

Deepfakes, Seeing is Believing?

Claudia López Frías

Universidad Complutense de Madrid, Spain.

Deepfakes, вярваш на това, което виждаш?

Клаудия Лопес Фриас

Университет Комплутенсе де Мадрид, Испания.

Author Note

Claudia López Frías  <https://orcid.org/0000-0001-8812-8890>

The author has no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to: Ph. D. Claudia López Frías, Facultad de Ciencias de la Información. Departamento de Periodismo y Nuevos Medios. Despacho 510 (12) - 5^a planta. Av. Complutense, 3 28040 – Madrid. Spain. Email: Claudia.lopez@ucm.es

Бележка на автора

Claudia López Frías  <https://orcid.org/0000-0001-8812-8890>

Авторът няма известен конфликт на интереси, който да разкрие.

Кореспонденцията относно тази статия трябва да бъде адресирана до Ph.D. Клаудия Лопес Фриас, Facultad de Ciencias de la Información. Departamento de Periodismo y Nuevos Medios. Деспачо 510 (12) - 5^a planta. Av. Complutense, 3 28040 – 3 28040 – Мадрид, Испания. Имейл: Claudia.lopez@ucm.es

Abstract

Artificial intelligence has burst into our lives and today it is a constant in our routines. Initially, it represented a great advance in everything that until now limited the human being. It also seems to have a series of risks that we are only beginning to be aware of. One of the technologies derived from AI is that of deepfakes, words that arise as an acronym when combining Deep, coming from how these intelligences learn, through deeplearning, and fake, whose translation would be false or falsification. It is a technology widely used in audiovisuals, both for cinema, television, or other types of videos, with variable results. However, the use of pornography of celebrities without their consent, or the deliberate interest in misinforming, at critical moments, are some of the problems faced by the use of this technology. Something that seems to reach a critical point when it comes to news since the presenters usually enjoy greater credibility for the viewer. Throughout this work, we will try to approach the use of artificial intelligence on television as well as Deepfakes, their uses, and risks.

Keywords: deepfake, artificial intelligence, disinformation, new technologies

Резюме

Изкуственият интелект нахлу в живота и днес е неизменна част от нашето ежедневие. Това, което в началото е означавало голям напредък във всичко, което досега е ограничавало човешкото същество, крие серия от рискове, за които едва се започва да осъзнава. Една от технологиите, извлечени от AI, са дълбоки фалшификати, широко използвани в аудиовизуалните продукти с различни резултати. Използването на порнография от известни личности без тяхното съгласие с цел печалба или умишленият интерес от дезинформиране, са някои от проблемите, с които се сблъсква използването на тази технология. В тази статия се опитваме да разгледаме използването на изкуствен интелект по телевизията, както и употребите и рисковете на Deepfakes.

Ключови думи: deepfake, изкуствен интелект, дезинформация, нови технологии

ARTICLE INFO:

Original Article

Received: 05, 09.2023

Revised: 15, 09.2023

Accepted: 01, 10.2023

Deepfakes, seeing is believing?

Introduction

Artificial intelligence (AI) has burst into our lives in recent years. We can find a multitude of definitions on a topic as complex as AI and there is no universally accepted definition. Definitions of AI include; «The ability of computers to do activities that normally require human intelligence» (Rouhiainen, 2018, p. 17) or «software (and possibly also hardware) systems designed by humans that, faced with a complex objective, act in the physical or digital dimension: perceiving your environment, through the acquisition and interpretation of structured or unstructured data, reasoning about the knowledge, processing the information derived from this data and deciding the best actions to achieve the given objective. AI systems can use symbolic rules or learn a numerical model and can also adapt their behavior by analyzing how the environment is affected by their previous actions» (European Commission, 2020, p. 4). These are definitions that give us a clue about the characteristics of the cognitive functions of this technology, such as the processing of information (in this case data) to solve different tasks. However, these definitions have evolved according to the application of the technology itself, which has allowed us to understand the different approaches and purposes that can be given to AI.

This is a concept that, although it may seem completely new to us, was proposed by the father of computing, the mathematician Alan Turing in the 1930s, with the Turing Machine, which could decipher all possible calculations, as well as carry out problem-solving for which there was no algorithmic solution, a thesis that was later supported by numerous experts such as Walter Pitts or Warren McCulloch. It was predicted in the 1840s by Ada Lovelace, at least regarding the technological part of it, describing various foundations of modern programming (Boden, 2016). If as an academic discipline, AI was born in 1956 (Ministry of Economic Affairs and Digital Transformation, 2020), it has been in the last decade where the most favorable conditions have been created for its exponential leap into society. These conditions have been the development of computers with greater power and capacity, the ease of access to data and the amount of data available, and the development of machine learning methods. Being a technology that aims to learn without requiring any human intervention at some point, it is constantly evolving. That is why these AIs are currently capable of seeing (image recognition, which we refer to as “artificial vision”), hearing (voice recognition), and understanding natural language. These are examples of

things that only 10 years ago seemed like something out of science fiction but that today are completely present in many areas of our lives. We are witnessing a technology of seemingly infinite possibilities, whose journey has already begun and many different experts from many different fields are working to advance, meaning that currently, the methodological scope of AI is incredibly broad.

If previously it was present only in certain niches, today AI is a constant in many aspects of our daily life. This is why AI is useful to a multitude of industries in completely different fields and is applied in practically all sectors, surpassing human beings. We are not just referring to disciplines such as computing, finance, or IT, but even in other more creative areas, such as art or writing, it is also a constant today. Anthropologically, creativity has always been understood as a trait of intelligence, and for many centuries it was believed that there was no logic in creative thinking, something widely surpassed not only by AI but even by an area in growing development called Computational Creativity. On many occasions, AI can perform tasks that previously required a human being, such as translating texts, designing logos, creating music, or driving. Something that is already implicit in the very definition of “Artificial Intelligence” is the possibility of learning on their own without the need to be programmed for specific tasks. And AI is inspired by the human brain itself including its neural networks, which work through knowledge nodes that can solve specific problems; when combined, they can conclude the data previously entered. The way these machines learn classifies both their operation and their practical applications.

AI represents a great advance in everything that limits human beings, such as the investigation of terminal diseases and their cure, the reduction of diagnostic errors in many disciplines, and when it is also used for climate efficiency, it also has different uses such as improving the thermal efficiency of buildings or reducing the environmental impact of energy. We must keep in mind that AI should be used to augment and amplify human capabilities, not to replace them, as this is one of its intended objectives. However, like any new technology, AI also brings with it a series of risks and conflicts of which we are only seeing a small part: "The management of big data (Big Data) can give rise to conflicts of privacy, responsibility or interests. For this reason, each government must address the issue by establishing the necessary legal frameworks to promote its development" (Estupiñán, *et al.*, 2021, p. 363). It is therefore necessary

for policies to be implemented to take advantage of the capabilities of AI without it putting at risk the privacy of its users. But some view this technology with even greater fear: "AI has called into question the concept of humanity and its future. Some even doubt whether we will have a future because they predict that AI will surpass human intelligence in all areas. Although some thinkers welcome this, most fear it: what place will there be, they ask, for human dignity and responsibility?" (Boden, 2016).

AI applied to journalism and television

Journalism, like many other disciplines, has also been affected by AI in many different aspects, the most impactful one being that due to data processing and language interpretation, AI has been able to transform data into the news. This is why different terms have already emerged to refer to the interference of this technology in journalism: "There is talk of computational, algorithmic, robot-journalism or augmented journalism" (Marconi, 2020). This is not something new, but it has been more than a decade since the first automated news production was launched "In 2010, The Big Ten Network, a partnership between Fox Networks and the Big Ten Conference, launched an automated sports news production service" (Lohr, 2011), and later, all the other media were automating different sectors of news production, especially those in the economic field "The fact that the Economy section was one of the first to be tested is not a coincidence: it is a specialty in which there is a lot of information online and where data is essential, which makes it easier to use PLN to convert it into text." (Canavilhas, 2022, p.2). "Algorithmic journalism is a semi-automatic natural language generation (NGL) process through which data is selected from databases, often public, from which information texts with a predictable structure are generated" (Gómez-Diago, 2022, p. 31).

However, this is only one of the areas of application of AI to journalism, there are many others such as the recommendation of the news based on our tastes or previous searches, the detection of information trends, or the increase in the technical possibilities of journalism... "it has also been proven that AI can be a complementary tool that improves journalistic processes and allows journalists to highlight their unique and creative skills" (Mondria, 2023, p. 57).

In the era of Smart TVs, AI has also adopted an increasingly relevant role in television, completely altering the way we consume it. It is already a reality to control television with your voice, for television to suggest content based on what we usually consume. The new televisions

with AI also adjust the sound depending on the acoustics where the television is located or the brightness and colors depending on the lighting environment of the room. But AI has not only changed the way television is consumed but also the way it is produced. Audiences can be analyzed using AI algorithms to determine what the audience wants to see, and thus this enables more efficiency in creating quality content for television. Advertising can also be optimized to reach a target audience thanks to data analysis.

In Spain, RTVE, the Spanish public television, carried out an experimental test, within the RTVE innovation project for the regional electoral campaign of the Community of Madrid: “Based on the data provided by official information sources, AI technology is capable of interpreting it and transforming it into news related to electoral information” (Corral, 2021). That is, thanks to the agile delivery of the data referring to the elections on May 4, the AI was able to generate narratives about the event so that all municipalities, regardless of their population, had access to the news about the electoral scrutiny. On this occasion not only was this technology useful to the public television, but also to complete complex narratives such as those of elections. In this way, AI can facilitate the work of professionals by detecting changes in voting trends or specifying the distribution of votes.

Deepfakes paradoxes and uses

The alteration of photographs or videos is not something new, it has practically occurred since the invention of the medium itself. However, the techniques to alter them are now much more advanced than in the past. Even in analog photography, there were manipulations, something that, naturally, passed on to digital photography and its retouching with Photoshop and of course to video editing. In the same way, the invention of fake news, as well as a fake documentary, is something that has also accompanied us for many years, even since the radio broadcast "The War of the Worlds", directed and narrated by Orson Welles in 1938, which caused panic among millions of radio listeners who took the story as real news. In recent years we also have the word mockumentaries to refer to those fictitious documentaries that, using the narrative techniques typical of documentaries about true events, give the feeling that the story being told is true.

The term deep fake emerged as an acronym by uniting “Deep,” coming from how this intelligence learns, through deep learning, and “fake”, which refers to what is false or falsification. With this technology, videos of real characters can be altered using existing videos and learning

algorithms known as RGA (Generative Antagonistic Network). The use of this technology has occurred in cinema on numerous occasions, especially in recent years such as *Rogue One, a Star Wars Story* (Gareth Edwards 2016) in which Princess Leia appears in the face of Carrie Fisher when she was young, or for example in the film *The Irishman* (Martin Scorsese, 2019) where the protagonist trio formed by Robert De Niro, Al Pacino, and Joe Pesci are rejuvenated during most of the footage by playing themselves as young people, instead of the choice of younger actors who look like them.

The porn industry immediately saw a possible opportunity for the use of this technology, using it to make pornographic videos with actors who had never participated in such scenes. Many videos were created of celebrities who, without giving their permission (the majority of these deepfake pornographic videos are not consensual, around 96% (Deeprtrace, 2019)), found themselves involved in all kinds of abuse regarding their image. For the first time, this occurred in 2017, when a Reddit user began posting clips that he had compiled on his computer at home combining images of celebrities with pornographic videos (Roettgers, 2018). Dozens of new clips appeared within days, prompting platforms like Twitter and Reddit to ban AI-based pornography on their platforms, updating pornography rules by adding a ban on "depictions that have been falsified" (Roettgers, 2018). However, the creators of this type of video have other ways of exchanging them and the adult film industry itself shares the idea of the difficulty of stopping this type of falsification. And if this can be a headache in our society, it is even more so when these types of videos are no longer made with celebrities, but with their classmates themselves, as happened in Almendralejo (Extremadura, Spain) this summer (Hedgecoe, 2023). Although there is no clear jurisprudence to combat this type of crime, it can be judged as a crime against honor and self-image, and even, according to certain jurists, it could be a case of "pseudo child pornography." In other cases, advertising is the only previous legal experience, since it is the brands that usually worry about legal problems. For example, in Spain, an advertisement in which the deceased Lola Flores appeared, had the support of her family so there were no legal problems. (Gonzalo, 2021) "Deepfakes are not illegal in themselves, but they can commit crimes that could be pursued through criminal, civil, and administrative means. In Spain, it is a crime due to the objective, not due to the technology." (Gonzalo, 2021). That is to say, the purpose for which it was

created must be considered since it is at this point where the law intervenes, which can regulate the purpose for which this technology is used.

Politics has not been immune to Deepfakes either, something that has been reported for years: "AI systems are capable of generating realistic-sounding synthetic voice recordings of any individual for whom there is a sufficiently large voice training dataset. The same is increasingly true for video. As of writing this, "deep fake" forged audio and video looks and sounds noticeably odd even to untrained individuals. However, at the pace these technologies are making progress, they are likely less than five years away from being able to fool the untrained ear and eye." (Horowitz *et al*, 2018, p. 6). And if computational propaganda was already a problem, what has contributed by entering social networks through anonymous users or even robots to hyperpolarize the political ecosystem, by spreading hoaxes or polarized content, AI is only aggravating it either by prioritizing the content that users expect or by sending personalized propaganda messages. Recently it was also used to launch a video in which Volodymyr Zelensky announced the surrender of Ukraine in the face of the advance of Russian troops. Zelensky appeared in the shot, and with his voice asked that they lay down their weapons (@Shayan86, 2022), in addition, the Ukrainian national news website was hacked to post the video. These hoaxes contribute to misinformation and although the reaction to it was extremely rapid, and prior notice work had been carried out to the population about the possibility of something like this happening. This type of video represents the highest echelon when it comes to sophisticated manipulation for specific purposes.

In television, the uses of this technology are also multiple; on the one hand, they carry an obvious danger that we will talk about later. However, there are other, more beneficial uses of this Intelligence, such as expert dubbing in any language. Uses that the BBC has already experienced with news presenter Matthew Amroliwala speaking different languages (Merino, 2018). The possibility of ending language barriers opens a new power for the media. On the other hand, in Spain, it is often used in the television program *El Intermedio* (La Sexta) with an intended humorous tone.

Deepfakes, real problems

However, everything described above is not the only controversy that deepfakes face. If identity theft and its use in pornographic videos have been a constant in the use (and birth) of deepfakes, where previously you needed a lot of knowledge to carry out something similar and the

result used to be mediocre, currently it is enough to a graphic processing unit and enough photos or video of the person to be replaced, something widely available on the Internet or social networks, so that anyone can make a video of this type, which represents a form of violence against women. In fact, of the total deepfake videos online, it is estimated that 96% are pornographic content while the remaining 4% are other types of content (Deeprace, 2019).

On the other hand, voice cloning has given rise to economic scams, such as the one that occurred in the United Kingdom, a subsidiary of another German energy company, where the CEO of a company asks his subordinates to urgently transfer 220,000 euros to a supplier in Hungary (Sanz, 2019). Something that, although it may seem paradoxical, is nothing more than an example of the use that criminals can give to these technologies. The number of scams of this style to invest in cryptocurrencies has also been known.

In news it is also reaching a tipping point since through AI news presenters are made to report false facts, make problematic value judgments, or even promote products without the presenters having done any of these things. These are also videos that go viral more and get more views sometimes even than the legitimate clips on which they are based. This is why the social network TikTok created a label that warns of realistic content generated by AI, since “AI enables incredible creative opportunities, but can potentially confuse or mislead viewers if they're not aware content was generated or edited with AI” (TikTok, 2023). In addition, it has also announced an algorithm with which it will automatically label publications with AI-generated content, although it has not yet been implemented. It is a worrying aspect regarding the use of AI since spreading false or propaganda messages with the faces of television presenters who are usually seen on news programs is one of the most feasible forms of deception since they are It presuppose a certain credibility.

Naturally, in such a polarized society, political activity can be altered through deepfakes as we have seen. Although now the consequences of these videos in this area seem to have been more limited than in other sectors, it is not difficult to imagine how it is possible that in the future this technology could be used to mobilize the masses or alter a country's elections, for example. Something that is part of information disorders and whose strategy consists of the fabrication of doubt and false controversies to achieve economic or ideological benefits (Del Fresno, 2019), something that should be pursued since “it is foreseeable that the difficulty to differentiate

intentional information disorders from truthful information, supported by a more exact knowledge of cognitive biases, does not stop growing, which is why it becomes more urgent to be able to quickly identify and penalize those responsible for misinformation disorders, defend the consensus between experts and socialize scientific knowledge” (Del Fresno, 2019. P. 9).

Another sensitive point when it comes to AI is intellectual property rights. A hitherto non-existent debate has been generated about the creative capabilities of the algorithm. If AI can create something, who is the author of that creation? Privacy has also been called into question with data protection. Regulatory entities must face all these challenges by updating all the rules on hitherto unregulated topics: “Governments must facilitate the development of AI and economic progress considering that citizens and companies expect that the law Protect your rights, also to the use of intelligent systems. It will be necessary to create synergies between policy and regulation to anticipate possible unknown implications of technology, which may lead to over- or under-regulation” (Estupiñán *et al.*, 2021. p. 367). At this point and given the lack of protection that AI could pose to creative professions, actors and screenwriters in Hollywood have expressed themselves, considering that AI and deepfakes could pose a threat to their profession. The concern is that deepfakes of themselves will be made and that dead actors revived by CGI will become the stars of tomorrow through faces and voices generated by computers or even that the time will come when neither actors nor scriptwriters are necessary for the making of a film (Sanchis, 2023). This is something that the film *The Congress* (Ari Folman, 2013) seemed to prophesy a decade ago, where an actress renounces working as an actress in exchange for money and, in exchange, the production company can generate films starring her, with her digitally scanned body. A prophecy that Bruce Willis seems to have fulfilled, the first actor to sell his image rights to appear in films (Zorrilla, 2022).

With all that has been said, are we facing the end of the saying “seeing is believing”?

References

- Belenguer, L. (2023). Una veintena de madres denuncia la difusión del falso desnudo de sus hijas realizado con inteligencia artificial [Twenty mothers denounce the spread of the fake nude of their daughters made with artificial intelligence]. *20 Minutos*.
<https://acortar.link/GiMFRs>
- Boden, Margaret A. (2016). AI. It's a nature and future. *Turner. Colección.NOEMA*

- Canavilhas, J. (2022). Inteligencia artificial aplicada al periodismo: traducción automática y recomendación de contenidos en el proyecto “A European Perspective” (UER) [Artificial intelligence applied to journalism: automatic translation and content recommendation in the “A European Perspective” (UER) project]. *Revista Latina de Comunicación Social*, 80, 1-13. <https://www.doi.org/10.4185/RLCS-2022-1534>
- Corral, D. (2021). Inteligencia Artificial elecciones en la comunidad de Madrid, tecnología y precisión excepcionales para cubrir una inesperada cita con las urnas [Artificial Intelligence elections in the community of Madrid, exceptional technology and precision to cover an unexpected appointment with the polls]. *RTVE*. <https://acortar.link/g9uQK7>
- Ajder, H., Patrini, P., Cavalli F., & Cullen, F. (2019). *The state of Deepfakes. Landscape, threats, and impact*. Deeptracelabs. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf
- Del-Fresno-García, M. (2019). Desórdenes informativos: sobreexpuestos e infrainformados en la era de la posverdad [Information disorders: overexposed and underinformed in the post-truth era]. *El profesional de la información*, 28(3) e280302. <https://doi.org/10.3145/epi.2019.may.02>
- Estupiñán Ricardo, J., Leyva Vázquez, M., Y., Peñafiel Palacios, A. J., & El Assafiri Ojeda, Y. (2021). Inteligencia artificial y propiedad intelectual [Artificial intelligence and intellectual property]. *Revista Universidad y Sociedad*, 13(S3), 362-368. <https://rus.ucf.edu.cu/index.php/rus/article/view/2490>
- European Commission. (2020). AI watch defining artificial intelligence. *Joint Research Centre (JRC)* <https://acortar.link/PmIIv1>
- Gonzalo, M.(2021). Deepfakes, mentiras y vídeo [Deepfakes, lies and video]. *Neutral*. <https://www.newtral.es/que-son-deepfakes-inteligencia-artificial/20210310/>
- Gómez-Diago, G. (2022). Perspectivas para abordar la inteligencia artificial en la enseñanza de periodismo. Una revisión de experiencias investigadoras y docentes [Perspectives for addressing artificial intelligence in journalism teaching. A review of research and teaching experiences]. *Revista Latina de Comunicación Social*, 80, 29-46. <https://www.doi.org/10.4185/RLCS-2022-1542>
- Hedgecoe, G. (2023). El escándalo en un pequeño pueblo de España por las imágenes de decenas de niñas y jóvenes desnudas generadas por IA [The scandal in a small town in Spain over

- the images of dozens of naked girls and young women generated by AI]. *BBC NEWS MUNDO* <https://www.bbc.com/mundo/articles/cz9r6792k13o>
- Horowitz, Michael C., Allen, C., Saravalle, E., Cho, A., Frederick, K., Scharre, Sch. (2018). National security-related applications of artificial intelligence. *Artificial Intelligence and International Security*. Center for a New American Security's.
- TikTok. (2023). New labels for disclosing AI-generated content. *News.tiktok.com*. <https://newsroom.tiktok.com/en-us/new-labels-for-disclosing-ai-generated-content>
- Lohr, S. (2011). In case you wondered, a real human wrote this column. *The New York Times*. <https://nyti.ms/3tG1RgK>
- Marconi, F. (2020). *Newsmakers: Artificial Intelligence and the future of journalism*. Columbia University Press.
- Merino, M. (2018). *La BBC experimenta con 'deep fakes' para doblar a un presentador en varios idiomas*. Xataca. <https://acortar.link/9EXnio>
- Ministry of Economic Affairs and Digital Transformation (Ministerio de Asuntos Económicos y Transformación Digital). (2020). Estrategia Nacional de Inteligencia Artificial (1.0). *Mineco.gob.es* <https://acortar.link/DEQoZc>
- Mondria Terol, T. (2023). Media innovation: applications of artificial intelligence in journalism in Spain. *Textual & Visual Media*, 17(1), <https://doi.org/10.56418/txt.17.1.2023.3>
- <https://www.cnas.org/publications/reports/artificial-intelligence-and-international-security>
- Roettgers, J. (2018). Porn producers offer to help Hollywood take down deepfake videos. *Variety*. <https://variety.com/2018/digital/news/deepfakes-porn-adult-industry-1202705749/>
- Roettgers, J (2018). Reddit, Twitter ban deepfake celebrity porn videos. *Variety*. <https://variety.com/2018/digital/news/reddit-twitter-deepfake-ban-1202690627/>
- Rouhiainen, L.(2018). *Inteligencia artificial 101 cosas que debes saber hoy sobre nuestro futuro* [Artificial intelligence 101 things you should know today about our future]. Alienta Editorial.
- Sanchis, A. (2023). Hollywood afronta una huelga de 160.000 actores sesenta años después [Hollywood faces a strike of 160,000 actors sixty years later]. *Xataca*. <https://acortar.link/dM9SGZ>

- Sanz Romero, M. (2019). Roban más de 200.000 euros falsificando mediante inteligencia artificial la voz de un CEO [They steal more than 200,000 euros by falsifying the voice of a CEO using artificial intelligence]. *Computer Hoy*. <https://acortar.link/NtIrqG>
- Zorrilla, M. (2022). Bruce Willis pone patas arriba Hollywood: es el primer actor que vende su imagen para ser recreado en películas a través de Inteligencia Artificial [Bruce Willis turns Hollywood upside down: he is the first actor to sell his image to be recreated in films through Artificial Intelligence]. *Espinof*. <https://acortar.link/E8M4JP>
- @Shayan86 (Shayan Sardarizadeh, 2022) President Zelensky has uploaded a video to refute the fake video of him in the tweet above or that he ever called on his troops to lay down their arms. TV24's hacked website still has a screenshot of the fake video along with a transcript of it. X. <https://twitter.com/Shayan86/status/1504106312115888130>