Research, Society and Development, v. 11, n. 6, e56211629533, 2022
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i6.29533

Deepfake: A evolução das fake news

Deepfake: The Evolution of fake news

Deepfake: La evolución de las fake news

Recebido: 23/04/2022 | Revisado: 01/05/2022 | Aceito: 06/05/2022 | Publicado: 10/05/2022

Adriano Cezar Molina
ORCID: https://orcid.org/0000-0001-8286-7449
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Brasil
E-mail: adrianocmolina@gmail.com
Orlando Leonardo Berenguel
ORCID: https://orcid.org/0000-0002-1498-3840
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Brasil
E-mail: oberenguel@hotmail.com

Resumo
O surgimento de novas técnicas como *deepfake* possibilitaram a manipulação e criação de novos conteúdos falsos de vídeos, áudios e imagens muito semelhantes ao original. A técnica pode ser usada para substituir o rosto de uma pessoa por outra com intuito de fazer uma pessoa dizer ou fazer coisas que nunca aconteceram. Os conteúdos falsos gerados são tão realistas quanto os originais, o que torna a tecnologia uma arma poderosa para construções de *fake news*. Por se tratar de uma técnica sofisticada, apresenta capacidade de distorcer a verdade e gerar conflitos, pondo em risco a reputação dos envolvidos e criando profundo prejuízo aos atingidos quando usado para fins escusos. É de suma importância popularizar como as *deepfakes* funcionam a fim de se ampliar o debate sobre o seu uso indevido. Nesse sentido, este artigo tem como objetivo aprofundar os estudos sobre a técnica *deepfake* e alertar sobre sua existência e seu uso indevido e quais são as soluções existentes para combatê-la. Foi realizada pesquisa bibliográfica baseada em diferentes fontes de dados, como sites especializados, leis, revistas científicas, e artigos. Conclui-se que *deepfakes* são variações e evolução das *fake news*, o uso indevido da tecnologia para a geração de conteúdos com o caráter de manipulação da opinião pública pode trazer danos graves para a sociedade. Os resultados demonstram que, embora a tecnologia seja neutra enquanto recurso, seu uso para fins de degradação de pessoas ou organizações não pode ser controlado, demandando a criação de leis que coíbam seu uso indevido.
Palavras-chave: Deepfake; Inteligência artificial; Fake news.

Abstract
The emergence of new techniques, such as deepfake, has made it possible to manipulate and create new fake video, audio, and image content very similar to the original. This technique can be used to replace the face of a person with another, so as to make somebody say or do things that have never happened. The fake contents generated by this technology are as realistic as the originals, which makes it a powerful weapon in the construction of fake news. Because it is a sophisticated technique, it presents the ability of distorting the truth and causing conflict, risking the reputation of those involved, and creating great harm to them when used for criminal purposes. It is crucial to popularize and spread how deepfakes work in order to broaden the discussion about their misuse. Therefore, this article aims at deepening the studies on deepfake techniques, warning about their existence, their misuse, and the existing solutions to fight them. A bibliographical research was carried out in different sources of information, such as specialized websites, laws, scientific journals, and articles. We have concluded that deepfakes are a variation and the evolution of fake news, with its misuse aiming at generating content to manipulate public opinion could bring serious harm to society. The results have showed that, although technology is neutral as a resource, its use with the purpose of harming people or organizations cannot be controlled, which calls for the creation of laws that curb its misuse.
Keywords: Deepfake; Artificial intelligence; Fake news.

Resumen
El surgimiento de nuevas técnicas como el deepfake hizo posible manipular y crear nuevos contenidos falsos de videos, audios e imágenes muy similares al original. La técnica se puede usar para reemplazar la cara de una persona con otra para hacer que una persona diga o haga cosas que nunca sucedieron. El contenido falso generado es tan realista como el original, lo que convierte a la tecnología en un arma poderosa para la construcción de fake news. Como es una técnica sofisticada, tiene la capacidad de distorsionar la verdad y generar conflictos, poniendo en peligro la reputación de los involucrados y creando un profundo daño a los afectados cuando se utiliza con fines nefastos. Es de suma importancia popularizar cómo funcionan los deepfakes para ampliar el debate sobre su uso indebido. En ese sentido, este artículo pretende profundizar en los estudios sobre la técnica del deepfake, advertir sobre su existencia, mal uso y cuáles son las soluciones existentes para combatirlo. La investigación bibliográfica se realizó a partir de

diferentes fuentes de datos, como sitios web especializados, leyes, revistas científicas y artículos. Se concluye que deepfakes son variaciones y evolución de las fake news, el mal uso de la tecnología para generar contenidos con carácter de manipulación de la opinión pública puede traer graves perjuicios a la sociedad. Los resultados muestran que, si bien la tecnología es neutral como recurso, no se puede controlar su uso con el fin de degradar a personas u organizaciones, exigiendo la creación de leyes que prohíban su mal uso.
Palabras clave: Deepfake; Inteligencia artificial; Fake news.

1. Introdução

Os avanços da inteligência artificial e o surgimento de novas técnicas como *deepfake* possibilitam a manipulação e criação de novos conteúdos falsos de vídeos, áudios e imagens muito semelhantes ao conteúdo original, apesar de trazer diversos benefícios, seu uso indevido é preocupante, as *deepfake* se tornaram uma ferramenta poderosa para influenciar ou distorcer a verdade, seja no âmbito político ou social, usada dessa forma a tecnologia passa a ser uma extensão das *fake news*.

De acordo com o dicionário Cambridge (s.d), *fake news* indica histórias falsas que parecem ser notícias, espalhadas na internet ou usando outros meios de comunicação, geralmente criadas para influenciar pessoas. As *fake news* são propagadas nas redes sociais e aplicativos de compartilhamento de mensagens com velocidade e escala, o que gera descontrole na disseminação.

Estrategicamente criadas com conteúdo que atraia a atenção dos usuários das redes, capazes de ironizar, divertir ou até mesmo provocar a opinião pública, se espalham como algo simples, pelo desejo gerado de compartilhamento fortemente presente em redes sociais. Para se ter dimensão do efeito das *fake news*, em 2019 o centro para inovação em Governança Internacional realizou pesquisa com usuários de internet em 25 países. Dos entrevistados 86% admitiram acreditar em pelo menos uma *fake news*. Outro dado importante é que a mesma pesquisa aponta que a maioria dessas *fake news* estavam em redes sociais como *Facebook* e *Twitter* (Agence France-Presse, 2019).

Outra pesquisa realizada com 70.333 brasileiros em 2020 pelo laboratório especializado em segurança digital da *Psafe* constatou que 75% dos entrevistados já foram impactados por *fake news*, 55% repassaram *fake news* sem saber e 80% dos entrevistados receberam essas *fake news* através do *Facebook* e *WhatsApp* (Pecsen, 2020).

Convencer pessoas, manipular opiniões, gerar degradação de imagem de pessoas e organizações, influenciar e obter ganhos forçaram uma verdadeira mutação tecnológica dos conteúdos de *fake news*. As *deepfakes* são evoluções tecnológicas na produção de *fake news*. A edição de vídeos, áudios e fotografias com o uso de inteligência artificial permite a criação e disseminação rápida de conteúdos modificados e com alta qualidade, dificultando a identificação de fraudes e adulterações. (Korshunov & Marcel, 2019). Com o passar dos anos, esses conteúdos ficarão tão precisos que será difícil distinguir se tal conteúdo é verdadeiro ou não.

Outro fator que contribui na disseminação desses conteúdos modificados é que muitos ainda desconhecem a existência dessa técnica, e o que aponta a pesquisa realizada pela *Kaspersky*, em parceria com a empresa de pesquisa CORPA. De acordo com a pesquisa 66% dos brasileiros desconhecem a técnica e 71% não reconhece quando um vídeo foi editado usando a técnica, tal resultado pode colaborar no sucesso de fraudes e ajudar em ataques de engenharia social. (Sica, 2022). Segundo Bestuzhev, (citado por Sica 2022) "À medida que a tecnologia se torna menos cara, podemos esperar o surgimento de seu uso ilícito".

As *deepfakes* representam um grande problema para a sociedade no futuro. A questão é que, quando as ferramentas de fabricação de *deepfakes* se tornam populares e acessíveis, os efeitos serão mais graves que o fenômeno atual das *fake news* (Patrini et al., 2018).

Nesse sentido, este artigo tem como objetivo aprofundar os estudos sobre a técnica *deepfake* discutindo o uso indevido de tal tecnologia e quais são as soluções existentes para combatê-la.

2. Metodologia

A fim de atender ao objetivo proposto, foi realizada pesquisa bibliográfica, que de acordo com Martins (2017, p. 22) envolve a leitura, análise e interpretação de diferentes fontes disponíveis sobre determinado tema, qualquer contribuição científica, seja ela impressa ou eletrônica pode se tornar uma fonte de consulta.

Foi feito levantamento de informações sobre o tema em diferentes fontes de dados, como sites especializados, leis, revistas científicas e artigos utilizando a ferramenta Google Acadêmico. Para critérios de inclusão, foram utilizados artigos disponíveis no idioma português e inglês relacionados à temática *deepfake*, publicados entre o período de 2019 a 2021, utilizando as palavras-chaves: *deepfakes* desinformação; detecção de *deepfakes*; *deepfakes* redes adversárias generativas; *deepfakes* ameaças; *fake news* e *deepfakes* em português e *deepfakes*; *deepfakes disinformation*; *deepfakes detection*; *deepfakes generative adversarial networks*; *deepfakes threat*; *fake news and deepfakes* em inglês.

Os artigos foram selecionados a partir do título e leitura do resumo, quando compatíveis realizaram-se as leituras e análises completas para o levantamento da pesquisa.

3. Das Fake News as Deepfakes

O termo *fake news* se refere aos relatos de diversos assuntos, na maioria com propósitos políticos alterados ou inventados com objetivos de enganar, manipular e caluniar uma pessoa ou uma instituição (Galhardi et al., 2020). Embora o dicionário *Merriam-Webster* (s.d) aponte que a expressão *fake news* surgiu no fim do século XIX, a utilização do termo se tornou popular durante a cobertura jornalística da eleição presidencial americana de 2016. No Brasil o termo se popularizou em 2018 devido ao cenário político (Viera, 2019).

Mesmo antes de serem evidenciadas nos meios de comunicação, as *fake news* já haviam causado danos sociais, exemplo disso é a disseminação de uma *fake news* em 2014 que acarretou a morte de uma mulher de 33 anos que foi confundida como uma sequestradora de crianças, na ocasião uma página do *Facebook*, além de circular um retrato falado, alertava sobre uma mulher que sequestrava crianças para fazer rituais de magia negra. Tudo não passava de uma *fake news* e, segundo a polícia, nenhum sequestro de criança havia sido relatado na cidade. Como resultado dessa *fake news*, a mulher foi linchada até a morte por moradores da cidade, deixou marido e dois filhos e cinco pessoas foram acusadas e condenadas pelo crime, mas o principal responsável pela criação da *fake news* ficou impune, uma vez que a legislação da época não previa punição a incitação à violência por meio da internet (Campos, 2019).

Divulgar informações falsas e arquivos digitais manipulados tem sido perigoso, gerando graves danos às pessoas envolvidas, como o exemplo citado anteriormente, as redes sociais formam um ambiente propício para a difusão em massa de notícias, sejam elas verdadeiras ou não (Campos, 2019).

Apesar das *fake news* já causarem um dano assustador, o uso da inteligência artificial e técnicas de aprendizado profundo que deram origem as *deepfakes*, permite a criação rápida e de alta qualidade de conteúdos digitais falsos, tal técnica tem o potencial de alterar a verdade e desgastar a confiança, dando "autenticidade" às *fake news* (Hasan & Salah, 2019).

O termo *deepfake* surgiu em 2017 quando um usuário do *Reddit* usou o apelido "*deepfakes*" para postar vídeos pornográficos alterados digitalmente com imagens de celebridades. A tecnologia foi aplicada usando como base inúmeras imagens e vídeos de celebridades para aprender a imitar as expressões faciais e sobrepor em um vídeo o rosto de uma celebridade no rosto de atrizes de filmes pornô (Hall, 2018).

A grande preocupação é que existem *softwares* de código aberto e aplicativos que tornam essa tecnologia acessível. Para se obter um bom resultado, ou seja, uma *deepfake* capaz de confundir os usuários de uma rede social, é preciso ter um grande acervo de material. A duração do processo de criação de conteúdos dependerá das configurações de *hardware* utilizados e das redes adversárias generativas (*GANs*). Essas redes podem ter limitações tanto na implementação quanto na

aplicação, por exemplo, para geração de imagens full HD é necessário mais tempo e um *hardware* potente, além de exigir um grande conjunto de dados (Huang, 2021). Com o processo realizado, a geração dos dados sintéticos se torna mais rápida e eficiente.

A maior parte das *deepfakes* são de natureza pornográfica, a tecnologia por trás das *deepfakes* é capaz de aprender e gerar novos dados a partir dos conjuntos de dados existentes como fotos, áudio e vídeo, esses dados são processados por meio de redes adversárias generativas (*GANs*) (Westerlund, 2019).

Redes adversárias generativas (*GANs*) são modelos generativos baseados em aprendizado profundo, uma *GAN* é composta por duas redes neurais: um modelo gerador para gerar novas instâncias de dados e o modelo discriminador que classifica se esses dados pertencem ou não a base de dados de treinamento real (Data Science Academy, 2021).

Segundo Goodfellow et al., (2020), a rede geradora pode ser considerada análoga à falsificadores que tentam produzir notas falsas de dinheiro e usar sem serem detectados, e a rede discriminadora é análoga à polícia que tenta detectar se as notas são falsas. Tanto "falsificadores" como a "polícia" se desafiam para melhorar seus métodos até que chegue um momento em que as falsificações sejam indistinguíveis das notas reais.

> A ideia de colocar dois algoritmos um contra o outro se originou com Arthur Samuel, um proeminente pesquisador no campo da ciência da computação que popularizou o termo "aprendizado de máquina". Enquanto estava na IBM, ele desenvolveu um jogo de damas - o Samuel Checkers-playing Program - que foi um dos primeiros a aprender com sucesso, em parte por estimar a chance de vitória de cada lado em uma determinada posição. Mas se Samuel é o avô dos GANs, Ian Goodfellow, ex-cientista pesquisador do Google Brain e diretor de aprendizado de máquina do Grupo de Projetos Especiais da Apple, pode ser o pai deles. Em um artigo de pesquisa intitulado "Redes adversárias generativas" de 2014, Goodfellow e seus colegas descrevem a primeira implementação prática de um modelo generativo baseado em redes adversárias (Wiggers, 2019).

Apesar da possibilidade de criar diversos tipos de arquivos, o uso mais comum da tecnologia e na produção de vídeos, muitas das *deepfakes* têm como alvo celebridades, políticos e *CEOs* de empresas, devido a facilidade de se encontrar facilmente dados como fotos e vídeos na internet, o que facilita o treinamento das *GANs* (Westerlund, 2019), mas nada impede que a tecnologia possa ser usada também para causar danos a qualquer outra pessoa, uma vez que as redes sociais colaboram com um grande acervo de fotos e vídeos.

O potencial das *GANs* é enorme uma vez que é possível ensinar a criar imagens, vídeos, áudios muito próximos da realidade, quanto mais realista esses conteúdos se tornarem mais difícil será diferenciar o que é real e o que é *deepfake* (Chesney & Citron, 2019). O que acaba contribuindo para que a tecnologia seja usada para fins indevidos, como vingança, desinformação, crimes, fabricação de provas falsificadas, chantagens e *bullying*, por exemplo, não é difícil achar relatos de exemplos nos quais a tecnologia já foi usada para esses fins, em um dos casos, a tecnologia foi usada para falsificar a voz de um dos *CEOs* de uma empresa britânica de energia para solicitar uma transferência bancária de € 220.000, que foi prontamente realizada, pois a voz era tão convincente que enganou o outro *CEO* da empresa (Damiani, 2019).

De acordo com relatório da *Sensity* (2019), 96% dos vídeos *deepfake* disponíveis na internet são de conteúdo pornográfico não consensual, e têm como alvo as mulheres, na maioria atrizes, assim como pessoas comuns, como é o caso da jornalista Rana Ayyub, que foi vítima de uma campanha de desinformação após uma entrevista contra o partido governante BJP da Índia e pedido de justiça para uma vítima de estupro. Segundo a jornalista, no dia seguinte começaram a circular diversos *tweets* falsos como se fossem seus e logo em seguida após esclarecer que esses *tweets* eram falsos, a campanha de desinformação continuou, dessa vez culminando em um vídeo pornô falso que estava circulando nas redes sociais (Ayyub, 2018).

São inúmeras as possibilidades do uso das *deepfakes*, a tecnologia pode ser aplicada para fazer pessoas a dizerem algo que nunca disseram como foi o caso do vídeo no qual Mark Zuckerberg, *CEO* do *Facebook* aparece dizendo que tem o

controle de bilhões de dados roubados, de seus segredos, de suas vidas e seu futuro, e que tudo isso graças a *Spectre* (uma exposição de artes) que mostrou que quem controla os dados controla o futuro (Alecrim, 2019). Assim como criar um perfil falso no *LinkedIn* para fins de espionagem, já que é possível gerar rostos de pessoas que não existem na vida real (Robitzski, 2019).

Nos Estados Unidos, um caso com grande repercussão foi de uma mãe que usou *deepfake* para favorecer sua filha que vinha tendo problemas com suas companheiras de equipe, a tecnologia foi usada na tentativa de prejudicar e causar a expulsão de algumas garotas que faziam parte de uma equipe de líderes de torcida, na intenção de constranger e forçar a saída das garotas da equipe. Foram criados e espalhados conteúdos falsos onde as garotas menores de idade apareciam nuas, com bebidas alcoólicas e fumando (Gogoni, 2021).

Também é possível usar a tecnologia para fins políticos, o que gera motivo de preocupação, até o momento as *deepfakes* foram usadas para fazer sátiras políticas. De acordo Atheniense (2019) as *deepfakes* podem ser uma grande ameaça à democracia, e com o ritmo em que a tecnologia avança tornará o combate ainda mais difícil, pelo fato de distinguir o que é real e o que e falso. “Se a gente já tem esse hábito em relação à disseminação de notícias falsas, quando você tem uma inovação tecnológica como *deepfake*, essa automação só potencializa aquele ato anterior que já existia” (Atheniense, 2019).

4. Métodos de Combate às Deepfakes

4.1 Tecnologias

O uso indevido de *deepfakes* é muito perigoso e preocupante, para impedir a propagação de conteúdos modificados e evitar que a tecnologia se torne uma arma de desinformação, empresas como *Amazon*, *Facebook* e *Microsoft* se reuniram em 2019 com acadêmicos e comitês de inteligência artificial, no intuito de estimular pesquisadores a criar novas tecnologias que ajudem a detectar *deepfakes* (Schroepfer, 2019).

Da mesma forma, outras empresas buscam por novas tecnologias anti *deepfake*, a empresa *Sensity*, por exemplo, desenvolveu uma plataforma que possibilita identificar vídeos e rostos gerados por redes adversárias generativas (*GANs*), basta apenas enviar um arquivo seja eles nos formatos mp4, mov, png, jpeg, jfif e tiff ou inserir um url de um vídeo para que a plataforma possa analisar e verificar se tal arquivo é uma *deepfake* (Sensity, 2021).

A *Google*, seguindo a mesma direção, criou e disponibilizou uma base de dados com mais de 3000 vídeos de *deepfake*, com a finalidade de combater conteúdos modificados pela tecnologia. A empresa espera que pesquisadores utilizem a base de dados para desenvolver e treinar ferramentas que ajudem a identificar *deepfakes* (Cancelier, 2019).

Muitas outras técnicas e soluções estão sendo abordadas para o combate do uso indevido de *deepfake*. Cientistas da Universidade de Buffalo nos Estados Unidos desenvolveram um algoritmo para analisar reflexos dos olhos nos vídeos, sendo que a luz emitida pelo cenário gera reflexos nas córneas, algo que não é tão preciso em vídeos alterados por *deepfake* (Almenara, 2021). O uso de *blockchain* também vem sendo estudado como solução de rastreamento de vídeos originais, o que acaba ajudando a determinar se o conteúdo veio de uma fonte confiável (Hasan & Salah, 2019).

A busca por novas tecnologias pode trazer diversas oportunidades para empresas de cibersegurança produzirem soluções para detecção de *deepfakes*, por outro lado, tais tecnologias não serão suficientes caso organizações e governos não adotem maneiras de se proteger (Westerlund, 2019).

4.2 Legislação

Assim como o surgimento de novas tecnologias, a legislação passa por adequações com o avanço da tecnologia. Em alguns países já existem leis referentes ao uso de *deepfakes*, estados como Califórnia e Virgínia, nos Estados Unidos atualizaram suas leis de combate a pornografia de vingança, proibindo a distribuição de imagens e vídeos pornográficos

modificados por *deepfake* (Parreira, 2021). O estado da Califórnia foi mais além proibindo juntamente *deepfakes* prejudiciais a candidatos políticos no período de 60 dias antes de uma eleição (Theobald, 2019).

O governo chinês, preocupado com riscos políticos e a segurança nacional, criminaliza a publicação e transmissão de *deepfake*. Caso esses conteúdos não sejam descritos como tal, será considerado violação da lei (Teixeira, 2019).

No Brasil ainda não se tem uma legislação específica referente às deepfakes, por enquanto as leis existentes dão amparo a possíveis crimes cometidos. Siqueira (2019) aponta que:

> A Legislação Brasileira não criminaliza especificamente o "Deep Fake'. Mas os intérpretes tem buscado amparo em tipos penais abertos descritos na Lei Federal n.º 12.735/2012 (Lei Azeredo), Lei Federal n.º 12.737/2012 popularmente conhecida como Lei Carolina Dickmann, Lei Federal n.º 12.965/2014 (Marco Civil da Internet); Lei Federal n.º 13.718/2018 oriunda do Projeto de Lei n.º 5.555/2013, Lei Federal n.º 13.709/2018 - Lei Geral de Proteção de Dados Pessoais, Lei Federal n.º 13.853/2019. Além dos tipos penais descritos na Lei de Crimes Financeiro (Lei Federal n.º 7.492/86), Lei de Falências (Lei Federal n.º 11.101/2005), Código Eleitoral (Lei Federal n.º 4737/65) e principalmente nos crimes contra a honra (artigos 138/145 do Código Penal) e dignidade sexual (artigos 213/235 'c' do Código Penal).

4.3 Convenção de Budapeste

A convenção de Budapeste foi criada em 2001 pelo conselho europeu e entrou em vigor em 2004, como instrumento internacional ao combate de cibercrimes e base para o desenvolvimento de legislação para qualquer país. Os crimes cibernéticos incluem todos os crimes praticados na internet como pornografia infantil, violação à segurança de redes, violações de direitos autorais e conexos, assim como fraudes relativas a dados e sistemas (Ferrari & Senna, 2021).

O Brasil caminha para sua adesão à Convenção de Budapeste; a adesão garante ao Brasil mais ferramentas jurídicas para acompanhar as inovações de tecnologias futuras como 5G e internet das coisas (Ferrari & Senna, 2021). Para Cristaldo (2021) a adesão do Brasil à convenção de Budapeste é um passo importante no combate de crimes cibernéticos e cooperação internacional eficiente. A adesão do Brasil facilitaria o compartilhamento de provas e investigação desses crimes entre os países integrantes.

4.4 Conscientização

As *deepfake*s podem ser usadas para mostrar candidatos políticos, celebridades e pessoas "comuns" fazendo algo que nunca fizeram ou dizendo algo que jamais disseram, além de tentar influenciar a opinião pública com a potencialização das *fake news*, o uso indevido da tecnologia é muito assustador e devastador quando usada para chantagens, uso pornográfico, produção de conteúdos sexuais falsos, como vingança de ex-parceiros, incriminar pessoas, falsificação de evidências, golpes, criação de narrativas, entre muitas outras possibilidades (Greengard, 2019).

Conforme a tecnologia avança, mais realista os conteúdos se tornam, dificultando ainda mais para uma pessoa identificar se tal conteúdo é real ou falso. Existem algumas imperfeições que podem ajudar a descobrir se tal conteúdo é falso, como distorção da face, pele muito lisa, falta de naturalidade ao piscar, falta de detalhes no cabelo e dentes, movimento dos lábios, presença de sombras em lugares errados, borrões. Também deve se identificar a fonte de compartilhamento e se tal compartilhamento é compatível com a pessoal retratada no vídeo. No entanto, futuramente talvez essas medidas já não sejam mais eficazes para ajudar a identificar tais conteúdos a olho nu (Westerlund, 2019).

É imprescindível popularizar como as *deepfakes* funcionam e aumentar a conscientização sobre o uso indevido da tecnologia e ao mesmo tempo fornecer ferramentas e métodos para ajudar a identificá-la. Assim, deve se dar uma atenção maior a população mais velha e as que têm pouca experiência em tecnologia; pois é importante que as pessoas sejam capazes de avaliar a autenticidade e confiabilidade de um conteúdo compartilhado (Westerlund, 2019).

5. Considerações Finais

A tecnologia pode trazer diversos benefícios para áreas da medicina devido à facilidade na geração de dados sintéticos e para áreas do cinema, arte, educação, dependendo da necessidade e da imaginação é possível criar ou modificar conteúdos. Mesmo com os benefícios que a tecnologia pode trazer, a grande preocupação é com seu uso indevido e consequentemente os riscos que podem trazer a sociedade.

Partindo da premissa que com a tecnologia é possível mapear o rosto de qualquer pessoa e substituir por outra em qualquer situação, ou até mesmo criar uma narrativa para ser usada como base para aplicar a tecnologia, são inúmeras as possibilidades de criar conteúdos falsos. Apesar das *deepfakes* serem mais comuns no formato de vídeos, fotos e áudios também podem ser manipuladas, expandindo assim o leque de opções para enganar alguém.

A *deepfake* é uma técnica poderosa para produção de *fake news*, com o potencial de criar conteúdos falsos quase imperceptíveis, e acaba de certa forma dando autenticidade às *fake news*, ou seja, as *deepfakes* futuramente têm potencial de se tornar a evolução das *fake news*, e com o avanço da tecnologia será muito difícil identificar conteúdos modificados.

Consequentemente, as redes sociais terão dificuldades para verificar tais conteúdos. A complexidade que a tecnologia traz em conjunto ao grande volume de conteúdos que são compartilhados diariamente, poderão contribuir para disseminação de conteúdos falsos não verificados.

A tecnologia ainda não é tão popular no Brasil e para ter um resultado de qualidade, é preciso de um banco de dados robusto (fotos, vídeos) além de tempo para "treinar" a inteligência artificial e bons requisitos de *hardware*, mas é só questão de tempo para que o uso da tecnologia seja cada vez mais comum. Muitos *softwares* de *deepfakes* são de código aberto, também é possível encontrar sites que já oferecem o serviço para criação de *deepfakes* e diversos aplicativos que possibilitam a troca de rostos em cenas de filmes e séries já pré-definidos pelos aplicativos. Esses aplicativos são dedicados ao entretenimento, mas futuramente com a evolução da tecnologia a técnica ficará mais acessível, facilitando ainda mais seu uso.

No período eleitoral se falou muito sobre *fake news*, o que acabou popularizando o termo, provavelmente tudo aponta que as *deepfakes* seguiram o mesmo caminho em virtude da polarização política e o poder da tecnologia como arma de desinformação. Outro grande problema é o uso para construção de pornografia não consensual, ou como forma de vingança de ex-companheiros, ou como extorsão e chantagem, já que grande parte dos conteúdos falsos encontrados na internet são de cunho pornográfico.

Ainda não existe nenhuma lei específica para *deepfakes*, a tecnologia na mão de pessoas mal intencionadas pode trazer grandes problemas para a sociedade, e, como apresentado no artigo são inúmeras as possibilidades para seu uso, com o decorrer do tempo novas leis terão que ser criadas e tanto empresas de tecnologia, quanto redes sociais devem estar preparadas para combatê-las.

Por fim, é preciso uma conscientização do tema frente à sociedade, pois, como apontado nas pesquisas citadas, é alarmante o número de pessoas que repassam *fake news* sem ao menos verificar a fonte ou tal informação. O uso de *deepfake* como técnica de produção de *fake news* pode agravar a situação, uma vez que muitos desconhecem a técnica, por isso é importante orientar a sociedade sobre essa tecnologia e as maneiras de combatê-las.

Como sugestão para trabalhos futuros sugere-se que novas pesquisas sejam realizadas levando em consideração os avanços da tecnologia, tal evolução implica no surgimento de novos métodos de detecção e combate cada vez mais eficazes.

Referências

Agence France-Presse. (2019). *Pesquisa global revela que 86% dos internautas já acreditaram "fake news"*. Exame. https://exame.com/brasil/pesquisa-global-revela-que-86-dos-internautas-ja-acreditaram-fake-news/

Alecrim, E. (2019). *Facebook decide não excluir deepfake de Mark Zuckerberg no Instagram*. Tecnoblog. https://tecnoblog.net/294405/facebook-vai-manter-deepfake-de-mark-zuckerberg/

Almenara, I. (2021). *Algoritmo é capaz de desmascarar deepfakes analisando o movimento dos olhos*. Canaltech. https://canaltech.com.br/inteligencia-artificial/algoritmo-e-capaz-de-desmascarar-deepfakes-analisando-o-movimento-dos-olhos-180574/

Atheniense, A. (2019) *O que é Deepfake? Saiba como funciona e porque tecnologia pode afetar a política*. Alexandre Atheniense Advogados. https://www.alexandreatheniense.com.br/o-que-e-deepfake-saiba-como-funciona-e-porque-tecnologia-pode-afetar-a-politica/

Ayyub, R. (2018). *I Was The Victim Of A Deepfake Porn Plot Intended To Silence Me*. Huffpost. https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316

Cambridge Dictionary. (n.d.). fake news. In *Cambridge Dictionary*. Retrieved August 1, 2021, from https://dictionary.cambridge.org/pt/dicionario/ingles/fake-news

Campos, L. V. (2021) "*O que são Fake News?*". Brasil Escola. https://brasilescola.uol.com.br/curiosidades/o-que-sao-fake-news.htm.

Cancelier, M.(2018). *Google cria base de dados com 3 mil deepfakes para ajudar a combatê–los*. Mundo Conectado. https://mundoconectado.com.br/noticias/v/10589/google-cria-base-de-dados-com-3-mil-deepfakes-para-ajudar-a-combate-los/mobile/

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif*. L. Rev., 107, 1753. https://doi.org/10.2139/ssrn.3213954

Cristaldo, H. (2021) *Câmara aprova adesão do Brasil à Convenção sobre Crime Cibernético*. Agência Brasil. https://agenciabrasil.ebc.com.br/politica/noticia/2021-10/camara-aprova-adesao-do-brasil-convencao-sobre-crime-cibernetico

Damiani, J. (2019). *A Voice Deepfake Was Used To Scam A CEO Out Of $243,000*. Forbes. https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=63e99eda2241.

Data Science Academy. (2021). *Capítulo 54 – Introdução às Redes Adversárias Generativas (GANs – Generative Adversarial Networks).* Deep Learning Book. https://www.deeplearningbook.com.br/introducao-as-redes-adversarias-generativas-gans-generative-adversarial-networks/

Ferrari, D & Senna, F. (2020) *Convenção de Budapeste e crimes cibernéticos no Brasil*. Migalhas. 2020. https://www.migalhas.com.br/depeso/335230/convencao-de-budapeste-e-crimes-ciberneticos-no-brasil

Galhardi, C. P., Freire, N. P., Minayo, M. C. D. S., & Fagundes, M. C. M. (2020). Fato ou Fake? Uma análise da desinformação frente à pandemia da Covid-19 no Brasil. *Ciência & Saúde Coletiva*, 25, 4201-4210.https://doi.org/10.1590/1413-812320202510.2.28922020

Gogoni, R. (2021). *Era inevitável: deepfake usado como ferramenta de bullying*. Meio Bit. https://tecnoblog.net/meiobit/434802/mae-usa-deepfake-prejudica-concorrentes-da-filha/

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM* 63, 11 (November 2020), 139–144. https://doi.org/10.1145/3422622

Greengard, S. (2019). Will deepfakes do deep damage? *Communications of the ACM*, 63(1), 17-19. https://doi.org/10.1145/3371409

Hall, H. K. (2018). Deepfake videos: When seeing isn't believing. *Cath. UJL & Tech*, 27, 51. https://scholarship.law.edu/jlt/vol27/iss1/4

Hasan, H. R., & Salah, K. (2019). Combating deepfake videos using blockchain and smart contracts. *Ieee Access*, 7, 41596-41606. https://doi.org/10.1109/ACCESS.2019.2905689

Huang, D. (2021) *Synthetic data generation using Generative Adversarial Networks (GANs): Part 1*. Data Science at Microsoft. https://medium.com/data-science-at-microsoft/synthetic-data-generation-using-generative-adversarial-networks-gans-part-1-47ecbf46b575

Korshunov, P., & Marcel, S. (2019, June). Vulnerability assessment and detection of deepfake videos. In *2019 International Conference on Biometrics (ICB)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICB45273.2019.8987375

Martins, J. (2017). Metodologia da pesquisa científica. Dowbis.

Merriam Webster. (sd) *The Real Story of 'Fake News'*. https://www.merriam-webster.com/words-at-play/the-real-story-of-fake-news

Parreira, R. (2021). *80% dos legisladores não sabem o que são deepfakes e as vítimas acumulam-se. Sobretudo mulheres.* Sapo. https://tek.sapo.pt/noticias/internet/artigos/80-dos-legisladores-nao-sabe-o-que-sao-deepfakes-e-as-vitimas-acumulam-se-sobretudo-mulheres

Patrini, G., Lini, S., Ivey-Law, H., & Dahl, M. (2018) *Commoditisation of AI, digital forgery and the end of trust: how we can fix it*. Giorgio Patrini. https://giorgiop.github.io/posts/2018/03/17/AI-and-digital-forgery/

Pecsen, T. (2020). *1 a cada 2 brasileiros afirma já ter compartilhado Fake News sem saber*. dfndr blog.https://www.psafe.com/blog/1-a-cada-2-brasileiros-afirma-ja-ter-compartilhado-fake-news-sem-saber/

Robitzski, D. *A Spy used a DeepFake photo to infiltrate Linkedin networks*. The Byte. https://futurism.com/the-byte/spy-deepfake-photo-infiltrate-linkedin-networks

Schroepfer, M. (2019). *Creating a dataset and a challenge for deepfakes*. Facebook A.I. https://ai.facebook.com/blog/deepfake-detection-challenge/

Sensity Team. (2019). *Mapping the Deepfake Landscape*. Sensity. https://sensity.ai/blog/deepfake-detection/mapping-the-deepfake-landscape/

Sensity Team. (2021). *How to Detect a Deepfake Online*. Sensity. https://sensity.ai/blog/deepfake-detection/how-to-detect-a-deepfake/

Sica, N. (2022). *Mais de 65% dos brasileiros não sabem o que é "deepfake"*. Kaspersky. https://www.kaspersky.com.br/blog/brasileiros-desconhecem-deepfake/18834/

Siqueira, P. A. R. d. (2019) *O 'Deep Fake' e a Legislação Brasileira - utilização de instrumentos legais para a proteção à imagem*. Conteúdo Jurídico. https://www.conteudojuridico.com.br/consulta/artigo/53256/o-deep-fake-e-a-legislao-brasileira-utilizao-de-instrumentos-legais-para-a-proteo-imagem.

Teixeira, L. A. (2019). *Nova lei na China criminaliza deepfakes*. GQ. https://gq.globo.com/Prazeres/Poder/noticia/2019/12/nova-lei-na-china-criminaliza-deepfakes.html

Theobald, B. (2019) *Deepfakers beware: Do it in California or Texas and you'll be in deep trouble*. Fulcrum. https://thefulcrum.us/deepfake-political-video

Viera, E. (2019) *Fake News: descentralização das informações e polarização política*. Observatório da imprensa. http://www.observatoriodaimprensa.com.br/desinformacao/fake-news-descentralizacao-das-informacoes-e-polarizacao-politica/

Westerlund, M. 2019. The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11): 40-53. http://doi.org/10.22215/timreview/1282

Wiggers, K. (2019). *Generative adversarial networks: What GANs are and how they've evolved*. VentureBeat. https://venturebeat.com/2019/12/26/gan-generative-adversarial-network-explainer-ai-machine-learning/