IMPLICAÇÕES POLÍTICAS, SOCIAIS E ÉTICAS DA POPULARIZAÇÃO DE *DEEPFAKES* POR MEIO DE SOFTWARES LIVRES

Bernardo Okazaki Kehdy, Gabriel Jerônimo Tavares, Gabriel Teixeira Lara Chaves, Wagner Felipe Patricio Maia

Universidade Federal de Minas Gerais

Resumo: Frente ao recente sucesso de Redes Adversariais Generativas e a popularização de suas aplicações por meio de softwares livres, torna-se necessário discutir quais são as implicações de um novo tipo de produção audiovisual cunhado *deepfake* na esfera política e social, analisando seu impacto por meio de exemplos concretos
Palavras-chave: *Deep Fake*, Software Livre, Fake News, Política, Indústria Pornográfica.

1. Introdução

Desde sua concepção em 1956 como disciplina acadêmica, o campo de Inteligência Artificial, definido em ciências da computação como o estudo de ‘’agentes inteligentes’’, passou por altos e baixos de financiamento e produção acadêmica, notavelmente atingindo um vale expressivo durante o ‘’Inverno de Inteligência Artificial’’ durante a década de 70.

A recente ressurgência do campo possibilitou o sucesso de Redes Adversariais Generativas (“GAN’s”), sistemas de aprendizado supervisionado onde duas redes trabalham em conjunto para gerar imagens que se aproximam de representações fiéis do mundo real, popularizou uma nova forma de produção audiovisual, a chamada indústria *deep fake.* Unindo os termos “deep learning” e “fake’’, a palavra *deep fake* é usada para se referir, principalmente, aos vídeos criados por GANs cuja produção se massificou a partir da circulação desses métodos em forma de software livre. Na prática, a produção de um vídeo *deepfake* se dá por meio de coleta de dados de um rosto humano, por meio de fotos e vídeos tradicionais existentes, e a transposição dessa informação para uma outra imagem existente, geralmente de uma pessoa diferente, de forma que, quando bem feito, emula-se com perfeição o humano de referência na situação desejada. Dessa forma, as inúmeras horas de esforço humano e técnica necessárias para falsificar uma imagem dessa maneira se reduzem à *hardware* bom, dados suficientes, e conhecimento técnico básico.

2. Indústria Pornográfica

Captura de Tela 1: DeepFake Nicolas Cage

Fonte: página de *Reddit* do usuário “derpfakes” (autor)

Na imagem acima, o usuário do site *Reddit* transpôs o rosto do ator Nicolas cage no corpo da atriz Amy Adams, em uma cena de *Homem de Aço* (2013). Construções cômicas como essa, no entanto, não foram as mais observadas. O mesmo site, até dia 07 de Fevereiro de 2018, possuía um sub-espaço de discussão intitulado “r/Deep Fakes”, cujo objetivo era produzir conteúdo pornográfico involuntário, inserindo rostos de atrizes famosas em vídeos pornôs. O subreddit tinha mais de 50,000 seguidores nos meses que antecederam seu desligamento pela comunidade gerenciadora do site.

Paralelamente, o site *adultdeepfakes.com* desenvolveu um repertório digital imenso de deep fakes pornográficos, composto quase que inteiramente por celebridades mulheres. A plataforma possui um botão de ‘’Make vídeo’’ de forma a explicitamente incentivar a contínua produção desse conteúdo. Esse tipo de imagem é produzida sem consenso do agente referencial e do agente cujo corpo é utilizado, e apesar de não ser absolutamente inédito - já que pornografia falsa existe há muito tempo - sua acessibilidade propõe um problema sério, como ressalta Megan Farrokhmanesh, escritora para *The Verge*: “Pornografia de photoshop já é uma ferramenta de assédio recorrentemente empregada contra mulheres na internet; um vídeo torna a violação muito mais ativa e difícil de identificar como falsa, além de ser mais fácil de produzir”. A quantidade de imagens de indivíduos disponíveis online, principalmente devido ao estabelecimento de redes sociais como o Instagram e a popularização de compartilhamento de “selfies” juntamente da crescente eficiência de algoritmos para falsificar vídeos potencializa a possibilidade de assédio que Megan descreve.

Quanto mais realista os vídeos se tornarem mais a diferença entre seu alcance e o conhecimento de que tais conteúdos são forjáveis se acentuará, fazendo com que eles sejam tidos como verdadeiros pela maioria de seus consumidores, configurando a situação que Adam Dodge, diretor de uma agência

de violência doméstica na Califórnia, descreve ao dizer que “Muitas pessoas não sabiam que essa tecnologia existia, muito menos que ela poderia ser usada como uma arma contra cidadãos comuns. Aí reside grande parte do problema [...]”.

Instituições e companhias foram pegas tipicamente despreparadas, sendo que sites pornográficos demoraram vários meses para se pronunciar sobre o nova “categoria” e a maioria dos países não possui legislação que discute eficientemente pornografia não consensual. A advogada Carrie Goldberg argumenta que pela ausência de leis específicas, advogados civis precisam ficar ‘’criativos’’, usando violação de direitos autorais e ações de difamação para trazer justiça à seus clientes.

3. Legislação brasileira sobre crimes cibernéticos (*deepfakes)*

Em 2004 entrou em vigor a primeira convenção sobre cibercrimes, que ficou conhecida como convenção de Budapeste, é um tratado internacional de Direito Penal e Direito Processual Penal firmado principalmente por países da Europa a fim de colaborarem entre si de forma harmônica a fim de combater os crimes praticados na Internet.

O Brasil resiste em aderir à convenção, sob a alegação de que não seria bom para o país participar de um acordo do qual não tenha ajudado a elaborar. Mas, apesar de manter este discurso, entrou em vigor no Brasil em 2013 as leis 12.737/12 e 12.735/12 , lei Carolina Dieckman e lei Azeredo, como ficaram conhecidas, que claramente possuem forte influência, para não dizer que são cópias, de artigos contidos na Convenção de Budapeste que data de 2004.

Isso significa, que o Brasil, por motivos de vaidade política, demorou 8 anos mais para perceber e legislar o que outros Estados já haviam visto como problema social tentaram uma união de interesse global para combater os cybercrimes, pois como foi explicado anteriormente o criminoso age dentro do cyberspace que não possui fronteiras.

Procurando coibir a prática dos crimes informáticos, foi sancionada, no ano de 2012, a lei 12.735, popularmente conhecida como Lei Azeredo. Oriunda do projeto de lei da Câmara dos Deputados de nº 84 de 1999, a proposição visava à inclusão de diversas tipificações no Código Penal, que não foram aprovadas.

A Lei Carolina Dieckmann é como ficou conhecida a Lei Brasileira 12.737/2012, sancionada em 30 de novembro de 2012, que promoveu alterações no Código Penal Brasileiro (Decreto-Lei 2.848 de 7 de dezembro de 1940), tipificando os chamados delitos ou crimes informáticos.

Além das Leis já criadas, existem outras propostas circulando pelo legislativo brasileiro para serem promulgadas, como as citadas a seguir:

O Projeto de Lei 5.555/2013, do deputado João Arruda (PMDB-PR) prevê alteração da Lei nº 11.340/06 no intuito de aproveitar a proteção especializada para combater a prática da pornografia não consensual e condutas ofensivas contra a mulher na internet.

O Projeto de Lei 6.630/2013 do deputado Romário (PSB-RJ) visa acrescer um artigo ao código penal para tipificar “a conduta de divulgar fotos ou vídeos com

cena de nudez ou ato sexual sem autorização da vítima” com o título de “Divulgação indevida de material íntimo”.

Portanto, o Brasil caminha em passos lentos na questão legislativa em relação aos crimes eletrônicos, o ordenamento jurídico passa por permanente atualização, porém lenta , surgem novos casos, novas condutas, novos valores forçando o direito a adquirir um dinamismo maior do que se tinha conceituado no passado a fim de que ele se mantenha aderente à realidade social. Fica também o alerta de que tanto usuários, como os provedores de aplicação e a sociedade em geral precisam agir de forma célere e em conjunto, unindo direito e tecnologia, para combater as *deep fake news* e resguardar os direitos dos usuários, inclusive das mulheres, para que a Internet não se torne um campo de desinformação e frequente violação de direitos.

4. Acessibilidade como software livre

DeepFakes e Softwares livres (uma classe de programas, os quais permitem ao usuário distribuir, modificar e estudar este de acordo com sua necessidade) são dois conceitos intrinsecamente conectados. Essa relação pode ser apontada pelo fato de que ao se analisar fatores impulsionadores para o desenvolvimento e aprimoramento das técnicas e softwares atualmente existentes para criação de *DeepFakes*, pode-se destacar a filosofia do Software Livre, que tem o intuito de promover uma comunidade de usuários altamente colaborativa em que liberdade de estudar o funcionamento de um programa, modificá-lo em busca de melhorias e tornar estas modificações públicas, traga benefício à todos. Tal hipótese é sustentada principalmente pelo ambiente na qual a tecnologia foi desenvolvida e aprimorada, fóruns como *DeepFakes* e CelebFakes hospedados no Reddit. Isso porque este ambiente foi de fundamental importância para que o trabalho colaborativo entre usuários possibilitasse o aperfeiçoamento dos modelos para troca facial, tornando-os mais convincentes, bem como a simplificação do uso da técnica a partir da criação do FakeApp, aplicativo que permite a qualquer pessoa sem grandes conhecimentos em programação criar vídeos manipulados, a partir de um conjunto de imagens de uma potencial vítima.

A democratização e difusão do uso desta tecnologia apresenta alto potencial destrutivo, ainda subestimado pelas autoridades e sociedade, podendo criar cenários devastadores, em que qualquer um em posse de um computador e um banco de imagens, como aqueles disponíveis em redes sociais, possa criar vídeos falsos extremamente difíceis de se distinguir, possibilitando que manipulações políticas e sociais sejam cada vez mais difundidas e convincentes, confundindo a linha da realidade e ficção, o que segundo Foer (2018) pode gerar um verdadeiro colapso à realidade.

Porém, um contraponto a tudo isso é que esta mesma difusão das técnicas de aprendizagem de máquina para a troca facial, juntamente com aqueles mesmos valores instituídos pelo conceito de software livre, apresentados acima, possibilitam

que a comunidade entusiastas bem como estudiosos da área de Inteligência Artificial trabalhem no desenvolvimento de tecnologias que detectem os vídeos manipulados, o que já está acontecendo, identificando mudanças que no futuro possam ser imperceptíveis aos olhos humanos. Outro ponto interessante é uso benéfico desta tecnologia, o qual já vem sendo discutido, como seu uso para manter pessoas que precisam de proteção de identidade no anonimato, a criação de conteúdos e comerciais personalizados para o usuário final e até um possível processo de licenciamento facial por atores no futuro.

5. Política e *DeepFakes*

No campo da política, as possíveis consequências de um vídeo *deepfake* podem ser ainda mais danosas que a imagem individual de uma pessoa ou grupo. Uma afirmação ou vídeo falsos poderiam levar até a uma guerra num caso extremo.

Imagine a situação num hipotético 2020: a inteligência americana recebe um vídeo com o líder norte-coreano Kim Jong-un preparando um ataque nuclear aos Estados Unidos. O presidente americano reage e ordena um contra-ataque. É iniciada uma guerra. A situação descrita poderia ser real, afirma o pesquisador Giorgio Patrini, da Universidade de Amsterdã, na Holanda. Patrini ressalta que um vídeo falso pode parecer tão realista que muitos podem acreditar nele, mesmo comprovado sua falsidade. Isso pode ocorrer pela repetição e exposição repetida e quando a informação falsa confirmar nossas crenças anteriores, ou seja, viés de confirmação (MONNERAT, 2018).

Outros exemplos de situações do uso político das *deepfakes* citados pelos professores de direito Chesney e Danielle Citron da Universidade de Maryland são a de um político “aceitando subornos” ou “praticando adultério”; soldados “assassinando civis inocentes em uma zona de guerra”; “autoridades anunciando um ataque iminente contra mísseis contra Los Angeles ou uma pandemia emergente na cidade de Nova York, provocando pânico e coisas piores”. *Deepfakes* políticos, em particular, representam um risco para a segurança nacional de nações. Eles podem forçar as relações já tensas entre governos e intensificar a falta de confiança no discurso oficial e nas instituições públicas (LIFHITS, 2018).

De acordo com os professores Chesney e Citron: “um dos pré-requisitos para o discurso democrático é um universo compartilhado de fatos e verdades apoiados por evidências empíricas”. Fraudes profundas e eficazes poderão permitir que os indivíduos vivam em suas próprias realidades subjetivas, onde as crenças podem ser apoiadas por" fatos "fabricados. O grande risco desta situação é que quando há dúvidas em fatos e dados empíricos básicos, o discurso democrático pode não prosseguir de forma sustentada.

No campo político, perguntas utilizadas no meio jornalístico devem, mais do que nunca, serem aplicadas e verificadas. Deve-se questionar sempre as fontes da notícia/vídeo, verificar se alguma pessoa ou grupo teria interesse na divulgação da

notícia/vídeo e se algum organismo independente já checou aquela informação previamente. São medidas básicas para evitar a proliferação das *fake news* e *deepfakes*.

Referências:

BRASIL. Câmara dos Deputados. Projeto de Lei Ordinária 5.555/2013. Altera a Lei nº 11.340, de 7 de agosto de 2006 – Lei Maria da Penha – criando mecanismos para o combate a condutas ofensivas contra a mulher na Internet ou em outros meios de propagação da informação. Disponível em <http://www.camara.gov.br/sileg/integras/1087309.pdf>. Acesso em 02 de outubro de 2018.

BRASIL. Câmara dos Deputados. Projeto de Lei Ordinária 6.630/2013. Acrescenta artigo ao Código Penal, tipificando a conduta de divulgar fotos ou vídeos com cena de nudez ou ato sexual sem autorização da vítima e dá outras providências. Disponível em <http://www.camara.gov.br/proposicoesWeb/prop_mostrarintegra;jsessionid=21DA571BA8765419CAECA9CB99780280.proposicoesWeb1?codteor=1166720 &filename=PL+6630/2013>. Acesso em 03 de outubro de 2018.

LEE, D. **Deepfakes porn has serious consequences** - BBC News. Disponível em: <https://www.bbc.com/news/technology-42912529>. Acesso em: 13 out. 2018.

Lei No - 12.737, de 30 de novembro de 2012. Brasília: Imprensa Nacional. Diário Oficial da União. CXLIX (232). 1 páginas. 3 de dezembro de 2012. ISSN 1677-7042. Consultado em 13 de setembro de 2018.

LIFHITS, Jenna. Deepfakes Are Coming. And They're Dangerous.**The Weekly Standard**, 20 jul. 2018. Disponível em: <https://www.weeklystandard.com/ jenna-lifhits/deepfake-videos-are-a-national-security-threat>. Acesso em: 14 out. 2018.

MARTÍNEZ, Antonio García. **The Blockchain Solution to Our Deepfake Problems**. 2018. Disponível em: <https://www.wired.com/story/the-blockchain-solution-to-our-deepfake-problems/>. Acesso em: 12 out. 2018.

MONNERAT, Alessandra; SARTORI, Caio. Vídeos falsos feitos com inteligência artificial podem ser danosos em campanhas eleitorais. **Estadão**, São Paulo, 25 jun. 2018. Disponível em: <https://politica.estadao.com.br/blogs/estadao-verifica/videos-falsos-feitos-com-inteligencia-artificial-podem-ser-danosos-em-campanhas-eleitorais/>. Acesso em: 29 set. 2018.

PL 2793/2011». Câmara dos Deputados. Consultado em 13 de setembro de 2018.. Conhecida como Lei Carolina Dieckmann.

ROMANO, Aja. **Why Reddit's face-swapping celebrity porn craze is a harbinger of dystopia**. 2018. Disponível em: <https://www.vox.com/2018/1/31/16932264/reddit-celebrity-porn-face-swapping-dystopia>. Acesso em: 12 out. 2018.

REBECCA, L. **Deepfakes are about to make revenge porn so much worse**. Disponível em: <https://mashable.com/article/deepfakes-revenge-porn-domestic-violence/#8RTBLZ1SxOqZ>. Acesso em: 13 out. 2018.

TAYLOR, Christie; BRANDON, Brandon. **The Reality-Distorting Tools Of The Future**. 2018. Disponível em: <https://www.sciencefriday.com/segments/the-reality-distorting-tools-of-the-future/>. Acesso em: 12 out. 2018.