

Fall 11-5-2022

Study of Deepfakes in Cyberspace, Impact, and Consequences on the Social Environment

Garcia Isaias Manuel
Universidade Católica Portuguesa, s-gimanuel@ucp.pt

Vítor J. Sá
Universidade Católica Portuguesa, Centro ALGORITMI / LASI, Universidade do Minho, vitor.sa@ucp.pt

Paulo Veloso Gomes
Politécnico do Porto, Escola Superior de Saúde, Laboratório de Reabilitação Psicossocial, Centro de Investigação em Reabilitação, pvg@ess.ipp.pt

António Marques
Politécnico do Porto, Escola Superior de Saúde, Laboratório de Reabilitação Psicossocial, Centro de Investigação em Reabilitação, ajmarques@ess.ipp.pt

Javier Pereira
Universidade de Coruña, CITIC, Centro de Investigación em Tecnologias de Informação e Comunicação, javier.pereira@udc.es

Follow this and additional works at: <https://aisel.aisnet.org/capsi2022>

Recommended Citation

Manuel, Garcia Isaias; Sá, Vítor J.; Gomes, Paulo Veloso; Marques, António; and Pereira, Javier, "Study of Deepfakes in Cyberspace, Impact, and Consequences on the Social Environment" (2022). *CAPSI 2022 Proceedings*. 55.

<https://aisel.aisnet.org/capsi2022/55>

This material is brought to you by the Portugal (CAPSI) at AIS Electronic Library (AISeL). It has been accepted for inclusion in CAPSI 2022 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Estudo da Utilização de Deepfakes no Ciberespaço, Impacto e suas Consequências no Ambiente Social

Study of Deepfakes in Cyberspace, Impact, and Consequences on the Social Environment

Garcia Isaias Manuel, Universidade Católica Portuguesa, Portugal, s-gimanuel@ucp.pt

Vítor J. Sá, Universidade Católica Portuguesa, Braga, Portugal, Centro ALGORITMI / LASI, Universidade do Minho, Guimarães, Portugal, vitor.sa@ucp.pt

Paulo Veloso Gomes, Politécnico do Porto, Escola Superior de Saúde, Laboratório de Reabilitação Psicossocial, Centro de Investigação em Reabilitação, Portugal, pvg@ess.ipp.pt,

António Marques, Politécnico do Porto, Escola Superior de Saúde, Laboratório de Reabilitação Psicossocial, Centro de Investigação em Reabilitação, Portugal, ajmarques@ess.ipp.pt

Javier Pereira, Universidade de Coruña, CITIC, Centro de Investigação em Tecnologias de Informação e Comunicação, Espanha, javier.pereira@udc.es

Resumo

Vivemos numa era inundados por informações vindas de ambos os hemisférios, e em que a economia da atenção nos torna distantes da verdade. O presente estudo tem como ângulo estudar a criação, utilização, bem como a partilha de vídeos originados pela inteligência artificial que podem fazer parecer que uma pessoa diz ou faz algo, embora na realidade ela nunca tenha dito ou feito nada desse tipo, esses tipos de conteúdos são denominados Deepfakes. O problema em causa, é a forma como são propagados esses conteúdos que para olho destreinado, podem ser vistos como autênticos. Para elaboração do artigo, optou-se por uma pesquisa quantitativa, por meio de um questionário, e ainda uma revisão bibliográfica em livros, artigos científicos e relatórios. Constatou-se através dos dados obtidos que, muitos utilizadores pretendem verificar a veracidade e a fonte das informações que circulam nas redes sociais digitais.

Palavras-chave: Deepfake; inteligência artificial; deep learning; ciberespaço; internet.

Abstract

We live in era in which we are inundated with information from both hemispheres, and in which the economy of attention makes us far from the truth. The present study has as its core to study the creation, use, as well as the sharing of videos originated by artificial intelligence that can make it appear that a person says or does something, although he has never said or done anything of the kind, this type of content is called deepfakes. The problem is the way in which these contents are propagated, which for the untrained eye, can be seem as authentic. To prepare the article, we opted

for quantitative research, through a survey, and a literature review in books, articles and reports. It was found through the obtained data that many users are linked to the desire to verify the veracity and source of information circulating in digital networks.

Keywords: *Deepfake; artificial intelligence; deep learning; cyberspace; internet.*

1. INTRODUÇÃO

É inegável que vivemos na era da transformação digital, pessoas e entidades têm migrado para o universo digital, desde os liberais até aos mais conservadores, iniciativas e visões de projetos idealizados no passado tais como: cidades inteligentes, internet das coisas, realidade virtual, robôs inteligentes e tantos outros têm se materializado nos dias de hoje, e todos partilham do mesmo denominador comum que é a internet. Segundo dados da *hotsuit* em parceria com a agência britânica *we are social*, a internet possuía mais de 3 bilhões de utilizadores (We are social, 2021). Para muitos a internet é um lugar de refúgio porque lhes dá a possibilidade de construir um mundo à sua imagem (E Silva & Gutiérrez, 2014, p. 47) “por exemplo por meio da adoção de avatares” (*apud* Wan & Chiou, 2006).

A prática que envolve a disseminação de conteúdos que visam a desinformação, não é tão prematura quanto aparenta, (Merriam-webster, 2017) “essa prática parece ter emergido por volta do final do século XIX”, período esse em que ainda não era usual as pessoas usarem o ciberespaço para partilhar informações com outrem.

Em tempos atuais, três décadas após o nascimento e popularização da internet, surgiram os *Deepfakes*, “uma combinação de “*deep learning* e *fake* [...]”(Westerlund, 2019), uma forma de desinformação mais poderosa, e com recursos extraordinários que fazem dela uma ferramenta de desinformação com um nível de credibilidade relativamente superior comparada as outras formas de desinformação como no caso das *fake news*. “Essas falsificações de pessoas reais costumam a ser altamente virais e tendem a se espalhar rapidamente pelas plataformas de *media* social, o que as torna uma ferramenta eficiente de desinformação” (Westerlund, 2019, p.46).

Nesta temática, o problema em causa é a rapidez com que esses produtos de vídeos que acarretam conteúdo duvidoso são partilhados no ciberespaço e conseguem atingir milhares de pessoas em fração de horas ou mesmo minutos em muitos casos. Algumas vezes, esses conteúdos são partilhados subitamente pelos internautas sem conhecer a fonte e sem saber que da sua ação podem surgir consequências gravíssimas para a comunidade em geral. “Por se apresentar como um recurso muito novo nas redes sociais digitais, muitos usuários da internet ainda não reconhecem a tecnologia produzida pelo *Deepfake*

e seu potencial poder de confundir e enganar a sociedade” (Robles-Lessa, Cabral, & Silvestre, 2020, p. 482).

2. CONHECIMENTO RELACIONADO

2.1 *Início dos deepfakes*

Muito antes surgimento dos *deepfakes* propriamente ditos, já no cinema eram utilizadas técnicas para manipular imagens e substituir rostos e corpos de atores em cenas, por causa da indisponibilidade dos mesmos, pela morte do ator, e ainda pela utilização de duplos. “[...] filmes de referência da história do cinema feitos nos Estados Unidos da América (EUA) como *The Wizard of Oz* (Victor Fleming, 1939), *Gone with the Wind* (Victor Fleming, 1939) e *Citizen Kane* (Orson Welles, 1941) utilizaram diversas técnicas de efeitos visuais na aplicação de cenários em cenas-chave [...]” (Tietzmann, 2017, p. 5).

Para a manipulação de vídeos no cinema, era necessário um grande investimento, mão de obra qualificada, muito tempo de produção e diversos procedimentos até chegar ao produto final “[...] Dependendo do tipo de efeito aplicado, as imagens passam por processos de segmentação, de transformações geométricas, de realce e de compressão, dentre outros” [...] (Gomide & Araújo, 2010). Com os *deepfakes* é exatamente o contrário, qualquer pessoa com um software e recursos computacionais adequados, consegue facilmente criar vídeos convincentes, (Harris, 2019) “o investimento e os recursos de uma produção de Hollywood é completamente desnecessário para se criara um *deepfake* em casa”.

(Kirchengast, 2020) o termo *deepfake* teve origem no final do ano 2017 quando um utilizador do *reddit* intitulado *deepfake*, começou a partilhar pornografia involuntária, que culminou num posterior banimento da conta por parte da rede social (Westerlund, 2019, p.39). *Deepfakes* são o produto de aplicativos de inteligência artificial (IA) que mesclam, combinam e sobrepõem imagens e vídeo clipes para criar vídeos falsos que parecem autênticos (*apud* Maras & Alexandrou, 2019). “Esses vídeos manipulados [...] podem fazer parecer que uma pessoa diz ou faz algo, embora, na realidade, ela nunca tenha dito ou feito nada desse tipo” (Dobber, et al., 2020).

No princípio, com o fim de popularizar a prática, eram tidos como alvos incipientes dos *deepfakes* figuras públicas do cinema, política, moda, entre outros seguimentos, muitas figuras públicas tiveram os seus rostos envolvidos em vídeos onde proferiam palavras que nunca tinham proferido ou fazendo coisas que nunca tinham feito na vida. (Westerlund, 2019, p.40.) “Deepfakes veio à tona ao público em 2017 quando um utilizador do *reddit* postou vídeos que exibiam celebridades em situações sexuais comprometedoras”, (Dobber et al., 2020, p.70) “Barack Obama, por exemplo foi uma vez ouvido e visto chamando Donald

Trump ‘um total e completo *dipshit*’ num vídeo online”, um outro caso segundo Westerlund (2019) é o de Scarlett Johansson que foi vista em *deepfake* de vídeo adulto, onde seu rosto foi sobreposto no de uma atriz porno. Hoje em dia, estão propensos a ser vítimas desta prática pessoas comuns desde que sejam utilizadores das redes e tenham conteúdos partilhados na web, (Westerlund, 2019, p.45) “Qualquer pessoa que poste uma única selfie ou um vídeo capturando 30 quadros por segundo num site ou rede social corre o risco de ser falsificado” (*apud* Brown, D. 2019).

Deepfakes tem crescido de forma exponencial e têm como principais vítimas as mulheres, são utilizadas como pornografia de vingança. Principalmente durante o confinamento provocado pela pandemia da Covid-19, aumentaram-se as partilhas e as reações no ciberespaço, o gráfico de novos utilizadores nas redes sociais dilatou de forma exponencial, em Portugal (Santos J. S., 2021, p. 18) “[...] 4/5 dos utilizadores participa nas redes sociais [...]”, é neste cenário em que os criadores de conteúdos *fakes* tiveram ou têm estado em momento de júbilo.

2.2 Como são produzidos os deepfakes?

Os *deepfakes* são produtos da inteligência artificial, sendo esta, “[...] resumidamente, é a capacidade de uma máquina através de algoritmos, possuir capacidades cognitivas semelhantes a um ser humano; com isso pode realizar atividades que só o homem era capaz” (Da Silva & Mairink, 2019, p. p.65).

A criação desses conteúdos é feita através de um repositório de arquivos já prontos, onde uma máquina é treinada a seguir determinado padrão. Essencialmente os *deepfakes* são produzidos através de Redes Adversárias Generativas (GAN’s), que não só têm a capacidade de identificar determinado padrão como também conseguem gerá-los.

Quanto a tecnologia, deepfakes são produtos de Redes Neurais Adversarias (GANs), nomeadamente duas redes neurais artificiais trabalham juntas para criar um vídeo convincente. Estas duas redes chamadas ‘the generator’ e ‘discriminator’ são treinadas na mesma base de dados de imagens, vídeos ou sons. A primeira tenta criar novas amostras que são boas o suficiente para enganar a segunda rede, que tem o papel de determinar se o novo vídeo gerado parece ser real. Desta forma, eles fundem-se um ao outro para aperfeiçoar o conteúdo. (Westerlund, 2019, p.41).

Para elaborar um vídeo falso de uma pessoa, não é necessário ter uma base de dados com imagens, áudios e vídeos da pessoa em causa. “usando muitos exemplos reais de fala e imagens em movimento, uma chamada rede neural é treinada pode ser usada para criar um *deepfake* e enganar os cidadãos” (Dobber et al., 2021, p.69-70). A existência do repositório já pronto, torna o processo de criação de vídeos mais fácil

e menos demoroso em relação as outras formas de manipulação de imagens. Sendo muitas vezes apenas necessário alterar alguns detalhes como expressão facial ou a voz, que segundo Dobber, at. all., (2021) contribui para o realismo do *deepfake*.

Os algoritmos dos *Deepfakes* possuem uma codificação automática, produzidos por meio de uma rede neural artificial utilizada para absorver dados em séries, através de um processo chamado de aprendizagem não supervisionada, este processo codifica os dados, corrigindo possíveis erros automaticamente, tornando-se próximo do conhecimento e da linguagem dos seres humanos, por isso, identificar uma imagem ou um áudio falso, tem-se tornado cada vez mais difícil. (Robles-Lessa, Cabral, & Silvestre, 2020, p. 484).

2.3 Efeito dos deepfakes no ambiente social

Devido a sua grande capacidade em dificultar a distinção do conteúdo real em relação ao conteúdo *fake*, *deepfakes* têm sido usados com mais frequência na política com o fim de alterar o pensamento e decisão dos eleitores em determinado escrutínio colocando em perigo a democracia. Dobber, at all, (2021, p.70) argumentam que “a desinformação transmitida por *deepfakes* pode representar um desafio durante as eleições, uma vez que para um olho destreinado, um *deepfake* pode ser difícil de distinguir de um vídeo real”.

Campanhas de desinformação massivas podem ser levados a cabo por agentes e partidos políticos com o fim de desacreditar seus oponentes, são usados principalmente as redes sociais para difundir essas informações desprovidas de autenticidade, tornando o processo mais eficiente. (Ferreira, 2018) “A destruição da credibilidade e o assassinato de personagens se tornam as armas políticas mais potentes” (*apud* Castells, 2007, p.243). (Vaccari & Chadwick, 2020). “[...] se não for controlada, a ascensão de *deepfakes* políticos provavelmente prejudicará a cultura cívica online, contribuindo para um clima de indeterminação sobre a verdade e a falsidade que, por sua vez, diminui a confiança nas notícias online”.

A falta de uma lei em vigor em diversos países que penalize os responsáveis dessa prática tem sido um dos fatores que fazem dela quase ilibada. (Robles-Lessa, Cabral, & Silvestre, 2020, p. 482) “A falta de conhecimento e desinformação, aliadas à ausência de uma legislação em vigor que regule esse novo mecanismo utilizado na internet, certamente irá aumentar os problemas que envolvem o uso da rede mundial de computadores quando associadas à vida privada, política e social”.

Os *deepfakes* atualmente, já não têm apenas como alvo as figuras publicamente conhecidas, há relatos de *deepfakes* apresentando pessoas que não são celebridades, a esses vídeos, Harris (2019) chamou de

“*personal deepfakes*”, que podem ser usados para ameaçar, extorquir, humilhar e chantagear as vítimas, devido ao elevado poder de credibilidade desses conteúdos, acabam por ser uma ferramenta perfeita para exercer poder sobre as pessoas supostamente expostas no vídeo.

De acordo com Anne (2020), Noelle Martin, jovem britânica, após ter feito algumas pesquisas no google, apercebeu-se que teve seu rosto envolvido em vídeos de conteúdos adultos, acompanhados de seus dados pessoais como nome e endereço. Este é daqueles casos em que mais uma vez leva a refletir a cerca do tipo de informações que as pessoas partilham em redes sociais digitais. Neste caso em particular, o fabricante teve pouco trabalho para realizar a falsificação com sucesso, já que atualmente é possível saber inúmeras informações de uma pessoa somente visitando os perfis pelos quais a pessoa atende no mundo digital.

O carácter aberto da internet e a falta de monitorização rigorosa nela, da uma certa invisibilidade ou anonimato às personagens que praticam tais atos relativos a desinformação, faz com que eventos dessa natureza se multipliquem cada vez mais, no caso das vítimas, são desencadeados vários impactos pessoais e psicológicos, bem como transtornos e desequilíbrios mentais, dentre eles o isolamento, depressão e até a possibilidade de pôr fim a própria vida, (Anne , 2020) “as implicações dos *deepfakes* pornográficos se infiltram em todas facetas da vida das vítimas”.

2.4 *Como são propagados os deepfakes*

A rapidez com que as informações, principalmente falsas fluem na rede, é fruto não apenas da partilha por parte dos utilizadores, como também de alguns mecanismos de difusão de informações nas redes sociais e outros *media*, um exemplo prático, é o caso dos *bots*, (Michalsk & De Paula, 2019, p. 3) “programas de computador criados para automatizar tarefas e procedimentos repetitivos em ambiente digital [...]” a propagação de informações por parte destes, vai depender de fatores tais como: a relevância do tópico em determinada tempo (*trending topics*) e também os agentes envolvidos (sua popularidade e engajamento *online*). (Robles-Lessa, Cabral, & Silvestre, 2020) “A maneira sofisticada pela qual informações falsas podem ser criadas, a rapidez com que podem ser criadas e o quão infinitamente elas podem ser disseminadas está em um nível diferente. No passado, eu poderia ter espalhado mentiras, mas meu alcance seria limitado” (*apud* Wachter, 2019).

Os produtores de conteúdos descontextualizados, tendem a ir longe quando o assunto é chamar à atenção dos internautas, são utilizados manchetes atraentes, títulos sensacionalistas, nomes e URL iguais ao de sites de *medias* tradicionais que gozam de alguma credibilidade, (Figueira & Oliveira, 2017, p. 819) “um

outro modelo de desinformação a realçar é aquele que consiste em publicar pedaços de notícias em *websites* com URL muito similares ao de estações de notícias populares e com boa reputação [...]”.

2.5 O comportamento dos internautas

Um grande problema das redes sociais, é a maneira com que são formados os ciclos de amizades, são criadas amizades com pessoas cujas opiniões são convergentes e são tidas como verdadeiras as notícias cujo teor está de acordo com as nossas convicções, crenças religiosas, ideologias e opiniões políticas

(Westerlund, 2019) “[...] os deepfakes não são apenas alimentados por um contexto político acalorado, mas também pelo contexto atual devido ao chamado apocalipse da informação, que faz com que as pessoas deixem de confiar em informações que não vem de suas redes sociais pessoais e são inconsistentes com as suas crenças anteriores [...]” (*apud* Britt et al., 2019; Hamborg et al., 2018; Zannettou et al., 2019).

Durante a estadia na web, somos invadidos por uma quantidade imensurável de dados e informações que ao nosso ver, parece ser sinónimo de sabedoria. O Papa Francisco em *Flatelli Tutti* adverte: “A sabedoria não se fabrica com buscas impacientes na Internet, nem é um somatório de informações cuja veracidade não está garantida [...] Assim, a liberdade transforma-se numa ilusão que nos vendem, confundindo-se com a liberdade de navegar frente a um visor” (Bergoglio, 2021).

As redes sociais em si, utilizam mecanismos de filtração de conteúdos que com os seus algoritmos avançados, fazem com os utilizadores possuam um *feed* que se limitam a exibir conteúdos com base nas pesquisas ou interações anteriores, ficando assim privados de muita informação que pode ser relevante, um dos mecanismos é o conhecido filtro bolha que basicamente faz das pessoas presas do *echo chamber* - que segundo Bozdag & Hoven (2015) resumidamente é um “universo personalizado de informações” (*apud* Pariser, 2015).

O comportamento mantido pelas pessoas enquanto utilizadores da internet é crucial para manter manutenção e estabilidade da ética no ciberespaço. Partilhar com amigos um *post* que chega através de anúncios, *feeds* ou *pop-ups*, torna-se uma atividade frequente nas redes sociais, esse ato permite a propagação de notícias verdadeiras e principalmente o de notícias falsas, (Dizikes, 2018) “Nas redes sociais, as notícias falsas viajam mais do que as notícias cujo teor é verdadeiro”.

3. METODOLOGIA

A criação de deepfakes, é algo que normalmente acontece nos bastidores, quase invisível aos olhos do público, a sua partilha ocorre com mais frequência nas redes sociais e aplicativos de mensagens

instantâneas, pelo seu elevado número de utilizadores ativos. Por este motivo, com o objetivo de perceber e analisar o comportamento dos internautas mediante a circulação de conteúdos de vídeos adulterados e descontextualizados que visam a desinformação, foi elaborado um questionário aos utilizadores do ciberespaço, em especial das redes sociais digitais.

O questionário foi constituído por um total de vinte e quatro (24) perguntas, algumas perguntas redirecionavam o questionário para diferentes secções de acordo com a resposta escolhida, sendo que nos resultados o foco foi exibir as perguntas de maior relevância ao estudo. O questionário continha perguntas introdutórias (tais como idade e género), perguntas específicas ao fenómeno *deepfakes* e outras mais genéricas.

Foi ainda feita uma vasta revisão da bibliografia em artigos científicos e relatórios.

Questão de investigação: em que medida os utilizadores que partilham deepfakes tem conhecimento da veracidade da informação que veiculam online e suas implicações?

Objetivo geral: entender o impacto e as consequências da utilização de deepfakes no ciberespaço.

Objetivos específicos

- Apresentar uma breve história e contextualização do fenómeno;
- Entender como são produzidos os *deepfakes*;
- Entender como são propagados os deepfakes no ciberespaço;
- Estudar as implicações relacionadas a partilha de deepfakes no ciberespaço.

4. RESULTADOS

4.1 *Informações gerais*

Num total de 57 pessoas inqueridas, 56.1 eram do sexo feminino e 43.8% do sexo masculino, 26.3% tinha idades compreendidas entre 21 e 25 anos de idade, 22.8% tinham de 15 a 20 anos de idade, 19.2% tinham de 26 a 30 anos, 12.3% tinham 51 ou mais anos de idade, 8% eram pessoas entre 46 e 50 anos, 5.3% entre 31 a 35 anos, e 5.3% eram pessoas de 41 a 45 anos de idade.

4.2 *Credibilidade nas notícias que circulam nas redes sociais*

Para o nível credibilidade nas notícias que veiculam em suas redes sociais, numa escala de 0 a 7, sendo zero (0) não acredito e sete (7) acredito plenamente, obteve-se o seguinte gráfico:

Sendo que as redes sociais são rodeadas de pessoas próximas ou das quais partilhamos as mesmas ideias, acredita nas notícias que vê nas suas redes sociais?

57 respostas

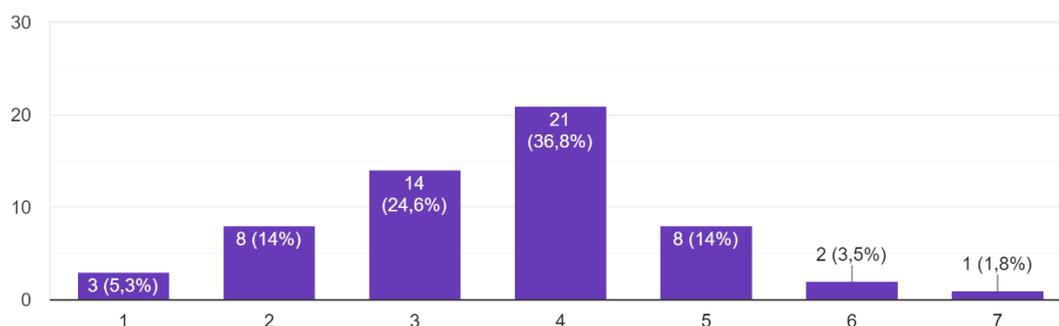


Figura 1 - Acredita nas notícias que vê nas suas redes sociais?

4.3 Conhecimento do fenómeno

Dos inqueridos, 57.9% da amostra afirmou ter conhecimentos do que são as *deepfakes* e 42.1% afirmou nunca ter ouvido falar do fenómeno.

Aos que disseram já ter conhecimento dos *deepfakes*. Com o fim de comprovar tal conhecimento, foram seleccionadas duas definições sobre *deepfakes*, da qual tinham de escolher a que acham mais certa para o fenómeno:

Qual dessas definições acha ser mais adequada para o termo Deepfake?

33 respostas



Figura 2 - O que são deepfakes?

4.4 *Contacto com as publicações de vídeo*

Sempre que te deparas com um vídeo aleatório nas redes sociais que pelo seu título parece ser interessante, qual tem sido a sua primeira atitude?

57 respostas

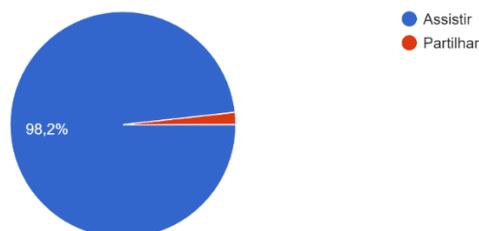


Figura 3 - Atitude tomada mediante contacto com publicações de vídeo

Aos que assistem os vídeos, 51.8% dizem reagir ou comentar o vídeo em função do seu conteúdo, 37.5 dizem que procuram obter mais informações e 10.7 dizem fazer a partilha dos vídeos com amigos.

Em relação ao possível arrependimento depois de partilhar conteúdos por conterem inverdades ou por outra razão, numa escala de um (1) a sete (7), sendo 1 nunca e 7 certamente, obteve-se o seguinte gráfico:

Numa escala de 1 a 5, já alguma vez partilhou vídeo nas redes sociais, mas teve que apagar porque descobriu que o conteúdo do vídeo não era verdadeiro ou por qualquer outro motivo?

20 respostas

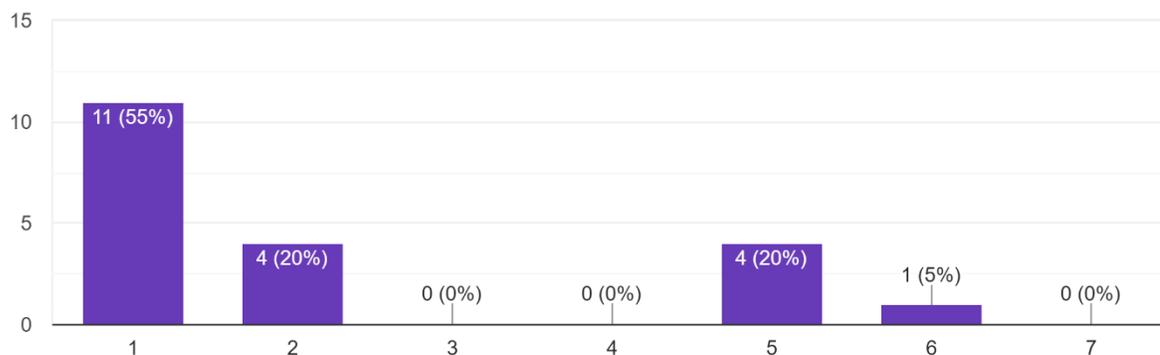


Figura 4 - Já alguma vez teve de apagar uma publicação de vídeo por conter inverdades?

4.5 *Abuso de imagem*

Dos inquiridos, 93% disseram nunca ter sofrido nenhum tipo de abuso da sua imagem, e os restantes, 7% disseram já ter sofrido abuso de imagem. Deste segundo grupo, 50% disse que foram partilhados os seus vídeos íntimos, e a outra metade, afirmou que foi editada a sua imagem com atributos diferentes da realidade. Em relação ao impacto na vida das vítimas pelo infortúnio vivido, 50% afirmou que a

experiência foi devastadora, 25% disse que a experiência foi muito devastadora e outros 25% afirmou que o impacto foi mínimo.

4.6 Perguntas gerais

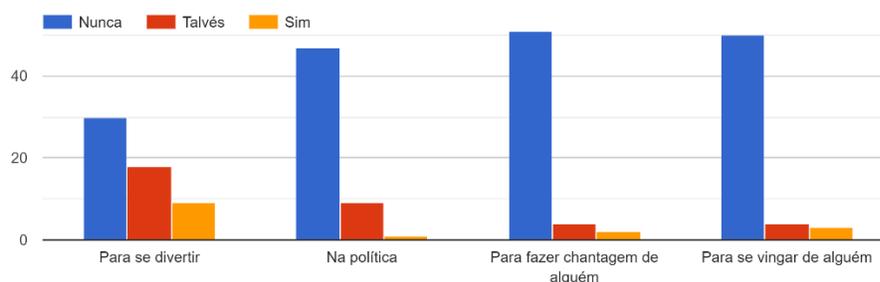


Figura 5 - Seria capaz de utilizar a tecnologia Deepfake?

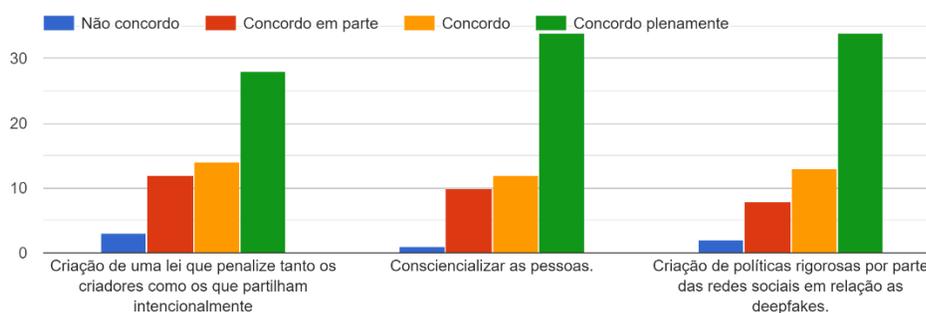


Figura 6 - Dessas medidas, quais devem ser tomadas para evitar a proliferação de Deepfakes maliciosos no ciberespaço?

5. DISCUSSÃO DOS RESULTADOS

Mediante o estudo e revisão da bibliografia realizados, é possível aferir que os Deepfakes apesar de inseridos no contexto tecnológico há muito pouco tempo, já não são uma promessa tecnológica, mas uma realidade vivida que a sociedade tem de encarar e com ela saber lidar. Podem ser auspiciosos em situações cujo ser humano é incapaz de atuar, bem como podem ser utilizados de forma anética. Mas, é crucial olhar para estas tecnologias emergentes como aliadas ao desenvolvimento socio tecnológico e não fazer delas um bode expiatório, já que imbróglis relacionados à partilha de informações estão presentes na sociedade desde o início da humanidade.

Segundo a literatura disponível, está provado que qualquer pessoa, por mais simples que seja, desde que partilhe seus conteúdos na internet, está sujeita ser vítima de uma falsificação nas redes sociais. É de realçar que a propagação de Deepfakes no ciberespaço é feita principalmente nas redes sociais, devido às

suas características aliadas a rapidez, personificação e um vasto público faminto por informações ultrajantes.

A pesquisa apontou que, por parte dos utilizadores há um equilíbrio de credibilidade nas notícias que veiculam online, cenário este que pode vir a mudar devido ao elevado número de falsificações existentes nestes, a tendência é que cada vez mais as pessoas desacreditem nas notícias e vídeos veiculados online. Os dados recolhidos apontam que em relação a propagação de vídeos nas redes sociais, maior parte dos inquiridos num primeiro contacto com vídeos, a primeira atitude é assistir aos mesmos, de seguida obtém informações sobre os vídeos para confirmar a veracidade do conteúdo que os mesmos acarretam, e só depois é que acontece uma eventual partilha. Tais atitudes mostram que os internautas são cuidadosos em relação aos conteúdos que partilham com os amigos.

Em relação às medidas a serem adotadas para evitar a propagação de Deepfakes maliciosos nas redes sociais, maior parte dos inquiridos concorda que consciencializar a sociedade e a implementação de políticas e tecnologias anti-deepfakes por parte das redes sociais são as melhores medidas a serem adotadas nesse quesito. Sendo que na segunda hipótese, já há alguns avanços, em resultado da pressão exercida por órgãos de governos, pela comunidade científica e até mesmo pelos utilizadores das redes sociais, algumas gigantes da tecnologia têm o combate a proliferação de *deepfakes* como uma prioridade, (Molina & Berenguel, 2022)“. A empresa *sensity* por exemplo desenvolveu uma plataforma que possibilita identificar vídeos e rostos gerados por GAN’s, basta apenas enviar o arquivo [...] ou inserir um URL do vídeo para que a plataforma possa analisar e verificar se tal arquivo é um deepfake (*apud* Sensity, 2021)“, (*Ibidem*) “O google, seguindo na mesma direção criou e disponibilizou uma base dados com mais de 3000 vídeos de deepfake, com a finalidade de combater conteúdos modificados pela tecnologia [...] (*apud* Cancelier, 2019)”.

A utilização de vídeos originados pelas GAN’s em Portugal ainda não é uma prática recorrente, porém, é necessário que haja uma consciencialização de que o fenómeno é real e que qualquer sujeito pode vir a ser uma vítima, principalmente em época de eleições, cenário esse em que qualquer arma desde que sirva para aniquilar o adversário, é tida como útil, mesmo sendo anética.

AGRADECIMENTOS

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

BIBLIOGRAFIA

- Anne , P. G. (2020). The New Weapon of Choice": Law's Current Inability t ent Inability to Properly. *Vanderbilt Law Review*.
- Bergoglio, J. M. (2021). *Flatelli Tutti*. Vaticano: Paulinas.
- Bozdog, E., & Hoven , J. v. (2015). Ethics and Information Technology. *Breaking the filter bubble: democracy and design*, pp. 249-265.
- Da Silva, J. A., & Mairink, C. H. (2019). Inteligência artificial: aliada ou inimiga. *Libertas*, 64-85.
- Dizikes, P. (8 de março de 2018). *Study: On Twitter, false news travels faster than true stories*. Obtido de MIT News: <https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>
- Dobber , T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (25 de julho de 2020). Sage journals. *Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?*
- E Silva, F. G., & Gutiérrez, J. C. (24 de fevereiro de 2014). Comunicar. *Internet como refugio y escudo social:* , p. 47.
- Ferreira, R. R. (2018). Observatorio. *Rede de mentiras: a propagação de fake news na pré-campanha presidencial brasileira*, pp. 139-162.
- Figueira, Á., & Oliveira, L. (2017). Procedia Computer Science. *The current state of fake news: challenges and opportunities*, pp. 817-825.
- Gomide, J. V., & Araújo, A. d. (2010). Revista de Informática Teórica e Aplicada. *Efeitos Visuais, uma Abordagem a Partir do Processamento Digital de Imagens*.
- Harris, D. (fevereiro de 2019). Duke Law & Technology Review. *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, pp. 99-12.
- Kirchengast, T. (16 de julho de 2020). Information & Communications Technology Law . *Deepfakes and image manipulation: criminalisation and control*.
- Merriam-webster. (23 de fevereiro de 2017). *The Real Story of 'Fake News'*. Obtido de Merriam-webster: <https://www.merriam-webster.com/words-at-play/the-real-story-of-fake-news>
- Michalsk, R., & De Paula, M. T. (julho de 2019). Múltiplos Olhares em Ciência da Informação. *Os bots de disseminação de informação na conjuntura das campanhas presidenciais de 2018 no Brasi*.
- Molina, A. C., & Berenguel, O. L. (2022). Research, Society and Development. *Deepfake: A evolução das fake news*.
- Robles-Lessa, M. M., Cabral, H. L., & Silvestre, G. F. (2020). Deepfake: a inteligência artificial e o algoritmo causando riscos à sociedade no ciberespaço. *Derecho y cambio social*, 482.
- Santos , J. S. (2021). *Adaptação aos novos tempos: novos hábitos de consumo online e consequente resposta das marcas*. Porto.
- Santos, H. (abril de 2005). Revista FAMECOS. *Alteridade, decepção e estigma no ciberespaço: desdobramentos da interação social mediada*, p. 42.
- Tietzmann , R. (9 de março de 2017). Intercom . *Cinema e efeitos visuais: quatro discursos a respeito de suas relações*, p. 5.
- We are social. (2021). *Global statshot report*. Londres: outubro.
- Westerlund, M. (novembro de 2019). Technology management review. *The Emergence of Deepfake Technology: A Review*.