Revista de Ciencias de la Comunicación e Información. 2024. Vol. 29, 1-14
ISSN 2695-5016

FAKE VIDEOS AND DISINFORMATION BEFORE THE IA: DEEPFAKE AS A POST-TRUTH VEHICLE

VÍDEOS FALSOS Y DESINFORMACIÓN ANTE LA IA: EL DEEPFAKE COMO VEHÍCULO DE LA POSVERDAD

Lucia Ballesteros Aguayo[1]: *University of Málaga. Spain.*

Francisco Javier Ruiz del Olmo: *University of Málaga. Spain.*

Funding. *This text is part of the project Posverdad a debate: reconstrucción social tras la pandemia (P020_00703) (Post-truth under debate: social reconstruction after the pandemic).*

How to cite this article:

Ballesteros Aguayo, Lucia, & Ruiz del Olmo, Francisco Javier (2024). Fake videos and disinformation before the AI: deepfake as a post-truth vehicle [Vídeos falsos y desinformación ante la IA: el *deepfake* como vehículo de la posverdad]. *Revista de Ciencias de la Comunicación e Información*, 29, 1-14. https://doi.org/10.35742/rcci.2024.29.e294

ABSTRACT

Introduction: Although the use of Artificial Intelligence in the generation of audiovisual content and narratives represents an opportunity in many fields such as art or visual and graphic creation, it also becomes a powerful tool to generate false stories and representations. **Methodology:** The Exploratory Systematic Review (ESR) is applied, providing references that show with empirical evidence the image of post-truth. **Results:** A critical review of the latest studies and trends in image creation through artificial intelligence related to disinformation is provided. This is part of the contemporary audiovisual ecosystem, threatening citizens' trust in the media, social or institutional environment. **Discussion:** Users, through social networks, generate false or distorted images, which once they have gone viral are again reinterpreted by other users. Fake videos can ruin both an individual's reputation and trust in social actors. These effects could be moderated by visual and digital literacy. **Conclusions:** The deep learning of artificial neural networks generates new forms of deepfake, disconcerting for their realism and verisimilitude, and which begin to question the media, delegitimizing the representation of reality and truthful information as the basis of a democratic society.

Keywords:

Disinformation; Artificial Intelligence; Deepfake; Post-truth.

[1] **Lucia Ballesteros Aguayo:** Professor in the School of Communication at the University of Málaga. Research lines: Communication, Disinformation, Propaganda and Media Studies.

Received: 27/12/2023 - Accepted: 12/03/2024 - Published: 04/04/2024

RESUMEN

Introducción: El uso de la Inteligencia Artificial en la generación de contenido y narraciones audiovisuales si bien representa una oportunidad en muchos campos como el artístico o en la creación visual y gráfica, también se convierte en un potente instrumento para generar relatos y representaciones falsos. **Metodología:** Se aplica la Revisión Sistemática Exploratoria (RSE), aportando referencias que radiografíen con evidencias empíricas la imagen de la posverdad. **Resultados:** Se aporta una revisión crítica de los últimos estudios y tendencias en la creación de imagen mediante inteligencia artificial relacionadas con la desinformación. Ésta forma parte del ecosistema audiovisual contemporáneo amenazando la confianza de la ciudadanía en el entorno mediático, social o institucional. **Discusión:** Los usuarios, a través de las redes sociales, generan imágenes falsas o distorsionadas, que una vez viralizadas son nuevamente reinterpretadas por otros usuarios. Los vídeos falsos pueden arruinar tanto la reputación del individuo como la confianza en los actores sociales. Estos efectos podrían estar moderados por la alfabetización visual y digital. **Conclusiones:** El aprendizaje profundo de las redes neuronales artificiales genera nuevas formas de *deepfake*, desconcertantes por su realismo y verosimilitud, y que empiezan a suponer un cuestionamiento hacia los medios de comunicación, deslegitimando la representación de la realidad y la información veraz como base de una sociedad democrática.

Palabras clave:

Desinformación; Inteligencia Artificial; *Deepfake*; Posverdad.

1. INTRODUCTION

The intensive and accelerated consumption of audiovisual content, both informative and entertaining, on social networks, together with the development of powerful technology based on algorithmic logic and automatic learning, are transforming our way of approaching reality and the truth.

The increasingly pressing viralization of false stories and representations with little control by authorities, institutions, citizens and the insufficient regulation by the platforms themselves -vehicle of such false products-, is the subject of lively debates within various supranational bodies such as the European Commission, among others. In December 2023, the Council and the European Parliament reached a provisional agreement on what will be the first Artificial Intelligence Act in the European framework to ensure safe and transparent AI. The ultimate goal is to promote the adoption of reliable and human-centered AI and to protect health, safety, fundamental rights and democracy from its harmful effects (European Parliament, 2023). Among the practices prohibited by AI for its intrusive and discriminatory use, MEPs criminalized the non-selective extraction of facial images from the Internet or from closed-circuit television (CCTV) images to create facial recognition databases. Penalties for non-compliance with the law range from 35 million euros or 7% of global turnover to a fine of 7.5 million euros.

The use of AI in the dissemination of content through social networks and more specifically in its application to audiovisual materials is particularly relevant; although this emerging technology represents an opportunity in many fields such as art or audiovisual and graphic creation, it also poses a challenge for citizens when it comes to distinguishing between the real and the fictitious, between truth and lies, that is, it becomes a powerful tool for generating false stories and representations. Authors such as Shahid et al. (2022) warn that

most users lack the skills and willingness to detect fake videos and are unaware of the risks and harms involved in this type of falsehood. Specifically in countries such as India - where these researchers focus their study - it stands out the fact that even when users know that it is a fake video, they prefer not to take any action and sometimes voluntarily share videos that are not truthful but favor their worldview.

For more than a decade, there has been a trend shift in the use and consumption of online discourse that prioritizes or relies on audiovisual discourse. The popularity of video platforms such as TikTok, Instagram or YouTube is stimulating video consumption at the expense of text. So the rise of the video format is not only unstoppable but permanent and consolidated. This is evidenced by the Digital News Report (Neuman et al., 2023), which shows how in 2023 there was a gradual fall in participation in traditional platforms such as Facebook, while it increased in TikTok and other networks based almost entirely on the video format. Video news consumption is also experiencing absolute growth in all markets.

This is even more relevant when we consider that young people disproportionately consume more news videos on social networks at the same time that they are less likely to access videos on news-specific websites or apps. Specifically, the report by Neuman et al. (2023) highlights that users between 18 and 24 years old are more likely to consume short videos on TikTok, Instagram Reels and YouTube Shorts. For all these reasons and in the face of these new practices of post-truth and disinformation with false images created using AI, this paper conducts an exploratory study of the main trends and studies that address this phenomenon as a basis for future research questions and processes.

In this web of informational disorders (Wardle and Derakhshan, 2017), the 2020 Digital News Report (Neuman et al., 2020) already warned that social networks were for users a major source of concern about the spread of disinformation (40%), ahead of news sites (20%), messaging applications such as WhatsApp (14%), and search engines such as Google (10%). The use of deepfake on these online platforms is particularly serious because of its ability to generate disinformation in a very powerful way and with great verisimilitude (Langguth et al., 2021)

Spoof videos or deepfakes are audiovisual forgeries deliberately created to suggest that someone did or said something that never happened (Chesney and Citron, 2018; Nelson and Lewis, 2019). It is worth noting the exponential leap that these manipulation mechanisms have made due to their ability to distort reality in a spectacular and shocking way. This is compounded by the possibility of rapid and widespread dissemination and the fact that they can be used by users who do not need to be technologically proficient. The result is increasingly realistic and detection-resistant deep fakes.

2. OBJECTIVES

Disinformation currently has a powerful ally in the preponderance of the image and audiovisual discourses, both in the transmission of information content in the various media and in the influence of social networks, whose relationship with reality is increasingly problematic, being altered or transformed.

By means of the exploratory and descriptive review method, it is intended to pursue a main objective that is specified in three specific objectives. The main objective is to evidence the power of the image as a referential element of reality in the processes of disinformation,

pointing out the main critical contributions of this communicative trend from the role of the existing literature on the role of the image created by artificial intelligence, fake videos and deepfakes, as elements that substantially alter reality. As secondary objectives, which are drawn from the exploratory critical analysis of relevant scientific texts published on the subject, are:

- To point out the characteristics of emerging technology based on artificial intelligence and machine learning and its capacity to visually falsify reality.
- To underline the influence of deepfake videos as an element of materialization and amplification of post-truth, especially with regard to the millennial generation.
- To point out the main risks and solutions to the phenomenon of fakes using software for generating fake videos.

3. METHODOLOGY

Since the generation of images for misinformation using artificial intelligence is a relatively recent phenomenon that is now beginning to become widespread, this paper has opted, in connection with the proposed objectives, for a methodology based on Exploratory Systematic Review (ESR), following Booth et al. (2012) and Munn et al. (2018). This is a type of critical review of studies, sources and references that allow exploring trends in the reflection on the proposed topic and synthesizing empirical evidence, in this case on the image of post-truth. This not only provides the scientific community with a critique of an emerging phenomenon, but also aims to generate new lines of research.

One of the objectives of ESR is to show a broader and more contextualized picture of the phenomenon under study, which differs from a traditional systematic review, which is more specific and exhaustive. ESR is applied in several stages that include the development of the research question and a search strategy, the literature search, the review and selection of studies, the extraction of data and trends, and the analysis and reporting of results. This research has opted for a CSR that shows a broad, rather than exhaustive, picture through "systematic approaches" (Booth et al., 2012) to address the trends and concerns raised by AI image dissemination.

An exploratory review of scientific literature was conducted since 2017, the year of the beginning of the massive dissemination and repercussion of fake videos through social networks. A search was made for both scientific articles, of applied type, experimental or non-experimental design, as well as theoretical discussion articles, and also a selection of monographs in the Scopus, Proquest, Ebsco and Dialnet databases. To process the search, the Boolean operator "and" was used in English and Spanish. The terms used were "Vídeos falsos" (Fake videos), "Deepfake", "Inteligencia Artificial" (Artificial Intelligence). A temporal delimitation was applied between the years 2017 and 2023 and a geographical delimitation referred to Europe, Latin America and Spain.

Most of the documents are logically shown in several databases at the same time; the selection process was based on the reading and analysis of both the title and the abstract of the document; and as for the selection criteria, it is necessary to point out that this is a critical and exploratory review, so it is not intended to show all the results, but a qualitative and interpretative selection of those that integrate the search terms and offer relevant

contributions to the specific objectives set. Thus, several scientific contributions of these texts are pointed out and integrated into this paper.

4. RESULTS AND DISCUSSION

Fake, decontextualized or altered images are an almost inescapable part of the contemporary audiovisual ecosystem. With different motivations and purposes, these images spread a permanent feeling of mistrust towards the media, including those specialized in information. Certainly, at times these same media have acted unethically, but in any case, today's visual disinformation, within the reach of any user, and the viralization of messages can raise this problem to a structural level and to the level of a social or institutional threat. For example, the proliferation of fake videos related to the performance of institutions and their democratic representatives is one of the main challenges of this technology (Dan et al., 2021).

Of all the threats reported in the literature, two broad categories seem particularly relevant, as established by Chesney and Citron (2018) or Vaccari and Chadwick (2020), among others. First, fake videos can ruin the reputation of the individual involved. For example, in the case of a political candidate, a fake video could affect public attitudes and threaten the individual's political success. Second, detrimental spillover effects may arise, such as widespread distrust of social and political actors and a sense of unreality and confusion about what is real and what is not. Such effects could be mediated by perceived (subjective) realism and moderated by visual and digital literacy, among others.

Some of the most clarifying definitions of deepfake are presented below, given the scarce literature on the subject. Selected contributions also highlight the challenges posed by the unethical use of this technology.

Table 1. Deepfake definitions.

Cover, R. (2022). *Deepfake culture: the emergence of audio-video deception as an object of social anxiety and regulation*.	Deepfakes draw on algorithmic powers, machine learning and modern information processing capabilities to allow users to insert the face, body and visual information of a real-world person into a fake scenario, producing very convincing videos that appear to be a "real" record.
Langguth, J., Pogorelov, K., Brenner, S., Filkuková, P., & Schroeder, D. T. (2021). *Don't Trust Your Eyes: Image Manipulation in the Age of DeepFakes.*	It is a novel technology that enables the economical manipulation of video footage using artificial intelligence.

Vaccari, C., & Chadwick, A. (2020). *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*.	Deepfakes are synthetic videos that closely resemble real videos. Integrating theories about the power of visual communication and the role uncertainty plays in undermining trust in public discourse.
Nelson, A., & Lewis, J. A. (2019). *Trust your eyes? Deepfakes policy brief.*	Deepfakes are near-perfect video and audio forgeries produced by artificial intelligence programs that produce seemingly realistic but fabricated images and sounds that show people doing and saying things that never happened.
Chesney, R., & Citron, D. K. (2018). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security.	Machine learning techniques are increasing the sophistication of the technology, making deep fakes increasingly realistic and increasingly resistant to detection.

Source: Elaborated by the authors.

While in countries like Spain the most notable scandals of videos and/or images faked using artificial intelligence do not appear until 2023 -this is the case for example of the minors in Almendralejo (Badajoz) who spread fake nude photographs of teenage girls, the fake nude of the singer Rosalía or that of the model Laura Escanes-, deepfake technology has its origin in 2017 when videos appeared on Reddit with faces of famous women like Gal Gadot or Scarlett Johanssen grafted onto bodies of porn actors (Cole, 2017; Langguth et al. , 2021).

Hence, and as of 2017, consumer image manipulation software that made use of machine learning gained public attention (Dan et al., 2021). However, in a paper published five years earlier, Krizhevsky et al. (2012) evidenced how deep learning, a refinement of artificial neural networks, established itself as a superior technology for image recognition. For Langguth et al. (2021) that paper demonstrated that when trained with a large number of suitable input images, convolutional neural networks (CNNs) can categorize the content of an image with high accuracy. Thus, a CNN can be trained to recognize specific people and reliably distinguish them in a wide range of images. The prerequisites for this are a powerful computer and a large number of images from which the CNN can learn. The result, say these authors, is an image created from the abstract features that the network has learned.

Thus, while the software is relatively affordable and requires little sophistication, it needs a large number of training images to function properly. Consequently, Langguth et al. (2021) warn that these programs are not practical for creating manipulated videos of an average person. For that reason, most deepfake videos that are made for entertainment purposes feature famous actors of whom there are many publicly available images.

The term cheapfake arises in line with this availability of image manipulation tools through AI. Gamir-Ríos and Tarullo (2022, p. 102) note that "the crudeness of their elaboration makes them more identifiable as fraudulent, cheapfakes can be produced without the need

for advanced technological skills or sophisticated software" and achieve a similar effect of disinformative pollution (Dowling, 2021), confirming pre-existing judgments (Weeks and Garrett, 2014).

The truth is that the manipulation of photographs and videos has a long history, but only recent technological advances have undermined the reliability of image evidence as a referential element of reality.

4.1. Audiovisual forms of post-truth.

An example of this type of online video is the fall 2018 posting on the U.S. entertainment media outlet BuzzFeed featuring Obama claiming that "President Trump is a total and complete moron" dubbed by actor Jordan Peele with the title "Hey you! I won't believe what Obama says in this video! 😉 " which garnered over 5 million views on YouTube.

In the context of today's media consumption, these practices are not only common, but increasingly elaborate; there is a proliferation of vicarious experiences mediated by their representation mainly through videos and images, fragmentary, subjective and emotional in nature, denying the absolute nature of knowledge. In the contemporary media system, post-truth would thus be a "media inflation" of the postmodern (Ferraris, 2019).

The cultural operations of post-truth and the image used as its main vehicle are often informed by a partial, biased, self-interested and relativistic perspective (Prozorov, 2019). So that everything that was once directly experienced has now been transferred to a representation (Debord, 1967). This is precisely what sustains the concept of the notarial image: it seems irrefutable when it is shown to us in a photograph (Sontag, 1981).

In relation to who and when disinforms, three major groups or types of disinformation, are highlighted, that is, the way in which post-truth is concretized in media discourses:

- Discourses that question reality through satirical or parodic elements: caricatures and memes with all their derivatives.
- Of an opportunistic or immediate nature: clinging to the speed and vertiginousness of the news and events narrated by the media.
- Malicious disinformation: that which seeks to radically distort public perception of a political, social or scientific issue, among others.

It is observed that image manipulations through deepfake or artificial intelligence are progressively placed in the last group, since they work with deep manipulations and highly distorting reality.

4.2. The image: vehicle of post-truth

Post-truth practices enhance the aforementioned characteristics through the still image or audiovisual narration. This is due to the polysemic character of the image: it is deeply ambivalent about its relationship with reality. Hence, it underlines its capacity to disguise itself or to appear similar to the truth. Added to this is the emotional and instantaneous character, so that it is a format that multiplies the distorting capabilities of post-truth: the primacy of the emotional over the rational (Hameleers et al., 2020).

The meaning of the images does not necessarily have to be related to the technical device, to the technology associated with the capture of the images, that is, it is not an automatic result of the materiality of the image, according to Comolli (2010), as detailed in his essay on Technique and Ideology, but to the interpretation of the image. Thus, the image reflects and directly informs about the reality represented, but at the same time it has a powerful capacity for evasion, subjectivity and interpretation, driving fuels for the distortion of that reality. For example, with regard to images that are based on false or distorted content, three general categories can be distinguished:

- The decontextualization images.
- Memes.
- Manipulated images: these correspond to highly sophisticated and profoundly transformative operations, which are generated through artificial intelligence or other radical alteration techniques.

In addition, and as another key element that consolidates the validation of image falsifications, media post-truth narratives tend to validate and reaffirm our pre-existing opinions, allowing us to adjust the facts to our personal belief system (Lynch, 2016). Thus, following Lilleker and Liefbroer (2018), emotions and opinions that appear credible are more powerful than reasoned and verified ones. In this context, it is worth noting the cognitive preeminence of emotions and intuition; automatic, non-reflective, and non-critical thinking; the trend towards quick and reactive decisions in the face of spontaneous stimuli; and the origin of biased or strategic judgment (Lewandowsky et al., 2017; McDermott, 2019).

Finally, the critical reflection around post-truth images, their nature and meaning, has been historically and pioneeringly undertaken by two instances, before it became a major social and institutional problem and is nowadays already addressed more systematically by the academic community and also by the journalistic profession:

A. The digitization of the image and its interested transformation, and analog photography, is also a representation and a manipulation (Marzal-Felici, 2021).
B. The intentional generation of fake images and reflection: it challenges the relationship of images with reality (political art) whose main objective is not to develop false narratives, but to allow viewers to develop a critical attitude towards the images they consume (Fontcuberta, 2017).

4.3. Fake images created with neural networks and their verification

Every communicative process requires a technological base to be carried out, due to its inherent character of representation and also because of its necessary transmission and dissemination. Post-truth discourses increase and go viral thanks to digitalization. Technology allows the creation of practically autonomous content -both in text and audiovisual format- through artificial intelligence and neural networks that generate profound alterations in the nature and meaning of discourse, currently being tools that generate fake images. The development of this type of tools supported by AI is a milestone in contemporary visual culture, since it produces substantial changes in the meaning of the image, its generation and distribution.

Moreover, the current media ecosystem is defined by the intervention of audiences in the production processes, in the specialized digital offer, enhancing interactivity in communication, multiscreen consumption and transmedia narratives. So, it is often the users, with or without malicious interest, who generate false or distorted images, which once they have gone viral are reinterpreted by other users.

Neural networks that apply artificial intelligence or other digital tools to transform and distort stories, news and information are both generators and verifiers of these discourses. Companies that own and manage this technology, aware of the challenges and potential dangers and other ethical considerations that its use entails, have put in place various self-monitoring and verification systems.

Some recent examples of this issue can be found in the case of ChatGPT software, which although it is not able to discern whether the information is true or false, it can provide additional information that will allow the user to obtain solid external criteria for verification. Therefore, it is not possible to create a fake news directly using ChatGPT. Upon the request to create a piece of false information or to create a piece of news that is poorly contrasted or lacks argument, the AI does not generate text, but warns the user that it cannot create such information by its protocols. The AI then has internal tools to identify fake news and responds to the basic parameters of media literacy (Dan et al., 2021; Qian et al., 2022).

Another recent example comes from the technology giant Intel, which offers Real-Time Deepfake Detector, a platform for analyzing videos, reaffirming the concept of responsible AI. The tool, one of the world's first real-time deepfake detection platforms, offers a high accuracy rate and is deployed in various industries and platforms, including social networking tools, news agencies, broadcasters, content creation tools, startups and non-profit organizations (Intel, 2022). This deepfake detector is based on the study of color changes in faces by inferring blood flow, a process called photoplethysmography. Researchers Ilke Demir and Umur Ciftci designed the software to focus on certain color patterns in certain facial areas and ignore others, making it possible to verify the reality or fakery of the images.

5. CONCLUSIONS

The massive, accelerated and persistent use, distribution and consumption of images in the contemporary communicative ecosystem entails substantial challenges for our societies. The need for a regulatory framework to govern the uses and limits of AI is one of the priorities of supranational bodies such as the European Union.

The polysemic capacities of the image, its emotional and representative fixation, give it a central power in the materialization and amplification of post-truth, hence the harmful and multiplying effects of the falsification of images and videos through deepfake.

Artificial intelligence can, with an increasing level of realism, construct, imagine and transform current events and narratives in order to subvert them in a profound way. There is an urgent need for critical reflection and the provision of solutions to prevent the harmful consequences of the use of this technology for the wrong purposes, such as electoral fraud, defamatory misuse, cyberbulling or pornography of public figures or anonymous citizens (Cover, 2022; Langguth et al., 2021).

In addition, the fact that hyper-realistic fake videos using face swapping go viral has a strong impact on the credibility that people have towards audiovisual evidence in journalism (Shin and Lee, 2022). Hence, journalists and the media are clearly concerned and feel a responsibility to debunk these fake videos and avoid manipulation of public opinion (Pérez Dasilva et al., 2021). This latest research focused on the use of deepfake on Twitter highlights how in all the tweets analyzed, the news media warn about the potential danger of this technology. This is consistent with recent studies such as those contributed by Yadlin-Segal and Oppenheim (2020) that show how journalists frame deepfakes as a destabilizing platform that undermines a shared sense of sociability and political reality.

The scientific literature (Chesney and Citron, 2018; Vaccari and Chadwick, 2020) evidences that the audiovisual deepfakes with the greatest impact are, on the one hand, synthetic political videos that generate deception and are intended to undermine the reputation of the politician or public personality involved. On the other hand, harmful indirect effects can also arise, such as distrust in social and political actors and a generalized sense of confusion about what is real and what is not. Second, detrimental spillover effects may arise, such as widespread distrust of social and political actors and a sense of unreality and confusion about what is real and what is not. All this invites us to re-evaluate the cultural ethics of communication, all the more so when a growing number of software tools are nowadays available to any user, making it possible to create synthetic images detached from the physical referent.

The widespread use of fake images in disinformation poses an additional risk to the credibility of institutions and constitutes an economic and creative challenge for contemporary societies. In contrast, the volume and speed of disinformation proliferating through online platforms are a source of enormous profits for those online platforms (Dan et al., 2021). The cultural operations of post-truth and the image used as its main vehicle are often informed by a partial, biased, self-interested and relativistic perspective (Prozorov, 2019).

Among the solutions, the effects of creating misrepresentations using AI could be mediated by perceived realism and moderated by visual and digital literacy (Dan et al., 2021; Qian et al., 2022). So digital media literacy intervention may motivate reverse search for out-of-context visual misinformation. Other solutions (Chesney and Citron, 2018) target industry involvement with technological responses, criminal sanctions, civil liability, or regulatory actions.

6. REFERENCIAS

Booth, A., Papaioannou, D. y Sutton, A. (2012). *Systematic approaches to a successful literature review*. Sage.

Chesney, R. y Citron, D. K. (2018). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3213954

Cole, S. (2017). *AI-assisted fake porn is here and we're all fucked. Motherboard*. VICE. https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn

Comolli, J. L. (2010). *Cine contra espectáculo seguido de Técnica e ideología: 1971-1972.* Manantial.

Cover, R. (2022). Deepfake culture: The emergence of audio-video deception as an object of social anxiety and regulation. *Continuum*, *36*(4), 609-621. https://doi.org/10.1080/10304312.2022.2084039

Dan, V., Paris, B., Donovan, J., Hameleers, M., Roozenbeek, J., Van Der Linden, S. y Von Sikorski, C. (2021). Visual Mis- and Disinformation, Social Media, and Democracy. *Journalism & Mass Communication Quarterly*, *98*(3), 641-664. https://doi.org/10.1177/10776990211035395

Debord, G. (1967). *The society of the spectacle*. Zone Books.

Dowling, M. E. (2021). Democracy under siege: Foreign interference in a digital era. En *Australian Journal of International Affairs*, *75*(4). Australian Institute of International Affairs (pp. 383-387). https://doi.org/10.1080/10357718.2021.1909534

European Parliament (2023). *MEPs ready to negotiate first-ever rules for safe and transparent AI.* European Parliament. https://acortar.link/a9Os2l

Ferraris, M. (2019). *Posverdad y otros enigmas*. Alianza.

Fontcuberta, J. (2017). *La furia de las imágenes. Notas sobre la postfotografía*. Galaxia Gutenberg.

Gamir-Ríos, J. y Tarullo, R. (2022). Predominio de las *cheapfakes* en redes sociales. Complejidad técnica y funciones textuales de la desinformación desmentida en Argentina durante 2020. *AdComunica*, 23, 97-118. https://doi.org/10.6035/adcomunica.6299

Hameleers, M., Powell, T. E., Van Der Meer, T. G. L. A. y Bos, L. (2020). A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication*, *37*(2), 281-301. https://doi.org/10.1080/10584609.2019.1674979

Intel (2022). *Intel Introduces Real-Time Deepfake Detector*. https://acortar.link/6RKmBp

Krizhevsky, A., Sutskever, I. y Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. En F. Pereira, C.J.C. Burges, L. Botton y K.Q. Weinberger (Eds.), *Advances in neural information processing systems*, 25 (pp. 1097–1105). Currant Associates, Inc. https://acortar.link/jGDz7g

Langguth, J., Pogorelov, K., Brenner, S., Filkuková, P. y Schroeder, D. T. (2021). Don't Trust Your Eyes: Image Manipulation in the Age of DeepFakes. *Frontiers in Communication*, 6, 632317. https://doi.org/10.3389/fcomm.2021.632317

Lewandowsky, S., Cook, J. y Ecker, U. K. (2017). Letting the gorilla emerge from the mist: Getting past post-truth. *Journal of Applied Research in Memory and Cognition*, *6*(4), 418-424. https://doi.org/10.1016/j.jarmac.2017.11.002

Lilleker, D. G. y Liefbroer, M. (2018). 'Searching for something to believe in': Voter uncertainty in a post-truth environment. *International Journal of Media & Cultural Politics*, *14*(3), 351-366. https://doi.org/10.1386/macp.14.3.351_1

Lynch, M. P. (2016). *The Internet of Us: Knowing More and Understanding Less in the Age of Big Data*. W. W. Norton & Company.

Marzal-Felici, J. (2021). Proposals for the study of images in the post-truth era. *El Profesional de la Información*, *30*(2). https://doi.org/10.3145/epi.2021.mar.01

McDermott, R. (2019). Psychological underpinnings of post-truth in political beliefs. *PS: Political Science & Politics*, *52*(2), 218-222. https://doi.org/10.1017/S104909651800207X

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A. y Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18. https://doi.org/10.1186/s12874-018-0611-x

Nelson, A. y Lewis, J. A. (2019). *Trust your eyes? Deepfakes policy brief*. Center for Strategic and International Studies.

Neuman, N., Fletcher, R., Eddy, K., Robertson, C. T. y Nielsen, R. K. (2023). *Reuters Institute Digital News Report 2023*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023

Neuman, N., Fletcher, R., Schulz, A., Andı, S. y Nielsen, R. K. (2020). *Reuters Institute Digital News Report 2020*. Reuters Institute for the Study of Journalism. https://acortar.link/p4gn2f

Pérez Dasilva, J., Meso Ayerdi, K. y Mendiguren Galdospin, T. (2021). Deepfakes on Twitter: Which Actors Control Their Spread? *Media and Communication*, *9*(1), 301-312. https://doi.org/10.17645/mac.v9i1.3433

Prozorov, S. (2019). Why is there truth? Foucault in the age of post-truth politics. *Constellations*, 26, 18-30. https://doi.org/10.1111/1467-8675.12396

Qian, S., Shen, C. y Zhang, J. (2022). Fighting cheapfakes: Using a digital media literacy intervention to motivate reverse search of out-of-context visual misinformation. *Journal of Computer-Mediated Communication*, *28*(1), zmac024. https://doi.org/10.1093/jcmc/zmac024

Shahid, F., Kamath, S., Sidotam, A., Jiang, V., Batino, A. y Vashistha, A. (2022). "It Matches My Worldview": Examining Perceptions and Attitudes Around Fake Videos. En *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-15) Association for Computing Machinery. https://doi.org/10.1145/3491102.3517646

Shin, S. Y. y Lee, J. (2022). The Effect of Deepfake Video on News Credibility and Corrective Influence of Cost-Based Knowledge about Deepfakes. *Digital Journalism*, *10*(3), 412-432. https://doi.org/10.1080/21670811.2022.2026797

Sontag, S. (1981). *Sobre la fotografía*. Edhasa.

Vaccari, C. y Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, *6*(1). https://doi.org/10.1177/2056305120903408

Wardle, C. y Derakhshan, H. (2017). *Information Disorder. Toward an interdisciplinary framework for research and policy making (DGI [2017] 09)*. Council of Europe report. https://acortar.link/2sjVWZ

Weeks, Brian E. y Garrett, R. K. (2014). Electoral Consequences of Political Rumors: Motivated Reasoning, Candidate Rumors, and Vote Choice during the 2008 U.S. Presidential Election. En: *International Journal of Public Opinion Research*, *26*(4). Oxford University Press (pp. 401–422). https://doi.org/10.1093/IJPOR/EDU005

Yadlin-Segal, A. y Oppenheim, Y. (2021). Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence*, *27*(1), 36-51. https://doi.org/10.1177/1354856520923963

AUTHORS' CONTRIBUTIONS, FUNDING AND ACKNOLEDGMENTS

Conceptualization: Ballesteros Aguayo, Lucía and Ruiz del Olmo, Francisco Javier. **Methodology:** Ruiz del Olmo, Francisco Javier and Ballesteros Aguayo, Lucía. **Validation:** Ruiz del Olmo, Francisco Javier and Ballesteros Aguayo, Lucía. **Formal analysis:** Ballesteros Aguayo, Lucía and Ruiz del Olmo, Francisco Javier. **Data curation:** Ballesteros Aguayo, Lucía and Ruiz del Olmo, Francisco Javier. **Drafting-Preparation of the original draft:** Ballesteros Aguayo, Lucía and Ruiz del Olmo, Francisco Javier. **Drafting-Revision and Editing:** Ballesteros Aguayo, Lucía and Ruiz del Olmo, Francisco Javier. **Visualization:** Ballesteros Aguayo, Lucía and Ruiz del Olmo, Francisco Javier. **Supervision:** Ballesteros Aguayo, Lucía. **Project management:** Ballesteros Aguayo, Lucía. **All authors have read and accepted the published version of the manuscript:** Ballesteros Aguayo, Lucía and Ruiz del Olmo, Francisco Javier.

Funding: This text is part of the project *Posverdad a debate: reconstrucción social tras la pandemia* (P020_00703) (Post-truth under debate: social reconstruction after the pandemic).

Acknowledgments: Thanks go to the members of the project *Posverdad a debate: reconstrucción social tras la pandemia* (P020_00703) (Post-truth under debate: social reconstruction after the pandemic).

Authors:

Lucia Ballesteros Aguayo
University of Malaga.
PhD *Cum Laude* and International Major, as a result of obtaining International Grants for research stays at the University La Sapienza (Rome) and the University of Aveiro (Portugal). Accredited by ANECA as a Hired Doctor. She graduated from Journalism and Advertising and PR and she holds a Master's Degree in Teacher Training from the Complutense University of Madrid. She is currently a professor at the School of Communication of the University of Malaga and Teaching Coordinator of subjects based on ICT. Research lines: Communication, Disinformation, Propaganda and Media Studies.
Índice H: 4.
Google Scholar: https://scholar.google.com/citations?user=wcxyZPUAAAAJ&hl=en

Orcid ID: https://orcid.org/0000-0003-1191-4070
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57211473463

Francisco Javier Ruiz del Olmo
University of Malaga.
Professor at the University of Malaga, in the area of Audiovisual Communication; his teaching and research work is developed in the Schools of Communication Sciences and Fine Arts. He has researched the communicational models and contemporary forms of the media, as well as their technical and social uses; a second line of research that he develops is related to communication He is currently co-directing the Knowledge Generation Research Project of the Ministry of Science and Innovation "AV-ADJUVEN:n and new media. Repertoires and media practices in adolescence and youth: reception and use of online audiovisual platforms."
Índice H: 12.
Google Scholar: https://scholar.google.es/citations?user=RphOTRkAAAAJ&hl=es&oi=ao
Orcid ID: https://orcid.org/0000-0002-1953-1798
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=55581055500

Revista de Ciencias de la Comunicación e Información
ISSN: 2695-5016

RELATED ARTICLES:

Anton-Bravo, A. y Serrano Tellería, A. (2021). Innovación en la docencia del periodismo a través de la ciencia de datos. *European Public & Social Innovation Review*, *6*(1), 70-84. https://pub.sinnergiak.org/esir/article/view/150

Encinillas García, M. y Martín Sabarís, R. (2023). Desinformación y Salud en la era PRECOVID: Una revisión sistemática. *Revista de Comunicación y Salud*, 13, 1-15. https://doi.org/10.35669/rcys.2023.13.e312

Fernández Fernández, M. (2021). Informaciones, crónicas y relaciones del siglo XVI como antecedentes de las fake news. El caso de la "invención" de San Segundo. *Historia y Comunicación Social*, *26*(2), 593-601. https://doi.org/10.5209/hics.68530

Toro González, S. y Pérez-Curiel, C. (2021). Populismo político en tiempos de COVID. Análisis de la estrategia de comunicación de Donald Trump y Boris Johnson en Twitter. *Revista de Comunicación de la SEECI*, 54, 1-24. https://doi.org/10.15198/seeci.2021.54.e700

Zúñiga, F., Mora Poveda, D. A. y Molina Mora, D. P. (2023). La importancia de la inteligencia artificial en las comunicaciones en los procesos marketing. *Vivat Academia*, 156, 19-39. https://doi.org/10.15178/va.2023.156.e1474