Kalpa Publications in Computing

Volume XXX, 2022, Pages 89–91

Proceedings of V XoveTIC Conference. XoveTIC 2022

The impact and consequences of deepfakes in cyberspace on the social environment

Garcia Manuel[1], Vítor J. Sá[2], Paulo Veloso Gomes[3], António Marques[3] and Javier Pereira[3]

[1] Universidade Católica Portuguesa – Faculty of Philosophy and Social Sciences, Braga, Portugal
[2] Centro ALGORITMI, School of Engineering, University of Minho, Guimarães, Portugal
[3] LabRP, Center for Rehabilitation Research, School of Health, Polytechnic of Porto, Portugal
[4] CITIC, Research Center of Information and Communication Technologies, Talionis Research Group, University of A Coruña, A Coruña, Spain

s-gimanuel@ucp.pt; vitor.sa@ucp.pt, pvg@ess.ipp.pt, ajmarques@ess.ipp.pt, javier.pereira@udc.es

Abstract

We live in an era daily inundated with information, and the economy of attention makes us far from the truth. The present study has its core to study the creation, use and 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 content is called deepfake. The problem is the way this content is propagated, which for the untrained eye it can be seen as authentic. Quantitative research was carried out, through an inquiry and a literature review.

1 Introduction

We live in the era of digital transformation, where people and entities have definitively migrated to the digital universe. In 2021 the internet had 4.88 billion users (We are social, 2021). For many, the internet is a place of refuge because it gives them the possibility to build a world in their own image (Silva & Sendín-Gutiérrez, 2014), for example through the adoption of avatars. The practice that involves the dissemination of content aimed at disinformation is not as premature as it appears. This practice seems to have emerged around the end of the 19th century (Merriam-webster, 2017). Later, deepfakes appeared, a combination of deep learning and fake (Westerlund, 2019), a form of disinformation, more powerful and with extraordinary resources that give it a relatively superior credibility compared to other techniques of disinformation as in the case of fake news. In this context, what is at stake is the speed with which these products that carry dubious content are shared in cyberspace and manage to reach thousands of people in fractions of hours or even minutes.

2 Related knowledge

Long before the emergence of deepfakes, techniques were already used in cinema to manipulate images and replace actors with other entities. Works such as The Wizard of Oz, Gone with the Wind and Citizen Kane used various visual effects techniques to transform sets and characters (Tietzmann, 2015, p. 5). For the manipulation of videos in cinema, it was necessary a great investment, qualified labor and still a lot of production time. In the process of creating deepfakes, these features become irrelevant (Harris, 2019).

Deepfakes are product of artificial intelligence (AI) applications that mix, match, replace and overlay images and video clips to create fake videos that look authentic (Westerlund, 2019, p. 39). Deepfakes are AI products that, in short, consist in the ability of a machine, through algorithms, to have cognitive capabilities like those of a human being and thus be able to perform activities that only humans were capable of (Silva & Mairink, 2019, p. 65).

The creation of these contents is done through a ready-made file repository, where a machine is trained to follow a certain pattern. Usually, Generative Adversarial Networks (GAN) are used, where two artificial neural networks work together to create a convincing media. These two networks called 'generator' and 'discriminator' are trained on the same image, video or sound database. The first tries to create new samples that are good enough to fool the second network, which has the role of determining whether the new generated media looks real or not. In this way, they "dive" into each other to perfect the content (Westerlund, 2019).

3 Methodology

For this work, a survey was carried out with users of digital networks in order to understand and analyze the behavior assumed through the circulation of adulterated video content. A review of bibliography was also carried out in scientific articles, reports, news, among others. The research question was the following: to what extent are users who share deepfakes aware of the veracity of the information they provide online and its implications?

4 Results

In 57 responses obtained, 56.1% were female and 43.9% male. Regarding using social networks as a source of news, 45% of people said yes, 22.8% said rarely, 21.1% said they see it often and 10.5% said they always use it.

Of those surveyed, 57.9% said they had heard of deepfakes and 42.1% said they had never heard of them. Regarding the frequency with which video content is shared, 42.1% of people said they receive video posts from friends on social networks, 22.8% said they rarely receive such content, 21.1% said they receive it often and 14% said they always receive such content.

In a first contact with videos that by their title prove to be interesting, 98.2% of those surveyed say they watch them first and 1.8% say they share the videos with friends right away. Of those who watch the videos, 51.8% say they make a comment on the video depending on its content, 37.5% say they go after more information about it and 10.7% say they share it with their friends after watching it.

In case of any regret after sharing videos, for containing untruths or for another reason, on a scale of 1 to 7, with 1 being "never" and 7 "certainly", we obtained the following chart:

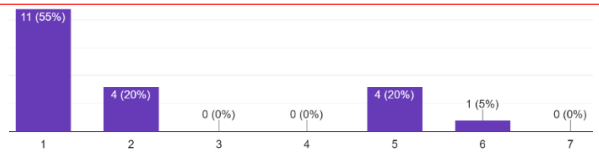

Figure 1 - Have you ever shared a video on social media, but had to delete it?

Of those surveyed, 93% said they had never suffered any abuse of their image, 7% said they had already suffered abuse. Of this second group, 50% were exposed because intimate videos were exposed, and the other half were because their images were edited with different attributes.

About a possible future use of deepfakes, we obtained the following chart:

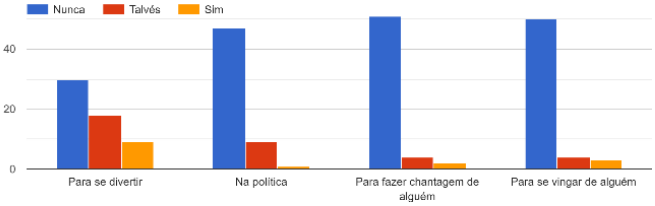

Figure 2 - Would you be able to use a deepfake?

Regarding the measures to be taken to prevent the massive proliferation of this content, we obtained the following chart:

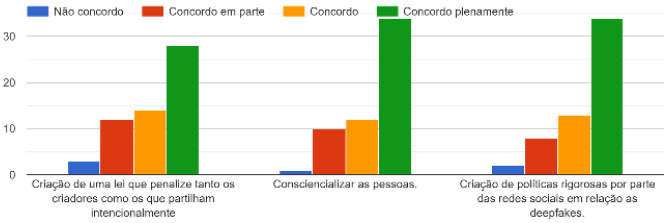

Figure 3 - Of these measures, which should be taken to prevent the proliferation of malicious deepfakes in cyberspace?

References

Harris, D. (7 de fevereiro de 2019). Duke Law & Technology Review. *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, pp. 99-127.

Merriam-webster. (23 de 03 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

Silva, F. G.-F., & Sendín-Gutiérrez, J. C. (2014). *Internet como refugio y escudo social: Usos problemáticos de la Red por jóvenes españoles*, pp. 45-53.

Silva, J. A., & Mairink, C. H. (13 de dezembro de 2019). *Inteligência artificial: aliada ou inimiga*, pp. 64-85.

Tietzmann, R. (7 de setembro de 2015). Intercom. *Cinema e efeitos visuais: quatro discursos a respeito de suas relações*.

We are social. (2021). *Digital 2021: October global statsshot report.* londres: we are social & hootsuit.

Westerlund, M. (novembro de 2019). Technology management review. *The Emergence of Deepfake Technology: A Review*, p. 40.