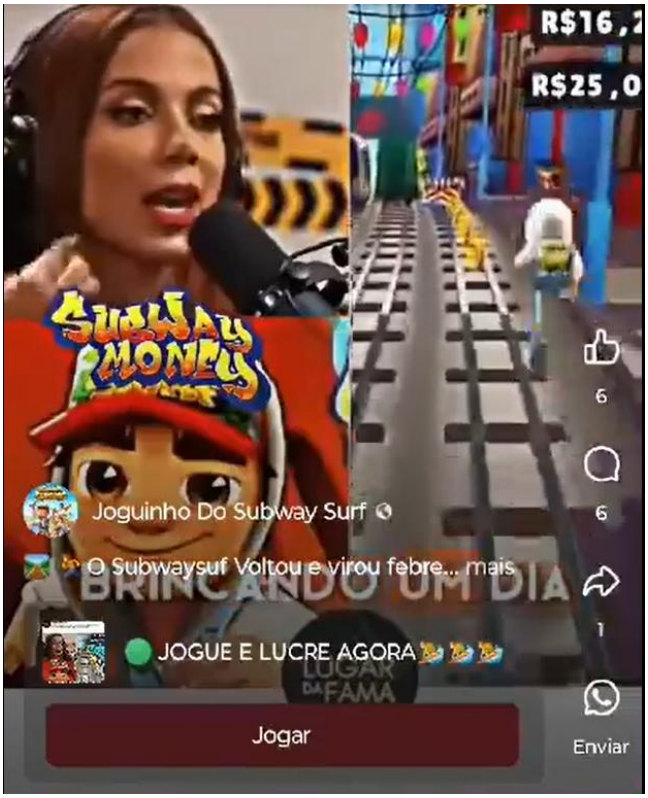


Figura 1: Publicidade do aplicativo Subway Money, exibindo vídeo deepfake da cantora Anitta recomendando o jogo.

Frente a isso, dois exemplos brasileiros que evidenciam o impacto das deepfakes em diferentes setores da sociedade são: a publicação da resolução 23.732/2024 pelo Tribunal Superior Eleitoral (TSE) para as eleições de 2024 e o caso Giovanna Ewbank, de 2025. No primeiro, frente aos iminentes riscos à integridade do processo eleitoral gerado pelas deepfakes, aprovou-se a referida resolução com o objetivo de se implementar inovações quanto ao uso de IA e ao enfrentamento da

[5] Folha de São Paulo. Golpistas usam imagem de Neymar, Anitta e famosos para divulgar jogo falso Subway Money. Folha de São Paulo, 2024. Disponível em: https://web.archive.org/web/20240704090808/https://www1.folha.uol.com.br/tec/2024/01/golpistas-usam-imagem-de-neymar-anitta-e-famosos-para-divulgar-jogo-falso-subway-money.shtml

[6] O Globo. Neymar, Paolla Oliveira e outros famosos viram vítimas de deepfake: Saiba os riscos e como se proteger. *O Globo*, 2025. Disponível em: https://oglobo.globo.com/ela/noticia/2025/04/05/neymar-paolla-oliveira-e-outros-famosos-viram-vitimas-de-deepfake-saiba-os-riscos-e-como-se-proteger.ghtml.

desinformação[7]. Em seu texto, constam medidas que visam tornar mais fáceis o processo de remoção de conteúdos danosos para o pleito eleitoral[8]. Além disso, o texto introduz um conjunto de obrigações para provedores de aplicações, exigindo ações proativas no combate a esse tipo de conteúdo, incluindo discurso de ódio e aqueles que configuram crime ao Estado democrático de Direito.

Em sua publicação, a referida resolução do TSE foi alvo de críticas, tendo sido apontada como contrária ao modelo de responsabilidade determinado pelo artigo 19 do Marco Civil da Internet (MCI) e mesmo como ofensa à liberdade de expressão dos usuários. Conforme apontam Curzi et al[9], o tribunal apresenta competência jurisdicional, normativa e administrativa, atribuídas pelo Código Eleitoral para a publicação e aplicação do referido texto. Ademais, a necessidade da resolução se fez frente a um contexto de vácuo regulatório em relação às novas ferramentas de IA, as quais precisam de contramedidas em um ritmo rápido e efetivo.

Ainda, vale trazer à tona que a maior responsabilização dos provedores de plataformas digitais nesse contexto é uma medida necessária. Frente ao modelo de negócios das plataformas, conteúdos virais e de fácil engajamento geram maiores lucros aos provedores. Observa-se assim que, enquanto geram danos à democracia e à integridade da informação, esses conteúdos são monetizados pelas plataformas. Nesse contexto, mecanismos de moderação de conteúdo são necessários para a sanitização do meio digital e promoção da democracia.

O segundo caso que envolve os impactos das deepfakes no Brasil diz respeito à situação que ocorreu com Giovanna Ewbank, em março de 2025[10]. A atriz foi vítima de um vídeo falso, divulgado em redes sociais, no qual sua imagem e voz foram manipuladas por IA para a promoção do uso de botox de forma não autorizada O conteúdo foi amplamente disseminado em plataformas digitais, utilizando anúncios

---

[7] TAVARES, Cláudio de Mello. Inteligência Artificial e Deepfakes: Desafios Jurídicos e Tecnológicos para a Integridade do Processo Democrático e as Implicações para as Eleições Municipais de 2024. *Revista Justiça Eleitoral em Debate,* v. 14, n. 1, p. 49-58.

[8] DIAS, Daniel Pires Novais; BELLI, Luca; ZINGALES, Nicolo; GASPAR, Walter; CURZI, Yasmin. TSE, plataformas digitais e competência normativa: Resolução n˚ 23.732/2024. *Consultor Jurídico,* 27 mar. 2024. Disponível em: https://www.conjur.com.br/2024-mar-27/tse-plataformas-digitais-e-competencia-normativa-uma-analise-da-resolucao-no-23-732-2024/.

[9] *Ibidem*

[10] O GLOBO. Caso Giovanna Ewbank: saiba como identificar deepfake e se proteger deste crime. *O Globo*, 24 mar. 2025. Disponível em: https://oglobo.globo.com/ela/noticia/2025/03/24/caso-giovanna-ewbank-saiba-como-identificar-deepfake-e-se-proteger-deste-crime.ghtml. Acesso em: 16 maio 2025.

pagos para alcançar um maior público, ilustrando por sua vez os efeitos das deepfakes envolvendo figuras públicas e a variedade de situações e grupos envolvidos na problemática.

Figura 2: Giovanna Ewbank em deepfake supostamente recomenda aplicação de botox.

Em vista aos casos apresentados, é possível observar que as deepfakes constituem um risco presente e iminente no contexto brasileiro. Frente a isso, como seria possível mitigar seus efeitos? Propõe-se aqui estratégias baseadas na Educação midiática e em seu fortalecimento, uma vez que o letramento digital dos usuários é um fator de extrema relevância para medidas que envolvem desde a identificação desses conteúdos danosos até a criação de uma cultura consciente e ativa em seu combate.

Dessa forma, tem-se que a educação midiática, ao desenvolver o pensamento crítico, a análise criteriosa de fontes e a compreensão dos processos de produção de conteúdo, capacita cidadãos – especialmente jovens – a reconhecerem sinais de manipulação em vídeos, áudios e imagens, tornando-os assim menos suscetíveis à desinformação e a fraudes digitais[11]. Nesse contexto, iniciativas como a Semana

[11] AGÊNCIA BRASIL. Com educação midiática, estudantes são aliados no combate a fake news. *Agência Brasil,* 30 out. 2024. Disponível em: https://agenciabrasil.ebc.com.br/educacao/noticia/2024-10/Com-educacao-midiatica-estudantes-sao-aliados-no-combate-a-fake-news.

Brasileira de Educação Midiática[12] e a Estratégia Nacional das Escolas Conectadas[13] promovem a integração de ferramentas e conteúdos interativos nas escolas, incentivando o uso responsável das mídias e a participação ativa dos estudantes no combate às fake news, colaborando para a criação de uma cultura de usuários conscientes e ativos.

Além disso, tem-se que a alfabetização midiática deve ser ampliada para além do ambiente escolar, envolvendo toda a sociedade em ações de conscientização e atualização constante sobre novas técnicas de manipulação digital. Dessa forma, a educação midiática, aliada à soluções tecnológicas de detecção e autenticação de conteúdos, constitui um pilar fundamental para mitigar os impactos negativos dos deepfakes e fortalecer a resiliência democrática no Brasil.

[12] BRASIL. Secretaria de Comunicação Social da Presidência da República. 2ª Semana Brasileira de Educação Midiática. Brasília: *SECOM,* 2024. Disponível em: https://www.gov.br/secom/pt-br/assuntos/educacao-midiatica/2asbem.

[13] BRASIL. Decreto nº 11.713, de 26 de setembro de 2023. Institui a Estratégia Nacional de Escolas Conectadas. *Diário Oficial da União:* seção 1, Brasília, DF, 27 set. 2023. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/decreto/d11713.htm.

É TUDO VERDADE? A MOÇA DO TÁXI E AS DEEPFAKES

Thiane Neves Barros

Como citar este artigo?
NEVES-BARROS, Thiane. É tudo verdade? A moça do táxi e as deepfakes. In: SOUZA, Gustavo; SILVA, Tarcizio (org.). **Enfrentando Deepfakes.** Desvelar: Brasília, 2025. Disponível em: desvelar.org/enfrentando-deepfakes.

Apresentação

Neste artigo, o objetivo é fazer uma relação entre lendas urbanas e as *deepfakes* como forma de contribuir para a nossa compreensão de que a manipulação de fotos e vídeos, sem consentimento de quem de direito, é uma violência e é um crime. Enquanto lendas urbanas surgem como resultado positivo de nossas relações sociais, as *deepfakes* vão na contramão da saúde da nossa sociedade, são um retrato do mau uso dos recursos tecnológicos e da falta de regulação das plataformas digitais.

O texto está dividido em duas partes, sendo a primeira a contextualização de uma lenda urbana que faz parte da cultura da minha cidade e na segunda parte um convite para refletirmos juntes sobre os riscos de deixarmos que as *deepfakes* atuem como se fossem lendas urbanas.

Pois bem, o imaginário belemense conta que uma vez por ano, todo ano, uma moça pegava um táxi na porta do Cemitério Municipal Santa Izabel, pedia que o motorista rodasse pela cidade e a deixasse de volta na calçada do cemitério, sempre na data de seu aniversário. O Santa Izabel é um cemitério público, um dos mais antigos de Belém e talvez o maior do seu tempo de criação, que data do século XIX, são aproximadamente 47 mil sepulturas[1].

Sua localização fica na fronteira entre dois bairros: São Brás e Guamá. Por ser um cemitério cujas sepulturas são antigas, existem dezenas de relatos de visagens e assombrações por pessoas que transitam ou transitavam pela calçada fosse de noite ou de madrugada.

A lenda de Josephina Conte (Figura 1), nossa famosa moça do táxi, rende até hoje dezenas de homenagens e pedidos por Graças em sua sepultura, é uma das recordistas de visitas em datas como o Dia de Finados, é de lei que a cobertura jornalística mostre

[1] LENNON, Raissa. Retrato do patrimônio histórico de Belém, cemitério Santa Izabel completa 143 anos. 2021. Disponível em: https://redepara.com.br/Noticia/220127/retrato-do-patrimonio-historico-de-belem-cemiterio-santa-izabel-completa-143-anos

sua sepultura. A moça do táxi é inquestionavelmente uma visagem que foi cristalizada como verdade.

A moça do táxi não é uma *deepfake*. Não é uma manipulação da verdade, a história ao redor dela não é um crime cometido de forma premeditada para causar qualquer prejuízo moral ou financeiro a ela ou à sua família. Mas é o tipo de história que é tão contada que nós acreditamos na sua veracidade, compartilhamos irrestritamente, multiplicamos versões e garantimos que conhecemos relatos reais. Não nos restam dúvidas de que a moça do táxi passeava de táxi nos seus aniversários. Essa história aconteceu no começo do século XX e é recontada de diversas formas, com variadas entonações e sempre com um ar de mistério que também requenta nosso interesse, ela é atualizada o tempo todo, modernizada, e Josephine segue quase imortal em sua lenda.

Figura 1 - Print do perfil Estado do Pará News no Instagram

Lendas urbanas são histórias fantásticas que além de serem parte do imaginário de uma coletividade, fortalecem identidades, contribuem em nossa socialização e, claro, muitas vezes também são criadas para direcionar comportamentos de grupos sociais. Então, você pode ter medo do Chupa Cabra, do Boi da Cara Preta, da Matinta Pereira ou da Pisadeira, mas deve conhecer e reconhecer lendas como da moça do táxi que contribuem para a cultura de seu território.

Deepfake não é uma lenda urbana: publicação de foto e vídeo sem consentimento é crime

Sabe aqueles vídeos que as pessoas colocam os rostos de outras e simulam suas vozes? Já aconteceu dezenas de vezes no Brasil. Isso é deepfake. Faz parte do pacote do que chamamos de desinformação, mas com um nível de refinamento muito avançado.

A desinformação passou a ser um lucrativo negócio no Brasil e as *deepfakes* também entraram no ramo das "encomendas" da inteligência artificial. Ao contrário de uma lenda urbana, as *deepfakes* são manipulações criminosas que comprometem pessoas, inventam situações ou discursos, simulam e adulteram a verdade a partir de valores ideológicos que cultivam ódio e violência. É tão grave o cenário no Brasil que a Agência Senado noticiou que 24% da população brasileira já caiu em algum tipo de golpe ou crime na internet[2].

As situações criadas em *deepfakes* são tão falsas quanto a nota de R$ 50 com o rosto de Pabllo Vitar[3] e prejudicam a sociedade de forma ampla, visto que quem cria *deepfakes* na maioria das vezes está incentivado por esses discursos e atos de ódio, pois se alguém distorce um vídeo para criar uma mentira, não está bem intencionado. Em reportagem no portal da BBC, a jornalista Shin Suzuki[4] mostra que inclusive existem grupos profissionais em vender deepfake, essas pessoas são contratadas principalmente para atacar e extorquir mulheres, políticos e ativistas.

> A reportagem da BBC News Brasil presenciou em tempo real participantes da comunidade no Telegram encomendando deepfakes pornô a partir de fotos de mulheres que diziam ser suas conhecidas.
> Depois de algum tempo, quem tinha feito o pedido recebia a imagem manipulada de volta. Uma das mensagens mostrava, por exemplo, uma selfie comum — sem grande produção ou cuidado com iluminação — de uma mulher vestida. A seguir, apareceu uma versão digitalmente modificada em que ela aparecia nua.
> Participante 1 - Quem é?
> Participante 2 - Mãe de um amigo meu kkkk

[2] AGÊNCIA SENADO. Golpes Digitais atingem 24% da população brasileira, revela Datasenado. Disponível em: https://www12.senado.leg.br/noticias/materias/2024/10/01/golpes-digitais-atingem-24-da-populacao-brasileira-revela-datasenado.
[3] A imagem pode ser vista aqui: https://caraoucoroa.blogosfera.uol.com.br/2018/02/21/pablo-vittar-vai-estampar-nova-nota-de-r-50-gente-isso-e-verdade/.
[4] O grupo brasileiro de ódio a mulheres que fabrica com IA imagens pornô falsas sob encomenda. 2024. Disponível em: https://www.bbc.com/portuguese/articles/cv2g3wvgldlo

> Nos diálogos travados no grupo, os participantes demonstravam ter uma postura de forte hostilidade em relação a mulheres.
> Participante 3 - Mulher calada ou tá lavando louça ou dando pro marido depois do trabalho
> Participante 4 - Ou apanhando
> Participante 5 - Mulher é um ser maligno e também é o ser mais falso da humanidade. Hoje em dia elas são verdadeiros demônios
> Outro membro da comunidade fala de forma obsessiva sobre a irmã, aborda a possibilidade de abusá-la e pede que façam um pornô deepfake dela. Alguns também se gabam de apreciar gore — vídeos e fotos de situações de extrema violência, quase sempre reais.[5]

São fotos e vídeos de mulheres nuas sem seu consentimento ou ciência caracterizam violência de gênero, isso é crime. Recortar ou editar vídeos de políticos para gerar duplo sentido ou mesmo para simular que determinada pessoa disse ou prometeu alguma coisa para sua base eleitoral, é violência política e também é crime. Se passar por outra pessoa por meio de manipulação de imagem é falsidade ideológica, outro crime. Simular que pessoas indígenas estão ateando fogo de forma criminosa em territórios na Amazônia também é crime. Bem como falsear uma foto que tenha o Movimento Sem Terra dizendo coisas que não estão alinhadas com seus princípios. Não duvide (ou duvide!): é deepfake. Os tipos mais comuns são: substituição de um rosto por outro, clonagem de voz, adulteração da região da boca (BRASIL, 2024).

Muitas deepfakes são tão bem feitas que não nos restam dúvidas, tal qual uma lenda urbana multiplicada tantas vezes ao ponto de virar verdade. No âmbito do *business* das *deepfakes*, o Projeto Amazônia Livre de Fake mapeou o ciclo de desinformação, e de *fakes* em geral, em três estados amazônicos: Mato Grosso, Amazonas e Pará. Entre as análises feitas no relatório, estão analisados os gastos com mídia e um dos achados destacados é a compra de serviço de licença de software por parte de parlamentares que ocupam cadeiras nas Assembléias Estaduais, Câmaras Municipais e políticos que estão em outros órgão públicos, como o chat GPT, Adobe Cloud, Envato Elements e Google Gsuite. Sabemos que alguns serviços são usados para essa produção de conteúdos “automatizados e tecnologias de geração de rostos digitais”[6]. Quando da Covid-19,

[5] Trecho de reportagem da BBC Brasil. Disponível em: https://www.bbc.com/portuguese/articles/cv2g3wvgldlo
[6] PROJETO AMAZÔNIA LIVRE DE FAKE. Relatório do Projeto Amazônia Livre de Fake. 2024. Disponível em: https://amazonialivredefake.intervozes.org.br/wp-content/uploads/2024/06/Relatorio-Amazonia-Livre-de-Fake-2024.pdf

muitas deepfakes foram criadas, inclusive com pessoas que morreram na pandemia[7] e até mesmo chefe de torcida organizada de um famoso Clube de Belém foi alvo de manipulação de vídeo[8].

E como podemos nos proteger? Pela informação. Sempre que tivermos dúvidas sobre determinada informação ou manipulação, o passo correto é verificar em fontes sérias, que usem de sensacionalismo em suas manchetes e abordagens jornalísticas. O Supremo Tribunal Federal criou um Guia Ilustrado contra as Deepfakes[9], que explica como identificar esses recursos em vídeo, áudio e imagem; aponta canais de denúncia; apresenta o programa nacional de combate à desinformação. É preciso mobilizar a sociedade para o combate a esses crimes e assim mitigar os danos causados às vítimas, bem como evitar que mais pessoas sejam atingidas.

Por fim, o cuidado digital também é um caminho para a nossa proteção e para combater essas práticas predatórias oferecidas por algumas plataformas de inteligência artificial, pensando em orientar nossa sociedade de uma forma carinhosa e fazendo isso usando nossos hábitos alimentares como referência, a Rede Transfeminista de Cuidados Digitais (RTCD) criou um site chamado PRATO DO DIA: a refeição dos cuidados digitais: https://pratododia.org, no qual é possível compreender também sobre os funcionamentos sobre algumas das plataformas que mais usamos em nossa rotina, o acesso e a reprodução são livres.

---

[7] MORAES, Mauricio. É falso que idosa filmada em maca foi resgatada viva do necrotério após hospital comunicar morte por Covid-19. 2020. Disponível em: https://lupa.uol.com.br/jornalismo/2020/05/07/verificamos-idosa-filmada-resgatada-viva-hospital-morte-covid-19

[8] Presidente da Trovão Azul denuncia vídeo falso gerado por Inteligência Artificial. 2025. Disponível em: https://parawebnews.com/presidente-da-trovao-azul-denuncia-video-falso-gerado-por-inteligencia-artificial/

[9] BRASIL. Guia Ilustrado Contra as Deepfakes. Supremo Tribunal Federal; Data Privacy Brasil. Brasília: STF, Coordenadoria de Combate à Desinformação, 2024. Disponível em: https://portal.stf.jus.br/desinformacao/doc/Guia%20ilustrado%20Contra%20DeepFakes_ebook%20(1).pdf

A LENDA DA MOÇA DO TÁXI

Cerca de 22 horas. Raimundo dirigia pela Av. Independência em direção ao Largo de Nazaré (praça Justo Chermont). Quase chegando a travessa 14 de Março, uma jovem fez o sinal para o táxi, Raimundo parou.

Por favor, deixe-me na Av. José Bonifácio de frente ao cemitério de Santa Isabel.

O motorista seguiu para o endereço dado. Ao chegar, a moça falou:

Estou sem dinheiro trocado, mas senhor faça favor de cobrar amanhã, nesse endereço.

Entregou um pedaço de papel a Raimundo, no qual estava escrito seu nome e seu endereço, fale com meu pai. Meio contrariado, o motorista segurou o papel que lhe era oferecido.

Mas logo hoje, que a renda está fraca. Pensou!

O taxista, meio aborrecido, pergunta:

Mas afinal, onde a senhora quer ficar?

Depois lhe direi. Não se aborreça comigo, por favor. O senhor cobrará depois o quanto quiser. No momento não vou a lugar nenhum. Estou apenas passeando. Sabe? Hoje é meu aniversário, e meu pai, todos os anos me dá de presente uma volta de táxi pela cidade. Ele pagará quanto o senhor pedir.

Afinal tudo é possível, pensou o motorista acompanhando a turista em sua própria cidade fazendo ele mesmo turismo forçado. Olhava toda cidade.

Era como se só saísse uma vez durante o ano. Ela relata as modificações nos vários bairros de Belém. Depois de tê-lo percorrido, pediu para ser deixada no cemitério de Santa Isabel.

Pode deixar-me aqui — disse a moça. Agora vou andar um pouco a pé. Muito obrigado pro tudo, principalmente pela sua paciência.

Muito bem moça — disse o motorista. — Feliz aniversário. Mas... e a corrida?

Ah! sim, desculpe, ia esquecendo. Cobre com meu pai neste endereço e diga-lhe que é meu presente de aniversário, muito obrigada de novo. Até logo.

No dia seguinte, pela manhã, o motorista foi ao endereço dado pela moça.

Bom dia! mora aqui o senhor fulano?

Bom dia! sim, mora, o que deseja?

Vim cobrar uma corrida de táxi da filha dele.

Mas ele não tem filha, ou melhor nós não temos filha porque sou a esposa dele.

Não é possível!

Ora não tenho porque lhe mentir...

Mas ontem uma moça assim, assim, correu toda a cidade em meu carro e me mandou cobrar aqui, dizendo ser sua filha e que o passeio era presente de aniversário.

A senhora ficou assustada.

Olhe já lhe disse que não temos filhas...

Nesse momento, pela porta entreaberta, o motorista nota o retrato de uma moça e, apontando-o diz:

A moça é aquela ali!

A senhora começa a chorar.

Não é possível... aquela moça era nossa filha... Mas ela já morreu a tanto tempo... e, realmente, o pai costumava dar de presente uma volta de táxi pela cidade.

O motorista começa a ficar nervoso já não lhe interessa a cobrança da corrida. Só quer esclarecer se a moça que pegou o táxi estava viva ou não. O caso é solucionado pela chegada do pai, que afirma a morte da moça. Levou o motorista ao cemitério, e lá mostrou o túmulo onde o motorista vê um retrato igual ao que havia na casa.

LISTA DE CASOS

Ao final desta publicação, oferecemos uma lista organizada a partir dos casos mencionados pelo autores em seus artigos. Esta lista foi sistematizada por Lauro Accioly. A iniciativa é baseada no mapeamento Danos e Discriminação Algorítmica, composto de casos, reportagens e reações sobre discriminação algorítmica desde 2010. A iniciativa de Tarcízio Silva é uma linha do tempo pode ser explorada de uma maneira interativa no site da Desvelar.

LISTA DE CASOS MENCIONADOS PELOS AUTORES

1. **Zelensky em deepfake de rendição. Vídeo manipulado.** Durante o conflito com a Rússia, o presidente ucraniano Volodymyr Zelensky estaria supostamente pedindo às tropas ucranianas que se rendessem, em uma tentativa de desinformação e desestabilização psicológica em tempos de guerra **(p. 22)**.

2. **Michelle Obama em postagem preconceituosa. Vídeo manipulado.** O rosto de Michelle Obama foi utilizado para criar um "meme" com conotações racistas, sexistas e transfóbicas, refletindo o uso da tecnologia para ataques misóginos e racializados contra figuras públicas negras **(p. 31)**.

3. **Vinicius Júnior profere ofensas. Vídeo manipulado.** O jogador Vinicius Jr. teve sua imagem manipulada digitalmente em um vídeo com informações falsas e distorcidas, o que gerou desinformação e ataques injustificados ao atleta nas redes sociais **(p. 33)**.

4. **Tabata Amaral em deepfakes de cunho sexual. Vídeo manipulado.** Houve a circulação de vídeos deepfake com conteúdo sexual falso envolvendo a deputada federal Tabata Amaral, usados para difamar e intimidar mulheres em espaços de poder político **(p. 46)**.

5. **Marcos Rocha, governador de Rondônia, em imagem abraçando Lula. Imagem manipulada.** O conteúdo foi utilizado para espalhar informações falsas sobre o governador Marcos Rocha, exemplificando o uso de manipulações digitais no contexto político-eleitoral brasileiro **(p. 48)**.

6. **Renata Vasconcellos. Vídeo manipulado.** A jornalista teve sua imagem manipulada em vídeo usado para fins de desinformação política, criando falas e posicionamentos inexistentes atribuídos a ela **(p. 49)**.

7. **Lula e Alexandre de Moraes em beijo forjado. Vídeo manipulado:** um deepfake simulando um beijo entre o presidente Lula e o ministro Alexandre de Moraes foi disseminado como forma de ataque político e deslegitimação moral **(p. 50)**.

8. **Descredibilização do ativista Raull Santiago. Imagem manipulada**. Um ativista brasileiro foi alvo de deepfake criado para descredibilizá-la, mostrando como a tecnologia pode ser instrumentalizada para silenciar e deslegitimar vozes dissidentes **(p. 55)**.

9. **Celebridades internacionais em deepnude não consentida. Vídeos manipulados.** Deepfakes inseriram rostos de celebridades internacionais em vídeos pornográficos sem consentimento, violando direitos de imagem e integridade pessoal **(p. 60)**.

10. **Barack Obama em discurso falso e satírico feito pelo humorista Jordan Peele. Vídeo manipulado.** Um vídeo deepfake de 2018, que foi amplamente divulgado, mostrava Barack Obama fazendo declarações que nunca ocorreram. Na época, o conteúdo revelava o potencial da tecnologia para falsificação de discursos políticos com aparência de autenticidade **(p. 60)**.

11. **Donald Trump com autopromoção e ataques. Vídeo manipulado.** Deepfakes foram utilizados nas redes sociais de Donald Trump tanto para sua autopromoção quanto para atacar adversários políticos, reforçando estratégias populistas e de manipulação de narrativas **(p. 63)**.

12. **Adolescentes realizam bullying com deepfakes.** casos de estudantes utilizando deepfakes para zombar, humilhar ou atacar colegas e professores, promovendo práticas de cyberbullying e violência digital entre jovens **(p. 41)**.

13. **Professora é vítima de deepnude pornográfico. Imagem manipulada.** Profissional foi alvo de montagem com o uso de deepnude, ferramenta que cria falsas imagens pornográficas, violando sua privacidade e gerando impacto psicológico e profissional significativo **(p. 42)**

14. **Indústria de games e uso de celebridades. Imagem manipulada.** Empresas de jogos utilizaram deepfakes com rostos de celebridades para campanhas de divulgação, levantando debates sobre ética, propriedade de imagem e consentimento na publicidade digital **(p. 67)**.

15. **Celebridades brasileiras e uso indevido de imagem.** Rostos de celebridades brasileiras foram usados em deepfakes diversos — de contextos desinformacionais a

comerciais — sem autorização, reforçando a banalização da violação de imagem no ambiente digital **(p. 68)**.

16. **Giovanna Ewbank e conteúdo publicitário enganoso.** Um vídeo deepfake da atriz Giovanna Ewbank foi usado para disseminar conteúdo falso e enganoso, exemplo de como celebridades são alvo frequente dessas manipulações **(p. 69)**.

Justiça racial, IA e tecnologias digitais

ISBN 978-65-985781-1-4