"*Deepfake*" como ferramenta manipulação e disseminação de "*fakenews*" em formato de vídeo nas redes sociais

Cristiane Pantoja de Moraes

0000-0002-3872-2717 + Universidade Federal do Pará, Belém, Pará. crikapj@gmail.com

Tipo de trabalho: comunicação

Palavras-chave: Manipulação de imagem; *deepfake*; *fake news*

1 Introdução

Este trabalho tem como objetivo fazer uma análise dos novos conteúdos digitais no âmbito da disseminação de falsas informações ou como são comumente conhecidas, as *fake news*, em formato de vídeo as chamadas *deepfake* nas redes sociais, uma vez que elas ganham rapidamente grande repercussão ao serem compartilhadas pelos usuários, viralizando na *Internet*. Com base nisso, através da literatura disponível sobre este tema, procurou-se discutir o que é *"deepfake"* e *"fake news"*, no contexto da produção e disseminação, bem como seu papel como ferramenta de manipulação das notícias no ciberespaço.

> [...] as notícias falsas ficaram em evidência. Sintetizando e simplificando a percepção geral: a epidemia de notícias falsas fez com que os eleitores e a opinião pública tomassem decisões equivocadas, baseadas na emoção e em crenças pessoais, ao invés de em fatos objetivos (Genesini, 2018, p.47)

De acordo com Bunk et al. (2017), o número de imagens digitais tem crescido exponencialmente com o advento de novas câmeras, *smartphones* e *tablets*. As mídias sociais, como *Facebook*, *Instagram* e *Twitter* contribuíram ainda mais para a sua distribuição. Dessa forma, as ferramentas para manipular digitalmente imagens evoluíram gradativamente e os *software* e aplicativos para *smartphones* tornam-se muito trivial para os usuários que manejam facilmente imagens.

As mudanças que ocorrem no mundo tecnológico, principalmente na manipulação de imagem e vídeos, geraram um campo para a criação de ferramentas modernas e livres, muitas vezes de acessos facilmente disponível, impulsionando o mercado e o paradigma quando se tratam de estruturas modernas de adulteração de imagem e vídeos que antes eram de domínio restrito para indivíduos treinados e hoje pode ser amplamente acessível a qualquer pessoa que

tenha um computador, *desktop* e até um *smartphon*e que contenham aplicativos como *FakeApp*[1].

O reconhecimento facial é uma tecnologia biométrica largamente utilizada porque é mais conveniente de usar do que outras abordagens biométricas, essa tecnologia de reconhecimento facial, tem se desenvolvido rapidamente nos últimos anos sendo uma das mais convenientes em comparação com outros métodos. No entanto, sistemas de biometria[2] para reconhecimento facial são ingênuos, ou seja, simples e não suportam qualquer tipo de detecção podendo ser facilmente falsificado usando apenas uma fotografia. Essa detecção de rosto é uma questão-chave no campo dos sistemas de segurança que utilizam câmeras, pois, infelizmente, existem falhas de impressão e borrão geral da imagem. No entanto, com desenvolvimento de dispositivos de exibição e tecnologia de captura de imagens, é possível reproduzir imagens de rostos semelhantes a rostos reais, o que gera número significativo de ataques usando uma fotografia ou vídeo exibido em uma tela (Miyoung & Youngsook, 2017).

A crescente sofisticação da tecnologia de câmera móvel e o alcance veloz das mídias sociais e mídia fizeram a criação e propagação de vídeos manipulados mais convincentes do que nunca. No que diz respeito aos números de vídeos falsos e seus graus de e sofisticação, o tempo de fabricação e manipulação destes tem diminuído significativamente nos últimos anos, graças à acessibilidade e ao grande volume e poder de computação e aplicativos *fakeapp*, e isso inclui também o crescimento do aprendizado em relação às máquinas e a visão mais detalhista de qualquer pessoa, que elimina a necessidade de um profissional (Li & Lyu, 2018; Maras & Alexandrou, 2018).

Hoje, as imagens digitais podem ser facilmente manipuladas e alteradas, as falsificações digitais, muitas vezes não deixam pistas visuais de alteração, podendo ser indistinguíveis aos autênticos (Popescu & Faridi , 2005).

Nas últimas décadas, a popularização dos *smartphones* e o crescimento das redes sociais fizeram as imagens e vídeos digitais objetos tão comuns, de acordo com vários relatórios, quase dois bilhões de fotos são enviadas todos os dias na internet, entretanto surgiu um aumento de técnicas para alterar o conteúdo da imagem, usando software de edição, o que gerou um campo de investigação forense digital dedicado à detecção de falsificações de imagem, a fim de regular a circulação tais conteúdos (Afchar, Nozick & Yamagishi, 2018).

1.1 O que é "*Deepfake*"

"*Deepfake*" é uma técnica que visa substituir o rosto de uma pessoa por outra em um vídeo. Em questão de datas, o primeiro ocorreu em Outono de 2017 utilizado para gerar conteúdos adultos. Posteriormente, essa técnica foi melhorada por uma pequena comunidade para criar nomeadamente uma aplicação chamada "*FakeApp*". O processo para gerar *deepfake* consiste em imagens que reúnem rostos alinhados de duas pessoas diferentes, nas quais há a reconstrução do rosto de uma em conjunto de dados de imagens faciais da outras e se auto-codifica para então reconstruir rostos com as imagens faciais. Na prática, os resultados são impressionantes, o que explica a popularidade da técnica. O último passo é levar o vídeo ao alvo, extrair e alinhar a face do alvo a partir de cada quadro, utilizando software ou aplicativos *"FaceApp*[3]*"* para gerar outra face com a mesma iluminação e expressão, e então fundir de volta no vídeo Afchar et al. (2018).

Atualmente, de acordo com Korshunov e Marcel (2018); Güera e Delp (2018), o "*FaceApp*" uma forma de "*FakeApp*" trata-se de uma ferramenta para *smartfones* que pode gerar automaticamente rostos em fotografias o que torna a imagem altamente realista, permitindo a mudança de rosto, cabelo, sexo, idade e muitos outros detalhes usando apenas o telefone móvel.

Os chamados "*Deepfakes*" são a mais nova forma de manipulação de mídia digital, atualmente existem diversas ferramentas de *softwares* livres aperfeiçoados em aprendizagem de máquinas que criam facilmente rostos em vídeos deixando poucos resquícios de manipulação.

Exemplo de uma imagem (esquerda) ao ser manipulado (à direita), utilizando a técnica *deepfake*. Note-se que a face trocada não apresenta a expressividade do original.

Figura 1: Exemplo de manipulação com a técnica *deepfake*

Fonte: (Afchar, Nozick & Yamagishi, 2018).

1.2 Como o "*Deepfake*" pode manipular e disseminar as "*Fake new*s" nas redes sociais

No panorama atual, é comum visualizar vídeos falsos e, por sinal, são totalmente realistas, o que os tornam perigosos pessoalmente como politicamente, às vezes são utilizados para falsas promoções, chantagens e até desmoralização pessoal.

> O novo é que estamos em uma nova era turbinada pela internet e pelas redes sociais, em que o crescimento é viral e o efeito, exponencialmente explosivo. O novo é o Facebook, o Google e o Twitter, não a tentativa de contar mentiras ou falsificar informações, o que sempre existiu na história do mundo (Genesini, 2018, p.46).

É comum nos divertimos usando o recurso *faceswap*[4] em aplicativos de celulares, uma opção que literalmente faz a trocar rostos, mesmo que consista em mudar de rostos com nossos amigos ou parecer como uma celebridade, essa tecnologia é frequentemente usada de maneira desagradável. No entanto, o reconhecimento facial tornou-se um pesadelo para alguns famosos com o surgimento dos *deepfake*, permitindo que os usuários criem vídeos com pessoas usando a aparência destes.

Frequentemente essa tecnologia tem sido usada para criar vídeos pornográficos falsos de celebridades causando imagens assustadoras, uma vez que é muito além de uma simples "troca de rostos", como as utilizadas nos aplicativos de celulares, isso acontece com *deepfake*,

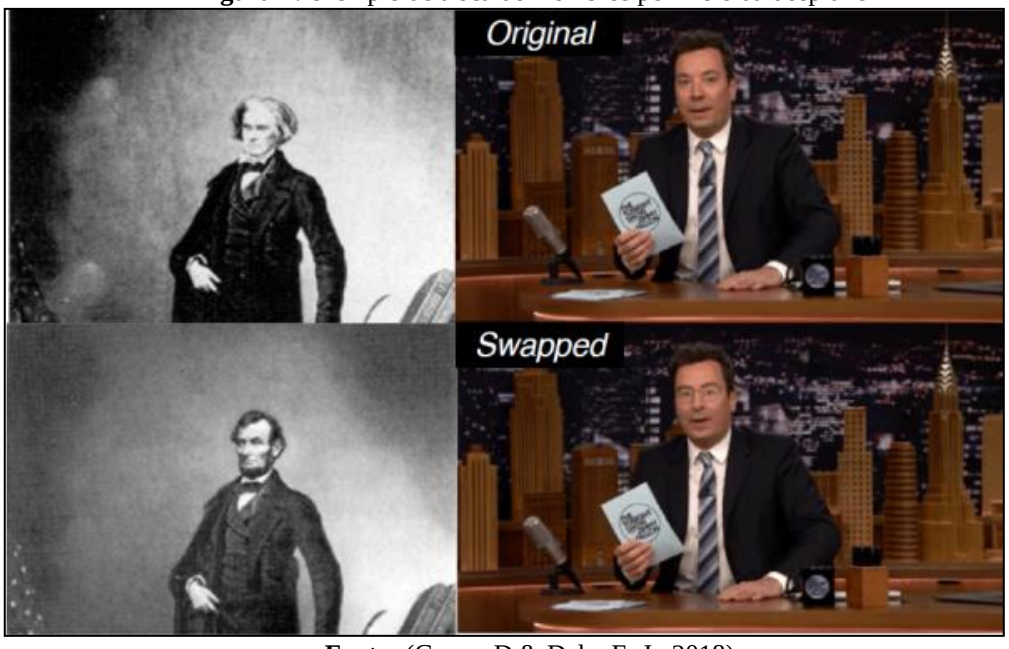
pois é um conjunto de técnicas utilizadas para sintetizar novos produtos visuais, por exemplo, substituindo rostos nos originais (Floridi , Korshunov & Marcel , 2018).

De acordo com Korshunov e Marcel (2018) são de grande preocupação pública a respeito dos *deepfakes*, existem atualmente software de código aberto acessível e aplicativos utilizados para a troca de rosto, o que gera uma grande quantidade de vídeos sinteticamente manipulados e distribuídos nas mídias sociais e notícias, o que representa um significante desafio técnico para detecção e filtragem de tal conteúdo.

Já não é de hoje as várias tentativas para a troca de rosto. Conforme Güera e Delp, (2018), explicam que, por volta de 1865, foram encontrados na litografia os traços do presidente dos Estados Unidos Abraham Lincoln de acordo com a figura a seguir, assim como utilização de *fakeapp* para a produção de *deefakes*.

Abaixo o exemplo de troca da cabeça do Presidente Lincoln com o corpo do político John Calhoun, criadas no século XIX (esquerda) e ferramentas modernas como, por exemplo, *fakeapp* viraram mais fácil para qualquer se produzir *deepfakes*, tais como a troca das cabeças do apresentador TV Jimmy Fallon[5] e John Oliver[6] (direita).

Figura 2: exemplo de troca de membros por meio da deepfake



Fonte: (Guera, D.& Delp, E. J., 2018).

O que distingue *deepfakes* de outras técnicas de manipulação de vídeo é, primeiramente, o seu potencial de resultados fotorrealista, com imagens em que os vídeos

resultantes podem ser extremamente convincentes. Em segundo lugar, a disponibilidade da técnica para leigos em um aplicativo chamado *fakeapp, que se trata* uma interface lançada e desenvolvida em torno do algoritmo *deepfake*, que permite que os usuários com conhecimento limitado de programação e pouca aprendizagem em tecnologia possam criar *deepfakes*.

Segundo (Koopman, Macarulla Rodriguez & Geradts, 2018) essa combinação de resultados fotorrealista é facilmente utilizada gera um aumento na necessidade de criar métodos autenticação para detecção de manipulação em *deepfake,* o que tem gerado um problema na atualidade das "notícias falsas", as chamadas *"fake news"*, que vão se tornando mais relevante também para os jornalistas em vídeo, hospedagem de sites e usuários de mídia social.

Hoje, o perigo de notícias falsas é amplamente reconhecido num contexto em que milhões de horas de conteúdo de vídeo são vistos diariamente em redes sociais, a propagação de vídeos falsificados aumentou o nível de preocupação. Por mais que melhorias sejam feitas para detecção de imagem falsificada, vídeo digital manipulados, o desvendamento continua a seu uma tarefa difícil.

Com os aparecimentos dos *deepfake* como cita Korshunov e Marcel (2018); Güera e Delp (2018), ferramentas que geram os *deepfakes* de vídeos têm sido amplamente utilizadas para criar falsas notícias de celebridades associados às pornografias ou até mesmo vídeos de vingança. Alguns sites como o *Twitter* já proibiram esse tipo de vídeos, embora a natureza desses vídeos quase que realista, torna-se alvo para gerar matérias de conteúdos ilícitos, falsos, pornográficos e maliciosos, utilizados até para criar tensão política e estão sendo alvos da atenção das entidades governamentais.

1.3 Como detectar os "*Deepfake*"

De acordo com Tan, Li, Liu e Jiang (2010) é comum que os dados faciais sejam roubados ou duplicados em um sistema de reconhecimento de face. Isto porque na internet, uma ou mais fotografias de um usuário pode ser facilmente obtida, sem contato físico o usuário através de download via internet ou simplesmente capturá-los usando uma câmera. Este sistema de reconhecimento facial baseado em imagens 2D pode ser facilmente manipulado por truques considerados bem simples, e que na verdade, é uma tarefa muito desafiadora para em que os esforço maior vem na pesquisa de reconhecimento facial que no

momento atual estaá focada mais na imagem correspondente do sistema sem se importar se o rosto combinado é de um ser humano.

Recentemente, o reconhecimento de rosto tornou-se cada vez mais importante devido aos rápidos avanços em dispositivos de captura de imagem (por exemplo, câmeras de vigilância e câmera em celulares), onde a disponibilidade de um número muito grande de imagens de rosto na internet, e as crescentes demandas por maior segurança (Li & Jain, 2011 como citado em Miyoung & Youngsook, 2017)

Os autores Korshunov e Marcel (2018) comentam que se considerarmos os vários sistemas de detecção *deepfake*, incluindo o sistema que utiliza dados de áudio-visual para detecção de inconsistências entre movimentos labiais e discurso de áudio,assim como, diversas variações nos sistemas baseados em imagem têm como objectivo distinguir os vídeos genuínos, onde o movimento da fala é sincronizado, no vídeo que foi modificado, e esses movimentos dos lábios e áudio, pode não ser necessariamente o discurso original.

A tecnologia *faceswap* é um programa *deepfake* em que se usa pesquisa de imagens na internet, explorando sites de mídia social, e depois, por conta própria, insere dados para substituir os rostos em vídeos quase que perfeitamente. O programa não precisa de nenhuma supervisão humana, seu processo de aprendizagem é independente e continua a melhorar o processo de forma autônoma. Qualquer pessoa pode criar vídeos pornográficos estrelados por celebridades, políticos, amigos ou inimigos (Cuthbertson, 2018 como citado em Maras & Alexandrou, 2018).

Existem diversos métodos para detectar a manipulações de imagem digitais baseados em artefatos de reamostragem de cor, variedade filtro, luz, câmeras forense, compressão JPEG, e muitos mais.

Alegações de “notícias falsas” já são tão comuns, tanto que a vítima de um vídeo falsificado pode não ser capaz de estabelecer a credibilidade ao afirmar que tenham sido falsificados, pois existe uma abundância de ferramentas sofisticadas para a detecção de falsificação de imagem. São susceptíveis de ser desenvolvidos, já que a demanda por eles tem aumentando. Floridi (2018).

Conforme Koopman et al. (2018) ,o algoritmo responsável por gerar o *deepfake* permite que um usuário possa mudar o rosto de um ator em um vídeo com o de um ator

diferente de uma forma Fotorrealista, isso desencadeia desafios aos profissionais forenses com relação à confiabilidade das provas de vídeo. Os testes realizados quanto à sua eficácia na detecção de manipulação de vídeo *deepfake* ainda mostram diferença significativa entre vídeos autênticos e "*Deepfakes*". Ultimamente, as provas fotográficas e de vídeo são comumente usados nas investigações tribunal e da polícia, que eram vistas como confiáveis de provas, no entanto, essas amostras de vídeo estão tornando-se potencialmente pouco viáveis e provavelmente serão necessários resquícios nesses vídeos, pois ao serem examinados cautelosamente, sempre deixam vestígios de adulteração antes de serem considerados admissíveis para qualquer tribunal.

Durante os últimos anos, métodos de aprendizagem profunda foi um sucesso empregado nas perícias de imagens digitais. Usar a aprendizagem profunda para detectar imagens, propõe uma rede para detectar o alvo da falsificação distinguindo gráficos de computador, executa muito bem pelos profissionais forenses digitais.

2 Considerações finais

Os "*Deepfakes*", quando comparadas a informações não rotuladas como "notícias falsas" ou *Fake news*, causam impactos quando são descobertas e um dos desafios da atualidade é rastrear "notícias falsas" na batalha contínua contra a desinformação.

Em seguida, se analisamos a relação entre a manipulação e disseminação de notícias falsas, as duas são reconhecidas como desinformação, haja vista que os usuários estão sujeitos a todo e qualquer tipo de imagem e vídeos compartilhados todos os dias nas redes sociais, fatos que mostram aumento nos resultados da polarização de "*fake news*" em que usuários associam a palavras-chave e *hashtags*[7] em notícias falsas, sem ao menos buscar a verdadeira fonte ou fonte confiável ao compartilhar a informação. O impacto dessa nova descoberta dos *deepfake* veio como desafio para rastrear "notícias falsas" no intuito de diminuir a propagação destas de acordo com peritos digitais.

As notícias apontaram a necessidade de manter um firme controle incidente sobre as notícias falsas, pois elas podem colocar em risco a liberdade de expressão, a honra das pessoas e o próprio processo democrático de um país. Foi possível notar, ainda, que para se combate as *fake news*, é obrigatório o respeito à liberdade de expressão, e contar com a cooperação e a ética dos usuários em compartilhar a notícia publicada.

Hodiernamente, o perigo da adulteração facial, ocasionado pela mudança no vídeo ainda não são amplamente reconhecidos pela população. Apesar de existirem produtos para detectar tais falsificações de forma eficiente e com um baixo custo computacional, muitas pessoas ainda desenformadas disseminam todo e qualquer tipo de vídeo sem se importar com sua veracidade ou fonte o que ocasiona uma disseminação desencadeada de falsas notícias que são rapidamente espalhadas em questão de segundos na internet

3 Referências

Afchar, D. , Nozick ,V., Yamagishi,J. (2018). MesoNet: A compact facial video forgery detection network. *arXiv*:1809.00888v1 [cs.CV] 4 Sep 2018. Recuperado de https://arxiv.org/abs/1809.00888.

Bunk, J. , Bappy, J. , Mohammed, T. , Nataraj, L. , Flenner, A. , Manjunath, B. , Chandrasekaran,S., … Peterson, L. (2017). Detection and localization of image forgeries using resampling features and deep learning. *arXiv*:1707.00433v1 [cs.CV] 3 Jul 2017. Recuperado de https://arxiv.org/abs/1707.00433

Floridi, F.(2018). Artificial Intelligence, Deepfakes and a Future of Ectype. *Springer Netherlands.* 31:317–321. Recuperado de doi:10.1007/s13347-018-0325-3

Genesini, S. (2018). A pós-verdade é uma notícia falsa. *Revista USP*. São Paulo:116, 45-58. Recuperado de http://www.revistas.usp.br/revusp/article/view/146577

Guera, D., Delp, E. J.(2018). Deepfake video detection using recurrent neural networks. In *IEEE International Conference on Advanced Video and Signal-based Surveillance (to appear)*. Recuperado de https://engineering.purdue.edu/~dgueraco/content/deepfake.pdf

Koopman, M., Rodriguez, A. M., Geradts, Z.(2018). Detection of Deepfake Video Manipulation.In: *Conference: IMVIP*, At Belfast. August 2018. Recuperado de https://www.researchgate.net/publication/329814168_Detection_of_Deepfake_Video_Manipulation

Korshunov,P., Marcel,S.(2018). DeepFakes: A new threat to face recognition? assessment and detection. *arXiv*:1812.08685v1 [cs.CV] 20 Dec 2018. Recuperado de http://publications.idiap.ch/downloads/reports/2018/Korshunov_Idiap-RR-18-2018.pdf

Li, Y., Lyu,S.(2018). Exposing deepfake videos by detecting face warping artifacts. *arXiv*:1811.00656v1 [cs.CV]. Recuperado de https://arxiv.org/abs/1811.00656

Maras,M-E., Alexandrou,A.(2018). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake vídeos. *The International Journal of Evidence & Proof*. 1–8. Recuperado de doi: 10.1177/1365712718807226

Miyoung, C., & Youngsook, J. (2017) Face recognition performance comparison between fake faces and live faces. *Soft Computing* 21(12), 3429-3437. Recuperado de https://link.springer.com/content/pdf/10.1007%2Fs00500-015-2019-4.pdf

Popescu, A. C, Farid,H.(2005). Exposing digital forgeries by detecting traces of resampling. in *IEEE Transactions on Signal Processing*. 53(2),758-767. Recuperado de doi**:** 10.1109/TSP.2004.839932

Tan , X., Li, Y., Liu, J., & Jiang, L. (2010) Face liveness detection from a single image with sparse low rank bilinear discriminative model, Computer Vision ECCV 2010. Springer, Berlin Heidelberg

[1] Aplicativo em como principal função é a possibilidade de modificar as suas fotos, mais precisamente os *selfies*, transformando seu rosto em uma versão completamente inusitada.

[2] Tecnologia que estabelecer a identidade de um indivíduo com base em um ou mais características fisiológicas ou comportamentais, tais como rostos, impressões digitais, íris e até vozes.

[3] Aplicativo desenvolvido para *smartphones* utiliza inteligência artificial para transformar a foto do usuário, com filtros que permitem simular alterações em sua aparência física.

[4] Ferramenta que utiliza aprendizado profundo para reconhecer e trocar faces em fotos e vídeos com base no código original do *deepfakes*.

[5] Ator, comediante, músico e apresentador de televisão famoso nos Estados Unidos .

[6] Ator e comediante britânico, também é o apresentador do programa de sátira política.

[7] É uma expressão bastante comum entre os usuários das redes sociais, na internet. Consiste de uma palavra-chave antecedida pelo símbolo #.