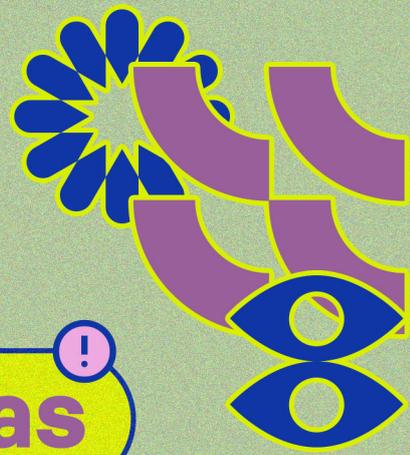


# Deepfakes Pornográficas

André Lucas Fernandes, Carolina Branco, Clarissa Mendes,  
Luana Batista, Lunara Santana, Raquel Saraiva



## I. O que são

*Deepfakes*, pornográficas ou não, são falsificações resultantes de programas de edição de imagem e vídeo que possuem (ou fazem uso de) um algoritmo de inteligência artificial em seu núcleo. Mas, antes de falarmos do que são *deepfakes* pornográficas, é preciso entender o funcionamento de sua tecnologia-base a fim de termos uma maior compreensão sobre como essas pornografias falsas são criadas e o que é possível fazer sobre o cenário de sua produção.

**Inteligência artificial** é uma expressão ambígua que se refere ao campo interdisciplinar acerca da representação e *mimesis* de atividades humanas ditas “inteligentes” (ou complexas, ou seja, que envolvem um raciocínio elaborado ou uma habilidade cognitiva humana) através de uma lógica matemática que possa ser executada computacionalmente por um programa, um robô, um dispositivo eletrônico, etc<sup>1</sup>. Isso pode ser obtido de diversas formas, mas para o escopo deste documento, vamos focar numa abordagem indutiva, fortemente estatística, chamada **aprendizagem de máquina** (*machine learning*, em inglês), que busca realizar essa atividade “inteligente” a partir de padrões estatísticos extraídos de um conjunto de dados.

A aprendizagem de máquina se baseia na ideia de que inferências e correlações podem ser feitas a partir de uma coleção de dados. Ao analisar esses dados, um

<sup>1</sup> “Quando máquina ‘aprende’, isso não quer dizer que máquina tem um cérebro feito de metal. Quer dizer que a máquina se tornou mais acurada em realizar uma única, específica tarefa de acordo com métricas definidas por uma pessoa” (BROUSSARD, Meredith. *Artificial Unitelligence*, p. 116, tradução livre)



algoritmo de aprendizagem de máquina (chamado também de **modelo**) pode ser treinado e testado para se aperfeiçoar em tarefas como classificação de imagens, reconhecimento de fala, detecção de objetos, processamento de linguagem natural, entre outras.

Até a primeira metade da década de 2010, essas atividades eram mais descritivas, discriminatórias e inferenciais. Porém, com (1) o crescimento do poder de processamento computacional; (2) a possibilidade de serviços de processamento de larga escala de dados ser feita em nuvem; e (3) o aumento de dados textuais, imagéticos e de diferentes modais coletados na internet, esses modelos passaram a ser criados para tarefas de geração de novos dados a partir dos anteriores, criando imagens, áudios, textos, etc. Esse crescimento expressivo ocorreu no final dos anos 2010 e início da década de 2020, trazendo à tona o campo da **inteligência artificial gerativa**, ou **generativa**, que já existia há décadas como um campo experimental e teórico, mas que só veio a ocupar um espaço no mercado recentemente.



Com o intuito de separar esses algoritmos mais comuns e antigos dos mais atuais, mais robustos, mais complexos (e mais opacos), o termo **aprendizagem profunda**, em inglês **deep learning**, foi cunhado para representar o último grupo. A ideia de profundidade (*deep*) vem da quantidade de camadas e etapas de processamento dos dados que esses modelos empregam. Essa complexidade maior faz com que esses modelos necessitem de um grande poder computacional e uma grande quantidade de dados para servir de base de aprendizagem, o que permite uma performance melhor (calculada a partir de um conjunto de métricas). Esses modelos também são muito pouco explicáveis, o que significa dizer que é difícil determinar os elementos que justificam uma decisão tomada pelo algoritmo.

Relacionando-se também a esse grupo de algoritmos, em 2017, o termo **deepfake** foi cunhado para definir programas de computador que usavam esses modelos profundos



de aprendizagem de máquina para falsificar imagens, áudios e vídeos<sup>2</sup>. O nome veio do perfil de um usuário da plataforma Reddit que publicou um vídeo falso da atriz Gal Gadot e criou um sub-fórum específico para divulgação dessas falsificações<sup>3</sup>. As falsificações são feitas por meio de imitações e edições da voz e do rosto de pessoas que não têm nada relacionado a aquele material, quase sempre com o intuito de divulgação na internet para fins diversos, dos mais satíricos e criativos, como os feitos por Bruno Sartori<sup>4</sup>, aos mais danosos, propagando mentiras e direcionados à vingança.



Chama-se, ademais, **deepfake pornográfica** a criação desses produtos falseados que colocam as vítimas em um contexto situacional de pornografia<sup>5</sup>, geralmente direcionada a atacar minorias como mulheres, pessoas trans e crianças, como será apresentado no texto - para efeitos de diferenciação deste fenômeno que tem como requisito a não consensualidade, entendemos por, exemplificado pelo caso mais famoso, chamar estas *deepfakes* de *deepnudes*.

Com o passar do tempo, a situação tornou-se ainda mais crítica com o surgimento do aplicativo *DeepNude*, que criava uma falsificação ao remover as roupas de mulheres em imagens, e muitos outros que vieram após o desligamento dele e que podem ser facilmente encontrados na internet e usados de forma gratuita, como é o caso do *Nudify.online*, *Undress.app*, *PornJoy AI*, entre outros. Até mesmo aplicações de inteligência artificial gerativa que não foram pensadas para *deepfakes*, como o Sora<sup>6</sup>, da OpenAI, possuem o risco de serem usados para esse fim.

Abordagens usadas para enfrentamento dessa tecnologia se concentram no campo legal e tecnológico (utilizando outros modelos de aprendizagem profunda), mas mere-

<sup>2</sup> **Deepfakes, explained | MIT Sloan**. Disponível em: [Acesse](#). Acesso em: 20 de fevereiro de 2024.

<sup>3</sup> COLE, Samantha. AI-Assisted Fake Porn Is Here and We're All Fucked. **Vice**, 11 dez. 2017. Disponível em: [Acesse](#). Acesso em: 20 de fevereiro de 2024.

<sup>4</sup> SARTORI, BRUNO. **Youtube**, 2023. Página do canal do jornalista Bruno Sartori no Youtube, onde vários vídeos humorísticos e satíricos feitos com *deepfakes* podem ser encontrados. Disponível em: [Acesse](#). Acesso em: 20 de fevereiro de 2024.

<sup>5</sup> A situacionalidade da deepfake pornográfica desaparece e se transforma em uma generalidade quando o produto do modelo é exposto na Internet. Importa notar que boa parte das aplicações que estão envolvidas no crescimento vertiginoso citado neste artigo curto fazem uso de uma plataforma de produção e divulgação massiva de conteúdo online.

<sup>6</sup> MOZELLI, Rodrigo. Sora: novo sistema de IA da OpenAI causa preocupações sobre deepfakes. **Olhar Digital**. Disponível em: [Acesse](#). Acesso em: 20 de fevereiro de 2024.



cem avançar em outras frentes de conhecimento e ação.

## II. Contextualização do Problema e Casos recentes

A criação de *deepfakes* vem crescendo exponencialmente no mundo: em 2023, foram encontrados 95.820 vídeos, um aumento de 550% em relação a 2019. Se esse tipo de mídia tem recebido atenção por seus potenciais riscos políticos, o que os dados têm apresentado revelam outra realidade - 98% dos vídeos *deepfakes* em 2023 são pornográficos; e dentre estes, 99% dos alvos são mulheres, de acordo com estudo do Home Security Heroes<sup>7</sup>.

Se a divulgação não consensual de imagens íntimas já vinha sendo uma preocupação crescente, muitas vezes atrelada a uma lógica de leitura pornográfica de mundo, o que os *deepfakes* pornográficos trazem para o cenário da violência de gênero digital e da violência de gênero facilitada pela tecnologia é que o teor sexual nem precisa estar presente na imagem original para que as pessoas o compartilhem - ele pode ser gerado artificialmente.

Dessa forma, qualquer mulher que tenha imagens suas disponíveis na internet estaria suscetível a sofrer um dano pela construção de um conteúdo falso que passaria a se confundir com o real.

Independente das imagens serem reais ou não, as consequências sobre as vítimas podem ser tão devastadoras quanto a pornografia de vingança, cujo impacto é bem documentado pela literatura. As vítimas de abuso digital têm taxas elevadas de problemas de saúde mental, como ansiedade, depressão, automutilação e suicídio, de acordo com pesquisa realizada por Asher Flynn, professora da Monash University<sup>8</sup>. Os efeitos podem ser sentidos tanto na saúde física como mental, com impactos sobre seu emprego, família e vida social, além de ter um efeito inibidor sobre a liberdade de expressão das mulheres.

Reconhecendo que, de acordo com os “Princípios Feministas da Internet”<sup>9</sup>, o con-

<sup>7</sup> HOME SECURITY HEROES. 2023 State Of Deepfakes: Realities, Threats, And Impact. Disponível em: [Acesse](#). Acesso em: 19 de fevereiro de 2024.

<sup>8</sup> Flynn, A., Powell, A., & Hindes, S. (2021). Technology-facilitated abuse: A survey of support services stakeholders (Research report, 02/2021). ANROWS. [Acesse](#).

<sup>9</sup> All Principles | Feminist Principles of the Internet. Disponível em: [Acesse](#). Acesso em: 19 de fevereiro de 2024.



sumo de conteúdo pornográfico não tem vínculo causal simples com a violência de gênero, e que existem possibilidades de conteúdo pornográfico alternativo que sejam saudáveis e opostos à visão patriarcal dominante, o tema merece uma análise dedicada e interdisciplinar. O propósito central desta análise é elucidar sobre a produção de conteúdo pornográfico de mulheres, empregando ferramentas de inteligência artificial, sem o devido consentimento.

Um dado revelador e preocupante a esse respeito é o de que, nos Estados Unidos, 73% dos homens usuários de pornografia *deepfake* não se sentem culpados a respeito disso (2023, Home Security Heroes). Os principais motivos apontados são: o fato de saber que não é realmente a pessoa; de acreditar que não machuca ninguém; de que seria simplesmente uma versão mais realista da imaginação sexual; e de que não seria muito diferente do pornô tradicional.

O fato de os usuários não entenderem como violência práticas que violam a noção de consentimento e que, de acordo com a literatura, têm efeitos violentos sobre as vítimas, ilustra a enorme lacuna existente sobre o tema e a necessidade de um processo educativo sobre os diferentes tipos de violência de gênero. Também levanta a necessidade de investigar de que forma essa percepção se daria no contexto brasileiro, considerando as nossas particularidades nacionais.



Recentemente, o Brasil tem enfrentado casos alarmantes de *deepfakes* pornográficos, com ocorrências que chamam a atenção, especialmente entre estudantes de escolas localizadas em Recife, Rio de Janeiro e Belo Horizonte. Segundo reportagem da BBC, jovens têm recorrido à inteligência artificial para a criação de montagens pornográficas de colegas<sup>10</sup>. Além disso, figuras públicas também têm se tornado alvo desse tipo de prática. De acordo com o G1, a atriz Isis Valverde teve suas fotos, originalmente postadas em re-

<sup>10</sup> LEMOS, V. **Nudes falsos se multiplicam e preocupam autoridades: “crueldade impensável com meninas tão novas”**. BBC, 16 nov. 2023. Disponível em: [Acesse](#). Acesso em: 19 de fevereiro de 2024.



des sociais, adulteradas e compartilhadas para simular um vazamento de “nudes”<sup>11</sup>. Esses ataques podem acarretar sérias consequências para as vítimas envolvidas, gerando danos emocionais e sociais significativos.

Neste contexto, as autoridades estão empenhadas na investigação desses incidentes, buscando não somente identificar os responsáveis como tomar medidas adequadas para evitar que se repitam, inclusive no âmbito educacional - uma vez que o ambiente escolar tem sido frequentemente cenário desses episódios. Como é possível perceber, a disseminação dessa prática destaca a urgência de uma investigação mais aprofundada. Diante disso, o Instituto de Pesquisa em Direito e Tecnologia do Recife (IP.rec) iniciou um projeto, dentro de seus esforços de atuação sobre o tema da inteligência artificial, que propõe não apenas a investigação do panorama nacional, mas também a produção de conteúdos educativos sobre o tema. A abordagem visa situar a prática de *deepfakes* dentro do escopo da violência de gênero, destacando a importância de informar sobre o tema e suas consequências, evitando usos danosos.

Posta na mesa a urgência de uma investigação sobre o cenário, este artigo resumido pretende analisar algumas iniciativas sobre o tema no Brasil, notadamente no aspecto legislativo e em medidas iniciais para produzir conteúdo educativo situando a prática dentro do escopo da violência de gênero.

### III. O que pode ser feito a respeito

Para os objetivos específicos deste estudo, pretendemos, primeiro, analisar três Projetos de Lei, propostos em 2023 como uma clara reação aos casos citados anteriormente. Além disso, numa segunda parte, abordaremos outros caminhos que extrapolam os projetos mencionados.

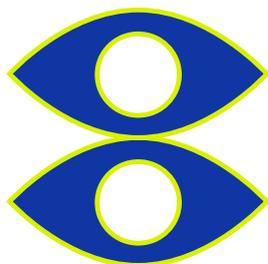
Os PLs escolhidos são: (a) o de número 5394, proposto pela Deputada Erika Kokay, do PT/DF; (b) o 5467, proposto pela Deputada Camila Jara, do PT/MS e, (c) o 5342, proposto pelo Deputado Marcelo Alvaro Antônio, do PL/MG. Todos os três textos versam, em linhas gerais, sobre a criação de tipos penais para criminalizar a prática relacionada às *deepfakes* pornográficas.

<sup>11</sup> Isis Valverde registra ocorrência após aparecer nua em montagens compartilhadas em redes sociais. G1, 26 out. 2023. Disponível em: [Acesse](#). Acesso em: 19 de fevereiro de 2024.



## O debate no legislativo (Projetos de Lei sobre o tema)

O PL 5394/23, proposto pela Dep. Erika Kokay, pretende “criminalizar a adulteração, montagem ou modificação de fotografia, vídeo ou qualquer outra forma de representação visual relativos à intimidade da pessoa, por meio de Inteligência Artificial”. A justificativa do projeto está atrelada a um argumento acerca dos “perigos” de tecnologias com “ares de modernidade” e que, usadas de forma maléfica, podem gerar danos e práticas de crimes cibernéticos. Citando os casos apontados anteriormente, o texto visa criminalizar a prática como forma de combate direto às práticas ilícitas. O tipo penal descrito no texto estabelece a seguinte hipótese de incidência:



Art. 216-C Simular nudez de pessoa ou a participação em cena de sexo explícito ou pornográfico mediante adulteração, montagem ou modificação de fotografia, vídeo ou qualquer outra forma de representação visual por meio do uso de tecnologias baseadas em Inteligência Artificial:

Pena – detenção, de 12 (doze) meses a 3 (três) anos, e multa.

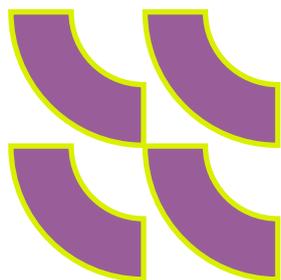
§ 1º Incorre nas mesmas penas quem adquirir, armazenar, disponibilizar, compartilhar, transmitir, distribuir, publicar ou divulgar por qualquer meio, inclusive por meio de sistema de informática ou telemático, fotografia, vídeo ou outro registro que simule nudez ou a participação de pessoa em cena de sexo explícito ou pornográfica, produzida mediante o uso de tecnologias baseadas em Inteligência Artificial.

O PL 5467/23, proposto pela Dep. Camila Jara, reconhecendo em texto legal que “divulgação de conteúdo falso sexual configura violência doméstica e familiar e criminaliza a divulgação de registro falso não autorizado de conteúdo com cena de nudez ou ato sexual ou libidinoso”, propõe a criação de um novo tipo penal e a alteração da Lei Maria da Penha. Eis os textos propostos, de forma expressa:

Art. 2º O inciso II do caput do art. 7º da Lei nº 11.340, de 7 de agosto de 2006 (Lei Maria da Penha), passa a vigorar com a seguinte redação:

“Art. 7º.....  
.....

II - a violência psicológica, entendida como qualquer conduta que lhe cause dano emocional e diminuição da autoestima ou que lhe prejudique e perturbe o pleno desenvolvimento ou que vise degradar ou controlar suas ações, comportamentos, crenças e decisões, mediante ameaça, constrangimento, humilhação, manipulação, isolamento, vigilância constante, perseguição contumaz, insulto, chantagem, violação de sua intimidade, ridicularização, divulgação de conteúdo falso sexual, exploração e limitação do direito de ir e vir ou qualquer outro meio que lhe cause prejuízo à saúde psicológica e à autodeterminação;” (NR)



Art. 2º Acrescenta o artigo 216-C no Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), com a seguinte redação:

Art. 216-C. Divulgar conteúdo falso sexual, por qualquer meio, com cena de nudez ou ato sexual ou libidinoso sem autorização da vítima:

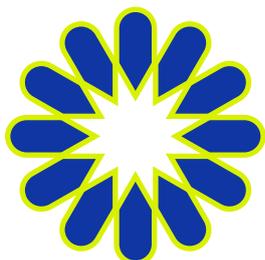
Pena - detenção, de 6 (seis) meses a 1 (um) ano, e multa.

§1º Se o crime for praticado contra menor de idade: Pena - reclusão, de 1 (um) a 3 (três) anos, e multa.

I - A plataforma digital que reproduzir deverá excluir imediatamente o conteúdo, sob pena de multa.

A alteração na Lei Maria da Penha cria nova hipótese de violência doméstica (violência psicológica) nos casos de divulgação de conteúdo falso sexual. Já o novo tipo penal é a própria divulgação de conteúdo falso sexual, com agravante quando o crime for praticado contra menor de idade. O texto, que não dialoga com o Marco Civil da Internet, cria hipótese de remoção imediata de conteúdo pelas plataformas e redes sociais.

O PL 5342/23, proposto pelo Dep. Marcelo Alvaro, situado em casos internacionais e nacionais, busca tipificar



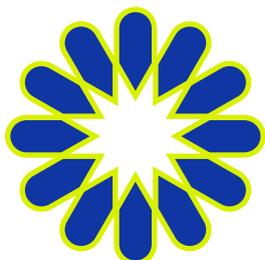
“[...] como Porno Fake, criar, oferecer, trocar, disponibilizar, transmitir, distribuir, publicar ou divulgar por qualquer meio, inclusive por meio de sistema de informática ou telemático, fotografia, vídeo ou outro registro que contenha fotos e vídeos de pessoas em cenas de sexo ou nudez, criada por meio de inteligência artificial, sem prévia autorização:

Pena - reclusão, de 6 (meses) a 1 (um) ano.”

Art. 2º A pena é aumentada de 2/3 (dois terços) a 3/3 (três terços) se as imagens criadas por meio de inteligência artificial forem de menores de idade.

Art. 3º A pena é aumentada de 1/3 (um terço) a 2/3 (dois terços) se as imagens forem utilizadas para chantagear a vítima.

Art. 4º Nos crimes descritos nos artigos 1, 2 e 3, proce-se (sic) mediante ação penal pública.



As propostas legislativas, optando por uma postura criminal, podem se mostrar inefetivas ou criadoras de externalidades negativas como em outros temas relativos ao debate sobre moderação de conteúdo na Internet. O tema aparece aqui potencializado pelo tipo de conteúdo e pela adoção de inteligência artificial na geração do mesmo.

A adoção de uma solução penalizadora parece motivada por dois elementos centrais, típicos ao processo legislativo atual: uma resposta simbólica a uma questão social



sensível, cujos danos são de relevo; ausência de debate entre especialistas e diferentes setores para o aperfeiçoamento da lei; ausência de diálogo normativo com fontes que toquem na mesma matéria - duas Leis e um Processo Legislativo, em andamento, merecem ser citados: o Marco Civil da Internet, a Lei Geral de Proteção de Dados e o PL 2338/23, que busca regular a adoção da IA no Brasil.

Dados relevantes apontam para uma abordagem mais madura e sincrônica das propostas: enquanto o PL 5467 parece acertar em alterar o tipo penal ligando-o à definição de “violência psicológica” - que melhor denota o que procede no uso de *deepfakes*

pornográficas, o PL 5342 cuidou de detalhes relevantes, tornando o tipo completo, quando do apontamento do tipo de ação (com acerto, pública) e das hipóteses de qualificação.



## IV. Conclusões Preliminares

Em que pese não existir um desaconselhamento prévio ao uso da criminalização de uma conduta cujo bem jurídico ofendido é de extrema relevância e que, portanto, demanda uma efetiva proteção do ordenamento jurídico, alguns pontos de melhoria para início de debate podem ser pontuados:

**A.** Os textos propostos podem pecar ao dialogar com um tema extremamente complexo, vendo uma solução de causalidade na penalização quando da tentativa de enfrentar uma questão multifatorial;

**B.** Os textos precisam dialogar, em suas justificativas, com legislações vigentes. O Marco Civil da Internet já permite à vítima de conteúdo íntimo divulgado sem consentimento acionar a justiça e exige das empresas que atuam na disseminação de conteúdo uma postura de prontidão. O tema difere de forma clara da remoção proativa de conteúdo, uma interpretação possível diante do texto do PL 5467;

**C.** A solução penalizadora pode vir acompanhada de uma discussão na esfera cível, permitindo maior agilidade no manejo de tutelas para remoção de conteúdo a partir da denúncia da vítima;

**D.** De acordo com a LGPD, os dados de pessoas identificáveis são protegidos pela lógica dos dados pessoais em geral, atrelados, no caso, a serem dados pessoais sensíveis. Assim sendo, a interpretação em diálogo normativo se impõe e deve estar presente nas propostas de enfrentamento ao problema pelos legisladores;

**E.** A remoção do conteúdo pressupõe o uso acabado de uma solução de IA na produção da chamada *deepfake* pornográfica - ou *deepnude*. Ou seja, não se aplica ao caso conteúdos produzidos com o consentimento das partes e não se analisa nada anterior ao evento a ser posto como crime;

**F.** O debate deste uso em específico da inteligência artificial, mesmo que possa ser feito de forma detida, merece um diálogo com medidas gerais para otimizar a governança de modelos em todo ciclo de vida, com base em padrões éticos globais e nacionais. Ou seja, é preciso pensar não apenas no uso ilícito de uma ferramenta, mas nas medidas gerais que possam desestimular o uso ilícito. Entre as medidas, situar o grau de risco elevado no manejo de *deepfakes* em contextos específicos como o da produção de conteúdos por-



nográficos falsos;

**G.** A formação e esforços no entorno da formação de governanças éticas, com a participação de profissionais diversos deve ser incentivada em todos os processos legislativos acerca da inteligência artificial, sob pena de não se resolver o problema do uso danoso, convertido em crime;

**H.** A propositura de regulação temática pelo uso pode vir acompanhada de outros comandos normativos a fim de promover um programa educativo no uso da inteligência artificial, em contextos escolares, universitários e sociais gerais. Parte considerável dos casos que têm surgido envolvendo *deepfakes* pornográficas no Brasil tem se dado no ambiente escolar, tendo sido criadas por menores de idade, e portanto escapam ao escopo dos Projetos de Lei. A inserção do tema em agendas da educação e saúde, orientando outras dimensões políticas que não a política criminal podem resultar num efeito preventivo maior do que o efeito proativo-criminal.

**I.** Em termos de boas práticas legislativas, a produção de normas jurídicas precisa incorporar a participação de interesses diversos e multissetoriais de modo a alcançar a participação pública significativa - instrumentos de consulta pública e audiências públicas são apenas um piso e devem ser incentivados.