# Vetting and verifying multimodal false information. A challenge for democratic societies*

Sebastián Sánchez-Castillo
Carlos López-Olano
Universitat de València
sebastian.sanchez@uv.es
clolano@uv.es

## 1. Introduction

The rapid proliferation of fake news is a challenge for free societies, founded on freedom of expression and information, endangering their democratic systems through audience confusion. Fake news is a type of disinformation with the ability to alter the logical orientation of readers, and as evidenced in different academic publications, it can radicalize the citizen and favour violence. This news is also created to confuse audiences and reinforce certain trends. This is important enough to produce a scientific interest in information science studies and urge the creation of tools capable of detecting and identifying this type of news, especially those disseminated by social networks, where a personal relationship is maintained with the group. In these close and friendly social spaces, users experience more induced forgetfulness than when interacting with strange social groups, that is, the convergence of memory is more likely to occur within the same group. There is a danger that false news and collective false memories could become the price of defending freedom of expression. The increase in misinformation can alter individual and collective memories in a worrying way. Understanding how and why false memories form could offer some protection the next time a massacre that never took

place was mentioned, such as when President Trump denounced bombings in Sweden in 2017 that never happened.

Disinformation is changing the way of communicating politics, favouring speeches that tend to mislead potential voters or questioning electoral results, as in the Brexit referendum in 2016 and the elections in the United States the same year (Giachanou, Zhang and Rosso, 2020). Detecting fake news is not an easy task due to the large number of technologies and methods employed. The content of the information is essential to detect false news, but it is not enough. The current multimedia informative content is constructed from a multimodal perspective, making approaches capable of combining textual, visual and semantic information to analyze the problem necessary (Song et al., 2020). Sometimes the content of the image and the text are constructed in a contradictory way, something that makes it even more difficult to detect their veracity.

It is necessary to go beyond from the academy in the construction of experimental models of control based on data on regular disinformation or conspiracy theories and evaluate whether the multimodal disinformation constructed by information professionals or that disseminated by non-professional communicators maintains correlation; and also, whether multimodal disclaimers can neutralize the persistence of such falsehoods. There is an academic debate about the depth of the fake news detection task. In the first place, if it is limited only to define whether the news is true or false, but also, as a possible multiclassification, regression or grouping problem.

The technological architecture that supports the current information system, the use of data mining and a news ecosystem based on immediacy are the perfect setting for quick propaganda and powerful falsification, capable of producing very negative consequences both for society as well as the individual. Fake news produces a persuasion aimed at favouring untrue or biased beliefs – in short, speeches without a clear definition between what is credible or what is false, that is, between information or propaganda.

## 2. Advanced detection modes

The low cost and the ability to disseminate news in real time makes social networks the ideal platform for a large part of society to search for and consume news. The rise of fake news on social media can discredit traditional sources of information, which have enjoyed high levels of public trust and credibility and damage the stability and harmony of society (Lazer et al., 2018).

The elimination of journalistic control based on objectivity and professionalism means that the spread of false news can undermine the assumed veracity enjoyed by traditional media, especially audio-visual media. The use of images to build a strategic disinformation model is likely to be based on the premise that images are a direct representation of reality and as such, are perceived as being more credible than textual information. Audiences may be less suspicious of multimodal communication processes, these models being

perceived as more plausible. This makes it necessary to go deeper into academic research, with the intention of showing whether images are instrumental elements in current disinformation, whether the free availability of image editing is favouring the extension of this process, and whether multimodal disinformation is going unnoticed by international verifiers among others *Politifact*, *FactCheck*, or in Spain such as *Newtral* or *Maldita.es*.

One of the methods used to mitigate the dangerous consequences of fake news is the manual verification of publications done by expert journalists, which, although benefitting from high precision, is an unviable process in the face of the huge amount of information that emerges and is generated online in real time (Zhou et al., 2019). This manual text analysis was organized under linguistic keys. However, the high speed at which new messages are generated makes it impossible to discriminate between the authentic and the false, making it necessary to have advanced automated systems to immediately detect these submissions or, at the very least, alert to messages that need a more detailed examination. Systems for detecting false information are based on the precision and speed of retrieving information, systems that must be continuously learned, especially those around social networks. The automated use of complex patterns has proven very useful in alleviating the shortcomings of traditional methods.

The new form of media consumption, based on the immediate, the ephemeral and excess, has been a challenge for scientific research in information science. It tends to identify false textual or visual news, without paying due attention to the general composition of the news in an integration of text and visual content as joint components. Fake news of a multimodal nature has a greater capacity to empirically influence readers through the juxtaposition of images (still or moving) and texts, making it a very effective strategy in the manipulation of ideas. The academic literature has very few in-depth studies on the detection of fake news under a multimodal approach, especially for long-term news. However, there is evidence of important academic publications that influence, from a technical perspective, the analysis of fake news through complex Recurrent Neural Networks (RNN) which combine textual, visual and semantic information with great success in learning the representation of images and texts (Khattar et al., 2019).

Since 2018, advanced automated systems such as the Elaboration Likelihood Model (ELM), the Machine Learning Model, the Multimodal Variational Autoencoder (MVAE), Limited Capacity Model of Mediated Motivated Message Processing (LC4MP) and the Crossmodal Attention Residual and Multichannel convolutional neural Networks (CARMN), have been capable of evaluating information in real time to identify textual and visual characteristics, photographs or videos, most likely to be associated with fake news, but without forgetting that they are systems that learn from probabilistic models (Singh, Ghosh and Sonagara, 2021). These are complex models with clear limitations when faced with a discourse that is difficult to detect, since disinformation can be in the nature of an unnoticed exchange of false information, or a deliberate action of information known to be false.

## 3. Multimodal misinformation and academic research

It is necessary to approach the problem of false news from a multidisciplinary perspective, in which information theories, computer science, statistics and systems engineering converge due to the need to try to stop or understand the scope of this information lacking the principle most basic of journalism – truthfulness. This is what the various authors participating in the monograph "Images for disinformation: multimodal falsehood models" do to shed light on the phenomenon.

In the fight against disinformation, which did not emerge in our time, but which has become viral thanks to social networks and with respect to which there is surely more widespread concern in the academic field, is basic popularization and the promotion of tools that facilitate media literacy, and the possibility of aborting the spread of fake news. And to talk about this phenomenon of the collaborative internet, perfect examples can be found in classic television programs such as *Yes Minister* (BBC, 1980-1984). This was political sitcom that can be useful in our times, many years after its broadcast, to combat disinformation speeches. The ironic dialogue between characters can be a powerful critical tool of political disinformation and something that fits well with the usual politainment on YouTube (Berrocal, Campos-Domínguez y Redondo, 2014).

The recovery of the messages of this political series through a network with the peculiarities of the most extensive video repository, has become viral and has updated its content. In today's society, information is cheap and the responsibility for its dissemination often falls on users, without centralized control. This promotes chain messages, in correspondence with the so-called Network Society. That some of the sequences of this nineteen eighties series reach 400,000 views alerts us to the validity that it still has in this web 2.0 environment (O'Reilly, 2006). Indeed, *Yes Minister* is a style book of political strategy, which strips and shows media manipulation, and which has been offering lessons in disinformation for nearly 40 years. A disinformation that now comes through social networks, which makes it much more effective and dangerous, but which in turn allows the acid discourse of this sitcom to also spread to denounce unethical politicians.

COVID-19 has undoubtedly had an important impact on many aspects of current and future society, and also in the expansion of hoaxes that have found currency in very dangerous health matters (Salaverría et al., 2020). In this panorama, therefore, the approach of the comparative study of the Instagram accounts of four Latin American candidates for the presidency is very pertinent. The omission of the use of the mask in messages with such wide diffusion at such a conflictive health moment can lead to the projection of a potentially very influential disinformative discourse. And the politicians analysed seem more concerned about other issues than about preventive measures related to the pandemic.

Multimodality is important in the analysis: beyond the message that one tries to launch, the image that the candidate projects is of interest here. A

gesture, a decision as seemingly simple as to whether wear the protective mask or not can be viewed from a communicative perspective with broad consequences when we consider its scope in society. Really, the multimodal discourse of candidates about the pandemic was mostly misinformative regarding COVID-19, being considered innocent, or without intention (Buckland, 1991). Even so, it could generate confusion and mistrust.

Misinformation and disinformation are undoubtedly related to the dissemination of racist, xenophobic and islamophobic discourse shared by the far-right or alt-right parties on the rise at a global level that Benkler, Faris and Roberts (2018) detected in what they consider an information epistemic crisis in contemporary democratic societies. In this publication we will see the analysis of fake news with this specific approach detected by the four Spanish information verification media entities accredited by the International Fact-Checking Network: *Maldita.es*, *Newtral*, *Efe Verifica* and *Verificat*.

Unlike what was found in previous studies, which concluded that fake news pieces operated mainly textually (Salaverría et al., 2020), in the study we will see the predominance of visual and audio-visual formats. The addition of texts and images is a basic tool in manipulation mechanisms, which serves to present the otherness as a group and not as individuals. This combination serves as a "fake news personalization", in which the user modifies the visual message enough to take the images out of context, creating multimodal disinformation.

The Arab world and its powerful Al Jazeera television are also an excellent field of work in which to see how large corporations incorporate labour organization dynamics to preserve their prestige in this media landscape dominated by social networks and immediate information. This TV station, which some consider to be the Islamic BBC, faces a future full of unknowns, in a crisis landscape in which some wonder how long the model based on conventional linear television broadcasting will continue (Schlesinger, 2016). It is significant to see how the advances in Artificial Intelligence can be used to create automated fact-checking, which allow the detection of hoaxes in the networks, but also the role that human intervention contributes, the qualitative factor of well-informed and specialized journalists. This verification work is even more important when introducing the variable of multimodality and image diversity.

Al Jazeera +was launched in 2014, as part of the broadcaster's diversification strategy, and offers videos and photos directly through practically all social networks and specific apps. It is visual material that largely comes from the users themselves, but which goes through a complex verification process carried out by a specific department before being disseminated. In this way, the media group offers in its content a prestige that it does not want to lose, and that is vital for its maintenance as an informative reference. The user-generated content becomes a vital source of information transmitted for the rest of the world, but always after the convenient filters that verify its origin, through geolocation, using various applications or platforms, or simply after

direct research the nine editors who work in his checking unit do. The credibility tools that can be used for this work are varied and effective. All these processes have led to crucial changes in the organization of the newsroom, and also the audio-visual material provides a quality bonus that contributes to the culture of participation from TV viewers.

Fake content, especially multimodal content expanded in the so-called Post-Truth Era (Keyes, 2004) also reaches the most recent social networks, such as TikTok. Beyond *challenges* and viral dance moves, we can verify through an international study that it is a platform capable of spreading misinformation, exactly like other networks, something that is evident after the analysis of the 2020 presidential election in the United States. The misinformation reality gains complexity with the hybridization of messages that support various conspiracy theories, which grow when related to the COVID-19 pandemic.

Individual profiles are more prone to the creation of false content, and in addition, the speech can be more effective when it comes from users who are not aware of its falsehood. The platform does not promote profiles but specific content: users build their messages with what the algorithm presents on their feed. The study also shows that regarding the TikTok's algorithm, the false content does not always obtain a better dissemination.

Finally, the so-called deepfake news is the protagonist of the latest proposal for analysis, as the last frontier and the greatest challenge in detection and verification. In this case it comes in the form of a thorough bibliographic review. These digitally manipulated videos, which make things that never happened seem real, have been installed for some time with increasing success and potential for viralization on social networks. Their importance as an object of study is great, since thanks to Artificial Intelligence, various free applications achieve credible results without the need for their creator to be an expert. In other words, they are a good example of user-generated content. The text highlights the dangers of these fake videos, for example in cybersecurity, and the difficulty of combating them using detection tools, since the applications that make deepfakes themselves are capable of learning and avoiding being detected by using Artificial Intelligence themselves. Beyond the jokes, comic effects and advertising strategies that become viral, the ability of these deepfakes to produce confusion and disinformation is predicted as very dangerous in our society in the immediate future.

However, the most damaging aspect of deepfake is probably not the disinformation *per se* but, rather, the lack of confidence in the news, even in audio-visual format, derived from constant contact with disinformation. All these different lines of analysis make up an excellent approach from diverse perspectives and ways of tackle disinformation, contemplating various narratives, techniques, themes or methodological approaches that seek to offer an image and interpretation faithful to this informative fact, that of the expansion of the multimodal fake news, associated with our times.

## Bibliographical references

BENKLER, Y.; FARIS, R. and ROBERTS, H. (2018). *Network propaganda*. New York, NY: Oxford University Press.

BERROCAL, S.; CAMPOS-DOMÍNGUEZ, E. y REDONDO, M. (2014). "Prosumidores mediáticos en la comunicación política: El 'politainment' en YouTube". *Comunicar*, 45, 65-72.
<http://dx.doi.org/10.3916/C43-2014-06>

BUCKLAND, M. K. (1991). "Information as a thing". *Journal of the American Society for Information Science*, 42 (5), 351-360.
<https://doi.org/10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3>

GIACHANOU, A.; ZHANG, G. and ROSSO, P. (2020). "Multimodal Fake News Detection with Textual, Visual and Semantic Information". In *International Conference on Text, Speech, and Dialogue*, 30-38. Springer, Cham.
<https://doi.org/10.1007/978-3-030-58323-1_3>

KEYES, R. (2004). *The Post-truth Era: Dishonesty and deception in contemporary life*. New York: St. Martins Pr.

KHATTAR, D.; GOUD, J. S.; GUPTA, M. and VARMA, V. (2019, May). "Mvae: Multimodal variational autoencoder for fake news detection". *The World Wide Web Conference* 1, 2915-2921.
<https://doi.org/10.1145/3308558.3313552>

LAZER, D. M. J.; BAUM, M. A.; BENKLER, Y.; BERINSKY, A. J.; GREENHILL, K. M.; MENCZER, F. and ZITTRAIN, J. L. (2018). "The science of fake news". *Science*, 359 (6380), 1094-1096.
<https://doi.org/10.1126/science.aao2998>

O'REILLY, T. (2006). "Qué es Web 2.0. Patrones del diseño y modelos del negocio para la siguiente generación del software". *Boletín de la Sociedad de la Información: Tecnología e Innovación*, 3, 177-201.

SALAVERRÍA, R.; BUSLÓN, N.; LÓPEZ-PAN, F.; LEÓN, B.; LÓPEZ-GOÑI, I. and ERVITI, M.C. (2020). "Desinformación en tiempos de pandemia: tipología de los bulos sobre la Covid-19". *El profesional de la información*, 29 (3), e290315.
<https://doi.org/10.3145/epi.2020.may.15>

SCHLESINGER, P. (2016). UK Public Service Media in Transition: Some Brief Reflections. *Gouvernance et service publique mediatique dans les sociétés de la connaissance*. Montreal, Quebec: CRICIS, 94-99. Retrieved from <https://eprints.gla.ac.uk/121055/1/121055.pdf>.

SINGH, V. K.; GHOSH, I. and SONAGARA, D. (2021). Detecting fake news stories via multimodal analysis. *Journal of the Association for Information Science and Technology*, 72 (1), 3-17.
<https://doi.org/10.1002/asi.24359>

SONG, C.; NING, N.; ZHANG, Y. and WU, B. (2021). A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. *Information Processing & Management*, 58 (1).
<https://doi.org/10.1016/j.ipm.2020.102437>

ZHOU, K.; SHU, C.; LI, B. and LAU, J. H. (2019). "Early rumour detection". *Proceedings of the 2019 Conference of the north American chapter of the association for computational linguistics: Human language technologies* (pp. 1614-1623). Minneapolis, Minnesota: Association for Computational Linguistics. Retrieved from <https://www.aclweb.org/anthology/N19-1163>.

# The handling of political disinformation in the TV series *Yes, Minister* (BBC, 1980-1984) and its impact on YouTube

Antonio M. Bañón Hernández
Universidad de Almería. CySOC
amhernan@ual.es

## Abstract

*Yes, Minister* is a series that has been part of the collective imagination of citizens in many English-speaking countries since the 1980s, in which disinformation is frequently used or mentioned by its main characters. Its enormous impact has been long-lasting, and in recent years it has gained special prominence on YouTube. The objectives of this paper are the following: a) to quantify the presence of fragments of the series *Yes, Minister* on YouTube, including their titles, the episodes to which they belong, their duration and the number of views and comments; and b) to analyse the processes, strategies and mechanisms of disinformation in these fragments. To this end, we first described the fragments with more than 200,000 views, of which there were forty. After this analysis, we chose the videos with more than 400,000 views and, in those, analysed the processes, strategies and mechanisms of disinformation. There were twenty-two such documents and they contained as many as 125 samples of disinformation: mostly associated with the process of concealment, followed by blurring and, thirdly, invention. We went on to check for the presence of the nine strategies linked to these processes (abolition, segmentation, deviation, saturation, alteration, divergence, impersonation, incorporation and transformation). Abolition and alteration predominated. Finally, we described the main mechanisms by which these strategies materialised, which included contradiction, confusion, ambiguity, exaggeration, interruption, separation and assignment. We conclude that the publication of the series fragments on the networks indicates public interest in political disinformation. Their use in formal educational contexts, based on analyses such as the one in this paper, is a valuable approach for dealing with discursive processes and mechanisms of disinformation in different areas of knowledge.

**Keywords:** political disinformation; disinformation strategies; TV series; *Yes, Minister*; YouTube

**Resum.** *El tractament de la desinformació política a la sèrie de televisió* Yes, Minister *(BBC, 1980-1984) i el seu impacte a YouTube*

*Yes, Minister* és una sèrie que forma part, des de la dècada dels vuitanta del segle passat, de l'imaginari col·lectiu dels ciutadans de nombrosos països de parla anglesa. La desinformació hi és freqüentment usada o esmentada pels seus personatges principals. El gran impacte d'aquesta sèrie s'ha mantingut al llarg del temps i en els últims anys ha cobrat un protagonisme especial a YouTube. Els objectius del present treball són els següents: *a*) quantificar la presència de fragments de la sèrie *Yes, Minister* a Youtube, incloent-hi el títol, el capítol al qual pertany, la durada i el nombre de reproduccions i de comentaris; i *b*) analitzar els processos, les estratègies i els mecanismes de desinformació en aquests fragments. Hem descrit els fragments de capítols inclosos en aquesta xarxa social que tinguessin més de 200.000 reproduccions. Després d'aquesta anàlisi, hem seleccionat els vídeos que superessin les 400.000 reproduccions i hi hem analitzat els processos, les estratègies i els mecanismes de desinformació. Hi ha vint-i-dos documents d'aquest tipus i incorporen fins a 125 mostres de desinformació associades, majoritàriament, al procés d'ocultació, seguit de la difuminació i, en tercer lloc, la invenció. A continuació, hem comprovat la presència de les nou estratègies vinculades a aquests processos (supressió, segmentació, desviació, saturació, alteració, desarranjament, suplantació, incorporació i transformació). Hi predominen la supressió i l'alteració. Finalment, hem descrit els principals mecanismes en què s'han materialitzat aquestes estratègies. Destaquen els següents: contradicció, confusió, ambigüitat, exageració, interrupció, escissió i cessió. L'exposició a les xarxes de fragments de la sèrie indica un interès social per la desinformació política. El seu ús en contextos educatius formals, a partir d'anàlisis com la inclosa en aquest treball, és una valuosa opció per tractar processos i mecanismes discursius de desinformació en diferents àrees de coneixement.

**Paraules clau:** desinformació política; estratègies de desinformació; sèries de televisió; *Yes, Minister*; YouTube

**Resumen.** *El tratamiento de la desinformación política en la serie de televisión* Yes, Minister *(BBC, 1980-1984) y su impacto en YouTube*

*Yes, Minister* es una serie que forma parte, desde la década de los ochenta del siglo pasado, del imaginario colectivo de los ciudadanos de numerosos países de habla inglesa. En ella, la desinformación es frecuentemente usada o mencionada por sus principales personajes. Su enorme impacto se ha mantenido a lo largo del tiempo y en los últimos años ha cobrado un protagonismo especial en YouTube. Los objetivos del presente trabajo son los siguientes: *a*) cuantificar la presencia de fragmentos de la serie *Yes, Minister* en Youtube, incluyendo el título, el capítulo al que pertenece, la duración y el número de reproducciones y de comentarios; y *b*) analizar los procesos, las estrategias y los mecanismos de desinformación en esos fragmentos. Hemos descrito los fragmentos de capítulos incluidos en esta red social que tuvieran más de 200.000 reproducciones. Tras ese análisis, hemos seleccionado los vídeos que superasen las 400.000 reproducciones y, en ellos, hemos analizado los procesos, las estrategias y los mecanismos de desinformación. Hay veintidós documentos de este tipo e incorporan hasta 125 muestras de desinformación asociadas, en su mayoría, al proceso de ocultación, seguido de la difuminación y, en tercer lugar, la invención. A continuación, hemos comprobado la presencia de las nueve estrategias vinculadas a esos procesos (supresión, segmentación, desviación, saturación, alteración, desacomodo, suplantación, incorporación y transformación). Predominan la supresión y la alteración. Finalmente, hemos descrito los principales mecanismos en los que se han

materializado esas estrategias. Destacan los siguientes: contradicción, confusión, ambigüedad, exageración, interrupción, escisión y cesión. La exposición en las redes de fragmentos de la serie indica un interés social por la desinformación política. Su uso en contextos educativos formales, a partir de análisis como el incluido en este trabajo, es una valiosa opción para tratar procesos y mecanismos discursivos de desinformación en distintas áreas de conocimiento.

**Palabras clave:** desinformación política; estrategias de desinformación; series de televisión; *Yes, Minister*; YouTube

## 1. Theoretical framework

*Disinformation* is the spreading of misleading or false information that leads to personal or collective gain if it is believed to be complete and true by others. If the message received is believed, the recipients may then become purveyors of the same disinformation. This is *how disinformation is propagated*. Deconstructing disinformation consists of decrying this attempt at manipulation and putting forward truthful and not just credible information alternatives. Education is the main deconstructive tool and, in this regard, the value of *media literacy* (literacy in and through the media) must of course be mentioned (Pérez Tornero, Tayie, Tejedor and Pulido, 2018; Aguaded and Romero, 2015). There are three primary *areas* of disinformation. The first is related to the *characteristics of the actors* (identity and responsibility, among others) (Bañón, 2010). The second is associated with the *content of the message* and is the most obvious. The third and final one is related to *contextual factors* (which must be spatio-temporal).

The disinformant is the actant or agent who devises the disinformation and the disinformer is the actor who appears as the person or representative of a group who puts this disinformation into the public domain. The addressees will be the potentially disinformed. The hierarchical relationship between the disinforming and disinformed actors is undoubtedly relevant in the political sphere. As such, we can speak of vertical and lateral disinformation; and, as part of the former, ascending or descending. Citizens trying to understand what policy-makers are doing are unlikely to identify the more implicit messages of disinformation. Indeed, there are many political interactions that remain 'out of the public eye' (Crowder, 2010: 72). Feeling disinformed does not mean having been intentionally led to that state by another interlocutor. Moreover, there are cases where, out of convenience or fear, one prefers to be disinformed or even appears to believe the disinformation.

We can speak of three major *processes* of disinformation, first identified decades ago (e.g., Durandín, 1995). The first is *concealment*. Information is hidden so that others do not learn about a reality which, whether immediately or not, the concealer thinks may be harmful, to a greater or lesser degree, to themself or their own people. The main concealment *strategies* are the following: a) the *removal* of actors, content or contextual factors; b) the *segmentation* of these same elements; and c) the *deviation* of responsibilities or iden-

tities of actors, or of content and spatio-temporal indicators. These strategies also allow us to identify new subtypes of actors, agents or actants that disinform: removers, segmenters and deviators.

The second basic process of disinformation is *blurring*. Information related to the characteristics of the actors, the message or the context reaches its recipient in its entirety, but with associated strategies that make it difficult to identify or interpret. Among these strategies, we consider the following to be the most important: (a) *saturation*, which entails disinformation through the accumulation of actors, information overload or excessive diversification of contextual references; (b) *altering* the proper pace of information, for example by bringing forward or postponing the presence or absence of actors, the presentation of content, or the spaces and times in which an interaction takes place; and (c) *divergence* which consists of the use of a paraphrastic or inaccessible communicative style, useful for blurring responsibilities or identities, content or the spatio-temporal framework itself. Actants, agents and actors can be blurring and therefore saturating, altering or diverging.

The third process is *invention*. In terms of the strategies associated with invention, we believe, based on Riker (1986: 34), Sorlin (2016: 109) and De Santiago-Guervós (2020: 111 and 126-127), that the main ones are a) the *impersonation* of identities and responsibilities, or of contents and spatio-temporal elements; b) the *incorporation* of actors that should not be present, as well as of themes or contextual factors; and c) the *transformation* of characteristics of the actors, the message or the context, falsified in a convincing way. Agents, actants and actors can therefore be impersonators, incorporators or transformers.

Alongside processes and strategies, there needs to be a final relevant discursive level: the verbal, paraverbal and non-verbal *mechanisms* used to express disinformation multimodally. Special attention should be paid to the mechanisms that serve to attenuate or intensify the image of actors, as well as the arguments and their essential parts. This is done by manipulating opinions and evaluations, justifications or demonstrations. Some of these mechanisms will be mentioned and exemplified in section 4.3 of this paper.

Television series help to educate viewers on socially relevant issues (Gil and Gil, 2020). Indeed, they have masterfully captured political contexts of disinformation (*House of Cards*, Andrew Davies and Michael Dobbs, 1990; *The West Wing*, Aaron Sorkin, 1999-2006). They serve as literacy educators on this topic, featuring models and anti-models (del Campo, Puebla and Ivars, 2016: 16-17). However, sometimes humour is also used to express these disinformative strategies in, seemingly, softer tones (*Veep*, Armando Iannuci, 2012-2019). It has been argued that humorous fiction can be a very useful method for examining the behaviour of individuals in political interactions (Considine, 2006: 57) and, in general, for broadening the contexts and intensity of criticism (Condren, 2017).

The literacy benefits of series can nowadays be measured not only by observing the national and international impact of the original broadcast, but

also by looking at their inclusion on video-on-demand platforms such as Netflix, HBO, Amazon Prime, Apple TV or Filmin, as well as their presence on social networks specialising in videos, such as YouTube.

Ruth Wodak has elaborated on the idea of *fictionalisating* politics through the media (2009: 19 and 185). This process inevitably leads to manipulative and disinformative communicative behaviour. Television series set in arenas of political management often illustrate these types of behaviour well. Sandrine Sorlin (2016) demonstrated this with *House of Cards*. Staci Beavers, for her part, highlighted the value of using excerpts from the political fiction series *The West Wing* to illustrate certain concepts related to political communication and to stimulate analysis and discussion of relevant issues in the field by students and teachers (2002: 213). We agree with this literacy function, whose impact in less formal settings increases if the fragments appear on social networks such as YouTube, where textual fragmentation is not only unsurprising, but actually forms part of their communicative identity. The chosen fragment has a galvanising force and can act as a vehicle for knowledge transmission. The same is true of our series, *Yes, Minister*. The rebroadcasting of fragments from this 1980s series on YouTube allows us to reflect, accordingly, not only on transmediation (from television to networks, in this case), a basic process in the shaping of socially shared content (Lacalle, 2011; Berrocal, Campos-Domínguez and Redondo, 2014: 66), but also on the transnational and transgenerational character of public debate about politicians. Similarly, it serves to multiply the number of participants (prosumers, one might say), in what Tomás Albaladejo, referring to political discourse in digital environments, calls the 'extension of polyacroasis' (2012). Ultimately, it involves a revitalisation of the power of ironic dialogue between characters as a tool of criticism in the face of political disinformation. Such irony fits well with *politainment* on YouTube (Berrocal et al., 2014) and with the literacy value of the analysis of anti-models. However, the following remain unclear: on the one hand, "the level of feedback that online citizen debate generates regarding the political class" (Gil-Ramírez, Gómez-De-Travesedo-Rojas and Almansa-Martínez, 2020: 6), and, on the other, how vigorously civic participation is encouraged as a result of the activity generated by the debate on YouTube. In fact, it is well known that there is still a large difference between the number of video views and the number of comments posted for each fragment (Berrocal et al., 2014: 70).

## 2. *Yes, Minister*

The series was created by Antony Jay and Jonathan Lynn. It was recognised not only for its success in terms of viewership, but also for winning several BAFTA awards. It is also a series that is part of Britain's collective imagination (Bonaut and Grandío, 2009: 4). It was even a favourite of Margaret Thatcher (Borins, 2014: 64), who had only come to power a few months earlier, in 1979, at a time in international politics when the *conservative revo-*

*lution* (then also led by Reagan in the USA and Köhl in Germany) was taking place. Carmelo Moreno (2012: 172) says that this new climate set the scene for the appearance of a political comedy series on television, with *Yes, Minister* being "the first political comedy series in history" (2012: 169). *Yes, Minister* went far beyond just being a pioneering series combining humour and political criticism: it was the first to use and mention examples of political disinformation to British viewers and therein lies its socio-educational value. Its uses have the power to teach literacy through exposure to disinformative interactions, while its mentions do so through the direct (and often ironic) reflection of the characters themselves, unsurprising since it was broadcast in more than 85 countries (Valbuena, 2010: 125). The title sequence of the series included three caricatures of the three main characters by the cartoonist Gerald Scarfe, with their very elongated noses. Disinformation is therefore being hinted at from the outset.

Among the series' main qualities, viewers have highlighted both its "visual verisimilitude" and its "authenticity" regarding the discourse of the leading characters (Crowder, 2010). That may well be because it fits well with the mental constructs that these viewers already had and still have about political-administrative activity (Borins, 2014: 63), or because they are persuaded that what is happening resonates well with what could actually be happening. It was revealed early on, by the creators of the series themselves, that there were advisors who had previously been carrying out political tasks around Westminster and Whitehall (Kamm, 2015: 115), which increased the credibility of the dialogue and the plots, regardless of the hyperbole that any comedy series may employ (Granville, 2009: 316).

The series is based on intelligent dialogue and brilliant performances by Paul Eddington (as Minister Jim Hacker), Nigel Hawthorne (as Sir Humphrey Appelby, the Permanent Secretary) and Derek Fowlds (as Bernard Woolley, Hacker's Private Secretary). While Appelby is the master of disinformation, Hacker and Woolley are the keen apprentices. All three, however, have extensive experience in political management and public administration. The plotlines on which the episodes were based were simple. Most of them followed one of three scenarios: a) the minister decides on a measure, the secretary tries to dissuade him by pointing out its catastrophic consequences and, in the end, the minister gives in; b) the minister is in a difficult situation and Humphrey is at first unwilling to intervene, but Hacker finds a way to win his permanent secretary over; and c) the minister and Humphrey work together to try to solve a problem that may affect them both (Botti, 1996: 13-14).

The researchers who have mentioned the series, from a wide range of disciplines, have not been sparing in their assessments of the characters, often using language that reveals the somewhat disinformative profile with which the latter are perceived (see Table 1).

**Table 1.** Language used by researchers to describe the characters

| | |
|---|---|
| James Hacker | "timidity and unwillingness to voice politically unpopular opinions" (Granville, 2009: 325) |
| | "hapless but ambitious minister" and "beleaguered" (Borins, 2014: 63) |
| | "a malleable, spineless minister" (Evans, 2013: 58) |
| | "hapless government minister" and "the pawn of […] Sir Humphrey Appleby" (Smith, 2009: 209) |
| | "relativement honnête, mais benêt, auto-satisfait, incompétent, indécis, toujours en quête de publicité personelle" (Botti, 1996: 15) |
| Humphrey Appelby | "simultaneously wily, obscure and over-clever, but never quite dishonest" (Granville, 2009: 324) |
| | "cynical, self-serving, manipulative civil servant" (Wall, 2008) |
| | "smoothly manipulative high-level", "imperturbable" (Borins, 2014: 63 and 65) |
| | "clever", "recalcitrant" (Teodoro, 2011: 198) |
| | "slick and conniving" (Botti, 1996: 15) |
| | "an effortlessly superior civil service mandarin" (Evans, 2013: 58) |
| | "prétentieux", "condescendant", "arrogant", "anti-européen", "anti-français" and "raciste" (Botti, 1996: 16) |
| | "bureaucratic doublespeak, boilerplate, dissimulation, and nonsense delivered with unflappable aplomb" (Borins, 2014: 64) |
| | the quintessential self-interested bureaucrat" (Considine, 2006: 55) |
| Bernard Woolly | "candide", "réspectueux" / "pédanterie maladive" (Botti, 1996: 17) |
| | "decent, perennially torn" (Borins, 2014: 63) |

Source: Own elaboration.

There are, then, five ways to describe Humphrey: cynical, manipulative, egotistical, arrogant and intelligent. 'Cynical' is defined in the RAE Dictionary as one who acts falsely or shamelessly. Humphrey strives to keeps his efforts to a minimum, an attitude he encourages in others (Considine, 2006: 59). He is the guardian of this classic model whose most obvious immobilism lies in foreign policy, the same as it has been for decades as he himself points out, a statement that has even been adopted by those who have studied the relationship between the United Kingdom and the European Union (Thody, 1997: 5). The terms 'Sir Humphrey-ism' (Granville, 2009: 329) or 'Sir Humphrey Appelby-ization' (van den Berg et al., 2019) reflect his immobile, elitist and manipulative spirit; as does the reference to his character as the epitome of the 'Mandarin scheme' of management: 'impenetrable', 'stagnant' and 'conservative' (Granville, 2009: 325). He handles language very well, combining arrogance and intelligence. Even his seemingly polite phrases can conceal an attack, as the minister himself discovers.

The minister is portrayed as both unhappy and ambitious at the same time. Pressed by events that seem to overtake him, he shies away from public opinion, unwilling to deal with it. Manipulable (especially by Humphrey) and indecisive, if anything his profile leans towards the "disinformable" (tendency to be disinformed) or disinformed participant. It has to be said that, as the series progresses, he gradually learns the basic techniques of disinforma-

tion. Bernard, for his part, is respectful and decent, but with an almost pathological pedantry. Nevertheless, he also understands and sometimes uses these techniques well.

Simon Chesterman (2011) revisits one of the most iconic dialogues from the first episode of the first season of *Yes, Minister*, entitled "Open Government". Bernard, Sir Humphrey and Sir Arnold (Cabinet Secretary) reflect on information transparency. The latter tells Bernard about it being contradictory to argue that it is necessary to both govern and be transparent at the same time. Bernard retorts that citizens of a democratic state have a right to know. Humphrey adds, in line with Arnold: "No, they have a right to be ignorant". The choice is clearly for secrecy and keeping government routines away from public scrutiny. Arnold himself also advocates secrecy, not only regarding the government's stance towards the people, but also that of the civil service towards politicians, as Orange and Turner (2013: 6) recall.

## 3. Objectives and methodology

Below are our main objectives:

1. To understand the *Yes, Minister* series in its historical context and relate it to political disinformation, based on the theoretical framework described previously (see section 2).
2. To delve deeper into the impact of the series on YouTube by identifying the most viewed fragments. Political action is often accompanied by disinformation. Given its characteristics, the Internet can propagate it exponentially and at great speed (Bañón, 2017). *Yes, Minister* has had an extraordinary impact on the social networks (Marrón, 2020) (Section 4.1)
3. To analyse the mentions and uses of processes, strategies and mechanisms of disinformation in a selection of the most viewed videos of the series. The uploading of these videos to this social network demonstrates an in-depth knowledge of the series on the part of those who have edited these fragments and the ability to create themes for the clips by using titles that, occasionally, coincide with those of the full episodes (Sections 4.2 and 4.3).

In order to fulfil the objectives, we used a mixed quantitative and qualitative methodological approach. For the archiving and coding of documents, as well as to obtain quantitative data, we used the text analysis programme *Atlas.ti 8*. We recorded all the types of processes, strategies and mechanisms mentioned in the theoretical framework, as well as the duration of the video, the number of views and also the associated comments. Qualitatively, apart from consulting scientific databases to compile a relevant bibliography on disinformation and the *Yes, Minister* series, as a starting point, we focused on three key areas for the critical analysis of discourse, in general and for the

study of disinformation processes and strategies, in particular: superstructure (related to contextualisation and the representation of relevant actors), macrostructure (regarding thematic-informative dynamics) and microstructure (oriented towards the identification of specific discursive mechanisms) (Bañón, 2017).

The steps followed to carry out our analysis were as follows:

1. Viewing of *Yes, Minister*. This series, like others from previous decades, has greatly increased its current popularity thanks to its availability on Filmin. All three seasons (21 episodes) are available on this platform.
2. Reading the book *The Complete Yes Minister* (Lynn and Jay, 1989) in Kindle format. One of the merits of this volume is that it helped its two creators to make the portrayals and intentions of the characters explicit, whereas in the series they were shown rather implicitly (Valbuena, 2010: 125). In the manner of a diary or notebook, the main characters narrate in detail what happens in each chapter.
3. Search for audiovisual fragments of the original version of the series appearing on YouTube. This search was based on the title, *Yes, Minister*, and the results were filtered, avoiding the inclusion of full episodes, as well as, by mistake, in our sample, clips from the series that succeeded it: *Yes, Prime Minister*. Another selection criterion was that of videos with a total of more than 200,000 views. The search was last updated on 15 December 2020.
4. Quantification, systematisation and interpretation of the data obtained, taking into account the following variables: title, number of views, duration, and number of comments posted for each fragment, plus identification of the episodes to which the fragments belong.
5. Analysis of the mentions and uses of disinformation processes, strategies and mechanisms in the videos on the list with more than 400,000 views. The analysis was preceded by the transcription of the dialogues. This screening of the total number of videos identified was done to help adjust the length of the paper.

## 4. Results

### 4.1. The videos from the series with the greatest impact on YouTube

The information regarding the fragments of the series to be found on YouTube with more than 200,000 views is presented systematically in Table 2 as follows; title (T), season (S) episode (E) to which it belongs, duration of the video (D), number of views of the video (V) and comments associated with each one (C). There are forty in total:

**Table 2.** Detailed description of the videos with the highest impact on YouTube

| Nº | T | D | V | Nº | T | D | V |
|----|---|---|---|----|---|---|---|
|   | S/E |   | C |   | S/E |   | C |
| 1 | Yes Minister explains the EEC (EU) | 03'59" | 3,197,162 | 21 | Explanatory note | 01'46" | 414,207 |
|   | 2/5 |   | 1,561 |   | 2/7 |   | 183 |
| 2 | Why the UK is in the EU | 03'30" | 3,025,963 | 22 | Planes, trains and boats | 01'54" | 407,797 |
|   | 1/5 |   | 3,517 |   | 3/5 |   | 88 |
| 3 | You're a banker | 03'07" | 2,103,301 | 23 | How to discredit a report | 06'48" | 393,615 |
|   | 2/6 |   | 748 |   | 2/4 |   | 334 |
| 4 | Sir Humphrey Appleby on the proper function … | 05'30" | 1,743,491 | 24 | Hushing up a mistake | 03'25" | 372,315 |
|   | 3/6 |   | 879 |   | 1/1 |   | 141 |
| 5 | Sir Humphrey explains Brexit | 01'37" | 1,656,608 | 25 | Pension calculations | 01'57" | 362,303 |
|   | 1/5 |   | 990 |   | 3/1 |   | 76 |
| 6 | The Rhodesia solution | 02'54" | 1,545,742 | 26 | Reshuffle rumours | 02'21" | 347,440 |
|   | 3/6 |   | 592 |   | 2/5 |   | 60 |
| 7 | Earn Your Honors | 03'07" | 1,130,119 | 27 | Mr Haig calling | 02'23" | 315,759 |
|   | 2/2 |   | 396 |   | 3/4 |   | 118 |
| 8 | Sir Humphrey and Jim Hacker discuss art subsidies | 03'29" | 1,033,719 | 28 | The Napoleon Award | 04'15" | 300,398 |
|   | 3/7 |   | 522 |   | 1/5 |   | 131 |
| 9 | Positive discrimination | 03'04" | 982,883 | 29 | Is this highly confidential? | 01'42" | 298,231 |
|   | 3/1 |   | 506 |   | 3/6 |   | 37 |
| 10 | How's the environment? | 03'01" | 934,895 | 30 | A scandal! | 04'18" | 287,795 |
|   | 1/6 |   | 288 |   | 1/7 |   | 110 |
| 11 | English customs | 02'44" | 819,133 | 31 | How to run a hospital | 04'19" | 286,191 |
|   | 3/4 |   | 252 |   | 2/1 |   | 224 |
| 12 | It wasn't me! | 03'00" | 699,391 | 32 | A phone call from the Prime Minister | 03'56" | 264,593 |
|   | 3/3 |   | 162 |   | 1/1 |   | 72 |
| 13 | The empty hospital | 02'01" | 690,647 | 33 | The Civil Service in short | 00'40" | 253,261 |
|   | 2/1 |   | 122 |   | 3/1 |   | 106 |
| 14 | One minister, two ideas | 02'10" | 580,634 | 34 | Jim's TV announcement | 04'02" | 242,089 |
|   | 3/2 |   | 163 |   | 1/4 |   | 125 |
| 15 | Get some patients | 02'10" | 580,351 | 35 | Local Council | 01'29" | 240,449 |
|   | 2/1 |   | 347 |   | 3/2 |   | 31 |
| 16 | The five standard excuses | 01'22" | 570,162 | 36 | Humphrey sets up a dinner | 04'13" | 222,887 |
|   | 2/7 |   | 161 |   | 2/2 |   | 113 |
| 17 | The six diplomatic options | 04'01" | 542,391 | 37 | Jim's going to Europe | 04'05" | 212,243 |
|   | 1/2 |   | 226 |   | 2/5 |   | 122 |

| 18 Government policy | 01'48" | 529,528 | 38 Jim's worst meeting | 04'07" | 208,576 |
|---|---|---|---|---|---|
| 2/7 | | 225 | 1/3 | | 93 |
| 19 Coffee at the university | 01'40" | 465,817 | 39 Getting the better of Humphrey | 3'34" | 207,125 |
| 2/2 | | 164 | 1/4 | | 79 |
| 20 The whisky priest | 04'50" | 452,989 | 40 Redrafting the Redraft of the Redraft | 4'16" | 203,091 |
| 3/6 | | 257 | 1/5 | | 101 |

Source: Own elaboration.

## 4.2. Processes and strategies of disinformation in the most viewed videos

There are twenty-two videos with more than 400,000 views, as can be seen in Table 2. In the following table, we see the presence of processes and strategies related to the three areas pointed out in the theoretical framework: thematic-informative content (T), characteristics of actors (A) and management of contextual factors (F). We also show the number of occasions in which processes and strategies are shown (C), as well as the number of the fragment in which such processes and strategies appear (N):

**Table 3.** Processes and strategies in the most viewed videos

| Concealment | | | | | |
|---|---|---|---|---|---|
| *Strategies* | *Areas* | *C* | *N* | *Subtotal 1* | *Subtotal 2* |
| Abolition | T | 14 | 2, 4, 6, 12, 13, 14, 20 | 23 | 48 |
| | A | 8 | 2, 3, 4, 6, 7, 11, 20 | | |
| | F | 1 | 13 | | |
| Segmentation | T | 1 | 6 | 13 | |
| | A | 12 | 1, 2, 3, 4, 5, 20 | | |
| | F | 0 | | | |
| Deviation | T | 3 | 4, 18, 20 | 12 | |
| | A | 9 | 2, 3, 4, 6, 13, 18 | | |
| | F | 0 | | | |
| Blurring | | | | | |
| *Strategies* | *Areas* | *C* | *N* | *Subtotal 1* | *Subtotal 2* |
| Saturation | T | 3 | 2, 13, 21 | 5 | 43 |
| | A | 2 | 17, 22 | | |
| | F | 0 | | | |
| Alteration | T | 17 | 1, 2, 4, 5, 3, 9, 13, 15, 22 | 20 | |
| | A | 3 | 1, 5, 7 | | |
| | F | 0 | | | |
| Divergence | T | 17 | 1, 2, 4, 5, 7, 8, 10, 18, 20 | 19 | |
| | A | 1 | 8 | | |
| | F | 1 | 6 | | |

| | | | Invention | | |
|---|---|---|---|---|---|
| *Strategies* | *Areas* | *C* | *N* | *Subtotal* 1 | *Subtotal* 2 |
| Impersonation | T | 2 | 1, 5 | 7 | 34 |
| | A | 5 | 1, 2, 3, 5, 14 | | |
| | F | 0 | | | |
| Incorporation | T | 6 | 7, 11, 12, 15, 16 | 19 | |
| | A | 13 | 1, 4, 5, 9, 13, 14 | | |
| | F | 0 | | | |
| Transformation | T | 6 | 7, 11, 12, 15, 16 | 8 | |
| | A | 1 | 10 | | |
| | F | 1 | 3 | | |

Source: Own elaboration.

## 4.3. The mechanisms of disinformation in the most viewed videos

Thematic alteration finds an essential mechanism in the appearance of *contradictions* with respect to the issues addressed. The minister is supposedly a defender of the European ideal, but does not hesitate to criticise the European model whenever his interests are affected by EU initiatives. The permanent secretaries, when they talk about equal opportunities for women and men, say they agree with the idea, but then use justifications and demonstrations to the contrary. At the minister's meeting with the transport representatives, the same thing happens: even though there is seemingly an agreement to bring the meeting to an end, the reality is quite different. With regard to pace as an element of disruption, it is worth noting the cases of *postponement* of decisions or initiatives as a tool to avoid dealing with conflictive issues, for example, by setting up external or internal committees that end up not deciding anything. Several actors try to convince the minister of the need to wait a couple of years to officially open the hospital, which has already been finished for more than a year, to the public. This despite the fact that it already has hundreds of administrative and maintenance workers. These delaying tactics are widely used by Humphrey, sometimes in the form of a simple request for patience when it comes to making decisions. Similarly, the pace of information can also be altered by *anticipating* possible conclusions. Humphrey argues that one should only investigate something if one knows the outcome of the investigation in advance.

Divergence plays an important role in the fragments of *Yes, Minister*. This is not surprising, since we understand that the characters know how to use language to generate mechanisms of *uncertainty, confusion, ambiguity* or *exaggeration*. The Tower of Babel is mentioned as an illustration of daily life in Europe. The minister says that Humphrey lives in an ivory tower. When something is said to be "officially impossible" it means that it is possible, if it falls outside the official boundaries. And of course, long sentences or cryptic administrative language are used by both Humphrey and Bernard to protect or flatter themselves. There are also examples of dressing up information to

make it more acceptable to the public, or answering questions by first analysing the wording of the questions so as to evade them. Exaggeration seems to be the exclusive property of the permanent secretaries, who respond to the minister's initiatives by alluding to the fact that, if what he proposes is done, we would be on the verge of a catastrophe and even the fall of the monarchy.

Abolition is expressed in mechanisms involving the *exclusion* of actors who are intentionally not informed, as happens with Humphrey in certain meetings or with the minister himself, who is not invited to relevant meetings. It also happens when the minister participates in a diplomatic meeting during which a trade agreement is reached with bribes, but he does not know anything about it. He is the only one who is disinformed because he has been excluded from these "negotiations". The same is true when it is stated that the Prime Minister is not to be informed about arms sales to terrorist groups. The actors are also in the process of being excluded when it emerges that the department to which the Minister, Humphrey and Bernard belong is about to be abolished behind their backs. Similarly, abolition is expressed in the confidentiality of alleged sources. As far as the abolition of content is concerned, this takes place, for example, through the flow of information being *interrupted*, as is the case when the minister has not been properly informed of the European intention to launch the Europass (a European ID card), which is, precisely, a key piece of information (there is even talk that it would be "political suicide" for Hacker). Or when a journalist from *The Mail* informs the minister of his suspicions regarding the destruction of part of the documentation he wants access to in order to investigate mismanagement carried out decades ago by Humphrey. Bernard also leaves out relevant information on occasion, when he briefs the minister on sensitive issues (such as the number of civil servants and health staff in a London hospital).

The mechanisms of actor segmentation are the *separation* of identities, the *division* of actors and the *splitting* of functions. Thus, for example, the minister confesses to being both pro-European and anti-Brussels, and believes that his Permanent Secretary acts more like a civil servant than a citizen on moral issues. Humphrey says that ministers are like impulsive toddlers and goes on to argue that if he had listened to the eleven governments that have come and gone in his thirty years of experience, he would have ended up schizophrenic. This idea fits well with his maxim when it comes to contentious issues: "This is not my problem". The segmentation of actors is also shown to be a relatively hidden objective of the UK in its policy on Europe, as – reportedly – it has always wanted to divide the partners in order to gain the upper hand.

The main mechanism of actor deviation is the *assignment* of responsibility. In the Europass dispute, the Foreign Office passes the problem on to the Home Office, which in turn does the same to Hacker's Department of Administrative Affairs. Hacker himself and his two advisers devise the most appropriate way to pass the buck to the Prime Minister's office regarding the information obtained about the sale of arms to terrorists, in such a way as to

prevent this from eventually coming to light. And Hacker himself surrenders responsibility for the welfare of a group of Cuban refugees arriving at the Treasury in London. As for Humphrey, he even goes as far as shifting responsibility for his department's underspending onto the minister himself when he is questioned in a parliamentary committee.

## 5. Discussion and conclusions

The European Union published a *Code of Practice on Disinformation* in 2018. One of its objectives was to "reduce the visibility of disinformation". On occasions, though, it seems to hold true that it is also advantageous to make it more visible, to deconstruct it and thus teach citizens about the processes, strategies and mechanisms on which disinformers rely. Although very important, it is not enough to offer media literacy "focused solely on the verification of discourses", but more ambitious models of "critical thinking" and "humanistic and civic values" must be proposed (Pérez Tornero et al., 2018: 229 and 231).

Since *Yes, Minister*'s original broadcast, many millions of viewers have learned about the main processes, mechanisms and strategies of political disinformation. It has often been considered a manual for political strategy, in which manipulation plays a leading role. In our view, however, it is a brilliant representation of the ways in which disinformation occurs in the political sphere and is then applied to diverse geographical contexts (Considine, 2006: 58).

In total, the forty videos chosen amount to 124 minutes, 56 seconds. The exact number of views is 29,125,291. This clearly illustrates the impact of the series, as well as its influence on the shaping of a shared mental model of political disinformation. Add to this the 14,422 comments: a low figure in proportion to the number of views, as is generally the case on YouTube (Gallardo and Jorge, 2010). All the seasons and all the episodes feature in this collection, with the exception of episode three from the second season. The episodes with the greatest impact, as can be seen, are "The Writing on the Wall" (1/5) and "The Whisky Priest" (3/6), with four fragments each. In third place, with three videos, are the following episodes: "The Devil You Know" (2/5); "Doing the Honours" (2/2), "Equal Opportunities" (3/1), "A Question of Loyalty" (2/7) and "The Compassionate Society" (2/1).

Disinformation strategies are used, mentioned or appear 125 times in our sample. Indeed, this is true for all the videos (except for 19, which shows a direct and unabashed interaction at Baillie College, where the dean and the bursar complain about the minister's projects). That means millions of YouTube users have been made literate in a certain way about this complex communicative macro-process. This is the case regardless of whether they have seen the clips for entertainment purposes or as a way to avoid being part of a politically disinformed society. Concealment is the most frequent process, followed by blurring and, lastly, invention. The most frequent strategies were, in descending order: thematic alteration, thematic divergence, thematic abolition, actor segmentation, actor deviation and actor abolition. Due to

lack of space, we describe only the mechanisms that fit these strategies: contradiction, confusion, ambiguity, exaggeration, interruption, separation and assignment.

The one hundred comments from users of the ten most popular videos confirm that for many of them this is a series that reflects reality (26%) and one of the best in history (18%). Some are written by civil service workers. But, most importantly for us, 9% explicitly state that these fragments have served as material for learning about political theory and practice. On one occasion, one of the users is actually a person who claims to be uneducated and yet demonstrates that they have been able to understand, in depth, the messages about political disinformation conveyed in the series. In general, it is common to read messages that highlight some of the key phrases in the videos, about which brief observations are made that demonstrate a knowledge of the strategies and mechanisms identified in this article: postponement, disturbance, contradiction or exaggeration, among others (30%). Without a shadow of doubt, *Yes, Minister* would serve as an ideal basis for the development of a specific media literacy project on political disinformation.

## Bibliographical references

AGUADED, I. and ROMERO, L.M. (2015). "Mediamorfosis y desinformación en la infoesfera: Alfabetización mediática, digital e informacional ante los cambios de hábitos de consumo informativo". *Education in the Knowledge Society*, 16 (1), 44-57.
<https://doi.org/10.14201/eks20151614457>

ALBALADEJO, T. (2012). "Retórica política y comunicación digital. La ampliación de la poliacroasis". In: E. DEL RÍO, M.ª C. RUIZ and T. ALBALADEJO (eds.). *Retórica y política. Los discursos de la construcción de la construcción de la sociedad*. Logroño: Instituto de Estudios Riojanos, 49-66.

BAÑÓN, A.M. (2010). "El debate político-electoral. Una aproximación desde el análisis del discurso". En M.ª J. MARÍN et al. (eds.). *Discurs polític i identitats (trans) nationals*. València: Universitat de València, 99-131.

— (2017). "Critical Discourse Analysis and New Media". In: K. BEDIJS and Ch. MAAB (eds.). *Manual of Romance Languages in the Media*. Berlin: De Gruyter, 203-244.
<https://doi.org/10.1515/9783110314755-011>

BEAVERS, S. (2002). "The West Wing as a Pedagogical Tool". *Political Science and Politics*, 35 (2), 213-216.
<https://doi.org/10.1017/S1049096502000537>

BERROCAL, S.; CAMPOS-DOMÍNGUEZ, E. and REDONDO, M. (2014). "Prosumidores mediáticos en la comunicación política: El 'politainment' en YouTube". *Comunicar*, 45, 65-72.
<http://dx.doi.org/10.3916/C43-2014-06>

BONAUT, J. and GRANDÍO, M.ª M. (2009). "Los nuevos horizontes de la comedia televisiva en el siglo XXI". *Revista Latina de Comunicación Social*, 12 (64), 753-765.
<https://doi.org/10.4185/RLCS-64-2009-859-753-765>

BORINS, S. (2014). *Governing Fables: Learning from Public Sector Narratives*. Charlotte, NC: Information Age Publishing.

BOTTI, N. (1996). *Yes Minister et Yes Prime Minister. La satire au service de la politique ? Un exemple britanique*. Strasbourg: Université Robert Schuman.

CHESTERMAN, S. (2011). *One Nation under Surveillance: A New Social Contract to Defend Freedom without Sacrifing Liberty*. Oxford: Oxford University Press.

CONDREN, C. (2017). "Yes Minister, Yes, Prime Minister: The Theoretical Dimension". In: MILNER DAVIS, J. (ed.). *Satire and politics. The Interplay of Heritage and Practice*. Cham: Palgrave Macmillan, 227-250.
<https://doi.org/10.1007/978-3-319-56774-7_8>

CONSIDINE, J. (2006). "Yes Minister: Invaluable Material for Teaching the public Choice of Bureaucracy". *Economic Affairs*, 26 (3), 55-61.
<https://doi.org/10.1111/j.1468-0270.2006.00650.x>

CROWDER, M. (2010). "Reading Political Comedy: Yes, Minister and Discursive Contexts". In: N. BOSE and L. GRIEVESON (eds.). *Using Moving Image Archives*, Published by Scope: An Online Journal of Film and Television Studies, 64-81.

DEL CAMPO, E.; PUEBLA, B. and IVARS, B. (2016). "Las series de televisión: 'multiverso' objeto de estudio en comunicación". *Index Comunicación*, 6 (2), 13-19.

DE SANTIAGO-GUERVÓS, J. (2020). "La orientación ideológica en los medios de comunicación social y la eficacia persuasiva de la desinformación". *Discurso & Sociedad*, 14 (1), 107-141.

DURANDÍN, G. (1995). *La información, la desinformación y la realidad*. Madrid: Paidós.

EVANS, E. (2013). *Thatcher and Thatcherism*. London: Routledge.

GALLARDO, J. and JORGE, A. (2010). "La baja interacción del espectador de vídeos en Internet: caso Youtube España". *Revista Latina de Comunicación Social*, 65, 421-435.
<https://doi.org/10.4185/RLCS-65-2010-910-421-435>

GIL, J. and GIL, S. (2020). "Series de ficción como medio de coeducación para la adolescencia. Estudio de caso: Las del Hockey". *Fonseca. Journal of Communication*, 21, 65-86.
<https://doi.org/10.14201/fjc2020216586>

GIL-RAMÍREZ, M.; GÓMEZ-DE-TRAVESEDO-ROJAS and ALMANSA-MARTÍNEZ, A. (2020). "Debate político en YouTube: ¿revitalización o degradación de la deliberación democrática?". *El Profesional de la Información*, 29 (6), e290638.
<https://doi.org/10.3145/epi.2020.nov.38>

GRANVILLE, S. (2009). "Downing Street's Favourite Soap Opera: Evaluating the Impact and Influence of Yes, Minister and Yes, Prime Minister". *Contemporary British History*, 23 (3), 315-336.
<https://doi.org/10.1080/13619460903080135>

KAMM, J. (2015). "Ignorant Master, Capable Servants: The Politics of *Yes Minister* and *Yes Prime Minister*". In: J. KAMM and B. NEUMAN (eds.). *British TV Comedies. Cultural Concepts, Contexts and Controversies*. London: Palgrave Macmillan,114-135.
<https://doi.org/10.1057/9781137552952_8>

LACALLE, CH. (2011). "La ficción interactiva: televisión y Web 2.0". *Ámbitos*, 20, 87-107.
<https://doi.org/10.12795/Ambitos.2011.i20.05>

LYNN, J. and JAY, A. (1989). *The Complete Yes, Minister*. London: BBC Books.

MARRÓN, M. (2020). "La escena de ficción que resume a la perfección la respuesta de los gobiernos a la crisis del coronavirus". *Nius Diario*. Retrieved from <https://www.niusdiario.es/vida/visto-oido/coronavirus-escena-viral-serie-yes-minister-predijo-respuesta-gobierno_18_2914020301.html> [Consulted: 15/12/20]

MORENO, C. (2012). "Humor, política y series de entretenimiento. El fenómeno 'Yes, Minister'". In: I. MARTÍNEZ and C. MORENO (eds.). *De Anatomía de Grey a The Wire. La realidad de la ficción televisiva*. Madrid: La Catarata, 68-192.

ORANGE, R. and TURNER, B. (2013). "Introduction". In: B. TURNER and R. ORANGE (eds.). *Specialist Journalism*. London/New York: Routledge, 1-10.

PÉREZ TORNERO, J.M.; TAYIE, S. S.; TEJEDOR, S. and PULIDO, C. (2018). "¿Cómo afrontar las noticias falseadas mediante la alfabetización periodística? Estado de la cuestión". *Doxa Comunicación*, 26, 211-235.
<https://doi.org/10.31921/doxacom.n26a10>

RIKER, W.H. (1986). *The art of political manipulation*. New Haven/London: Yale University Press.

SMITH, J. (2009). *The Presidents We Imagine: Two Centuries of White House Fictions on the Page, on the Stage, Onscreen, and Online*. Madison: University of Wisconsin Press.

SORLIN, S. (2016). *Language and Manipulation in House of Cards. A Pragma-Stylistic Perspective*. Basingstoke: Palgrave/Macmillan.

THODY, P. (1997). *An Historical Introduction to the European Union*. London: Routledge.

TEODORO, M. (2011). *Bureaucratic Ambition: Careers, Motives, and the Innovative Administrator*. Baltimore: Johns Hopkins University Press.

VALBUENA, F. (2010). "El humor en la Comunicación Política". *Cuadernos de Información y Comunicación*, 15, 123-164.

VAN DEN BERG, C.; HOWLETT, M.; MIGONE, A.; HOWARD, M.; PEMER, F. and GUNTER, H. (2019). *Policy Consultancy in Comparative Perspective: Patterns Nuances and Implications of the Contractor State*. Cambridge: Cambridge University Press.

WALL, S. (2008). *A Stranger in Europe: Britain and the EU from Thatcher to Blair*. Oxford: Oxford University Press.

WODAK, R. (2009). *The Discourse of Politics in Action: Politics as Usual*. Basingstoke: Palgrave/Macmillan.

# Misleading Discourse on Instagram: A Multimodal Study of Latin American Presidential Candidates in the Face of COVID-19

Marga Cabrera-Méndez

Universitat Politècnica de València. Department of Audio-Visual Communication, Documentation, and Art History

mcabrera@upv.es

Rodrigo Cisternas Osorio

Universidad Casa Grande. School of Communications

rcisternas@casagrande.edu.ec

Alberto J. López-Navarrete
Rebeca Díez-Somavilla

Universitat Politècnica de València. Department of Audio-Visual Communication, Documentation, and Art History

allona2@upv.es

rdiez@har.upv.es

## Abstract

Instagram as a multimodal information network has helped politicians to position both their brand and their campaign. We analyzed whether the images and texts published during the pandemic contained misinformation. We studied from a multimodal perspective the Instagram accounts of the presidential candidates of four Latin American countries which held elections in 2021 to identify how much of the discourse was related to controlling the pandemic. The discourse was analyzed using different taxonomies. In the correlation between the discourse and following the recommendations of the World Health Organization (WHO), Chile stood out with the highest level of pandemic compliance; Peru and Ecuador were placed in the middle, while Honduras showed little if any interest. The conclusion was that politicians focused primarily on their campaigns

and marginally on the pandemic. The omission of COVID-19 from most publications reflected a misinformative discourse which could potentially confuse the public.

**Keywords:** COVID-19; Instagram; misinformation; multimodal; politics

**Resum.** *El discurs amb informació errònia a Instagram. Estudi multimodal dels candidats presidencials a l'Amèrica Llatina davant la COVID-19*

Instagram com a xarxa multimodal ajuda els polítics a posicionar la seva marca i campanya. Analitzem si existeix informació errònia en les imatges i textos publicats en plena pandèmia. S'estudien les publicacions a Instagram dels candidats presidencials de 4 països llatinoamericans amb eleccions el 2021 i es pretén identificar quant del discurs està relacionat amb la prevenció davant la pandèmia des d'una perspectiva multimodal. S'analitza el discurs a través de diferents taxonomies. En la correlació entre el discurs i el seguiment de les recomanacions de l'OMS destaca Xile amb el seguiment més gran, el Perú i l'Equador queden en un terme mitjà, i Hondures no desprèn cap interès. Es conclou que els polítics se centren en la seva campanya i molt poc en la pandèmia. L'omissió de la COVID-19 en la majoria de les publicacions projecta un discurs amb informació errònia que pot confondre la ciutadania.

**Paraules clau:** COVID-19; Instagram; informació errònia; multimodal; política

**Resumen.** *El discurso con información errónea en Instagram. Estudio multimodal de los candidatos presidenciales en Latinoamérica frente a la COVID-19*

Instagram como red multimodal ayuda a los políticos a posicionar su marca y campaña. Analizamos si existe información errónea en las imágenes y textos publicados en plena pandemia. Se estudian las publicaciones en Instagram de los candidatos presidenciales de 4 países latinoamericanos con elecciones en 2021 y se pretende identificar cuánto del discurso está relacionado con la prevención frente a la pandemia desde una perspectiva multimodal. Se analiza el discurso a través de diferentes taxonomías. En la correlación entre el discurso y el seguimiento de las recomendaciones de la OMS destaca Chile con mayor seguimiento, Perú y Ecuador quedan en un término medio, y Honduras desprende interés nulo. Se concluye que los políticos se centran en su campaña y muy poco en la pandemia. La omisión de la COVID-19 en la mayoría de las publicaciones proyecta un discurso con información errónea que puede confundir a la ciudadanía.

**Palabras clave:** COVID-19; Instagram; información errónea; multimodal; política

## 1. Introduction

The recent health crisis favored the use of a presidential style supported by social networks to communicate decisions, gauge the environment, and validate policies (Manfredi-Sánchez, Amado-Suárez and Waisbord, 2021), but what about presidential candidates in particular? What was their stance in the midst of campaigning? Social networks lend themselves to populist practices (Moffitt and Tormey, 2014; Bevelander and Wodak, 2019), although pandemic-related information was not always honest and consistent on the Instagram accounts of candidates who sometimes posted inaccurate news reports or items tailored to a personal goal.

Instagram was created as a multimodal information network, i.e., audiovisual and textual publications. Due to its popularity, public figures, including politicians, resorted to Instagram as a communication tool. We considered it imperative to analyze the symbolic elites of this multimodal channel where they were able to present their discourse, often with a misinformative result (Van Dijk, 2006).

Other research in political communication has also focused on Instagram (Sampietro and Sánchez-Castillo, 2020) and there have been numerous studies on the personal content of political candidates on this network, in both electoral and post-electoral periods (Lalancette and Raynauld, 2017; Larsson, 2017; Muñoz and Towner, 2017; Quevedo-Redondo and Portalés-Oliva, 2017). Campaigns based on social networks resulted from new thoughts on political communication and the management of the personal and public images of politicians in digital environments (Lees-Marshment, 2001; Paré and Berger, 2008; Marland, 2012). Instagram constituted one of the best platforms for a communicative model based on the denotative power of photographs, on the emergence of a conversation type based on portraits, selfies, and videos (Quevedo-Redondo and Portalés-Oliva, 2017), and on the combination of photography and politics (Ruiz-del-Olmo and Bustos-Díaz, 2016; López-Rabadán, López-Meri and Doménech-Fabregat, 2016).

## 1.1. Instagram and COVID-19 in Latin America

This research focused on misinformation in the discourse of political candidates on Instagram regarding COVID-19. A multimodal analysis was conducted of the graphic and written discourse of the main presidential candidates in a collection of Latin American countries which held presidential elections in 2021, Chile, Ecuador, Honduras, and Peru. The number of Instagram users was 7.3M in Chile, 5.2M in Ecuador where it was the fourth most used social network, 5.16M in Peru, and 0.8M in Honduras (Librero, 2020), Instagram being one of the four most-used networks in each of the four countries analyzed.

Politicians were able to use the content of their networks to promote public participation in the prevention of and the fight against the pandemic (Niknam et al., 2020; Castillo-Esparcia, Fernández-Souto and Puentes-Rivera, 2020). Research, such as a study of the use of Twitter as a presidential communication channel in the initial period of COVID-19 (Manfredi-Sánchez et al., 2021), indicated Presidents Sánchez and Bolsonaro deployed a health management communication strategy, while López Obrador and Fernández paid little attention to health policy. A study of the Instagram account of the former vice president of Ecuador, Otto Sonnenholzner, during the first emergency phase resulting from the pandemic, showed this politician used his account mainly to promote his personal image rather than having objectives linked to the national health emergency, which echoed political promotion (Lozano-Recalde, 2021).

In 2020 in Latin America, Peru had the highest number of deaths due to COVID-19 across the entire population at 118.6 per 100,000 inhabitants, while in Chile the figure was 89.8, in Ecuador 82.8, and in Honduras 38.0 (Pasquali, 2021).

From the initial declaration of the pandemic, the World Health Organization (WHO) continued issuing updated information on the disease; however, there were indications and recommendations consistent over time that were assumed by governments, national and international organizations, and the media, e.g., wash and disinfect hands frequently, cover your mouth when coughing, refrain from touching your nose or mouth, ventilate rooms well, limit stays in closed and poorly ventilated spaces, maintain respiratory hygiene, and avoid direct physical contact with others. The two measures of most interest in this study were social distancing and the use of masks, as they could both be analyzed by means of images of the candidates.

## 1.2. Misinformation in Multimodal Discourse

We characterized misinformation as ambiguous or vague information that can generate confusion and mistrust in the receiver, can make it difficult to use the information, and can cast doubts in receivers on processing actions and decision-making (Karlova and Fisher, 2013). Discussion surrounding the terms disinformation and misinformation suffered from imprecision in their definitions (Karlova and Lee, 2011). Sometimes the terms were used interchangeably as though there was no distinction between them (Fox, 1983; Losee, 1997), or one was used as a variation of the other (Zhou and Zhang, 2007).

Buckland (1991) defined misinformation as accidental or innocent, while disinformation can reveal the malicious intent of the speaker, if the receiver was aware of being misled. According to Bednar and Welch (2008) and Stahl (2006), the difference between disinformation and misinformation depended on the intention of the speaker. Misinformation occurred when the information provided was not correct but the intention of the deception was benevolent, while disinformation presupposes a malevolent intention on the part of the issuer. However, the intentionality of the speaker may be ambiguous and/or unknown, and it was not therefore satisfactory to differentiate the two terms solely on the basis of this criterion (Karlova and Lee, 2011).

Authors such as Losee (1997) and Zhou et al. (2004) defined one of the aspects of misinformation as incomplete, distorted, or ambivalent information. In addition, Zhou and Zhang (2007) added various types of misinformation, such as concealment, ambivalence, distortion, and falsification. However, these authors did not conduct any disambiguation between the two terms misinformation and disinformation. However, Fallis (2009) concluded that misinformation may or may not be inaccurate, but must be misleading, further adding that the misleadingness of misinformation may be due to the context of the situation. This approach supported the constructivist and subjective view of information of Hjørland (2007). Thus, disinformation is not a proper subset

of inaccurate information (which would be misinformation). For their part, Karlova and Lee (2011) added that misinformation could be inaccurate, uncertain (e.g., by presenting more than one possibility or choice), vague (unclear), or ambiguous (open to multiple interpretations). Information that is incomplete can also be a form of deception, often referred to as misinformation.

As discussed above, misinformation is a multifaceted concept, more complex than simply incomplete information (Karlova and Lee, 2011). Karlova and Fisher (2013) attributed much of the problem of misinformation to its ability to create confusion and mistrust among recipients, making it difficult to use the information.

The analyzed discourse was approached from a multimodal perspective, seeking to detect a process of misinformation in the graphic and textual discourse of the candidates on Instagram. Multimodal Discourse Analysis (MDA) is based on the systemic-functional linguistics of Halliday (1978). This approach understood language as being embedded in social semiotics (van Leeuwen, 2005), where meaning systems are shaped by their contextual inscription.

The multimodal perspective of discourse analysis asserts that communication is inherently multimodal. MDA analyzes the multiple modalities (language, image, or audio) that combine to create meaning in different contexts (O'Halloran et al., 2014). Various modalities beyond language come into play when creating discourse (O'Halloran, 2012). These modalities are also called semiotic resources (Parodi, 2010). Each semiotic resource is understood as a system of options that interact with each other, creating a discursive semantic unit (Menéndez, 2012).

The objective of this research was to study the multimodal narrative of candidates and to determine whether it carried implicit misinformation in the coherence of the coronavirus prevention discourse with respect to the images and text projected by candidates on their Instagram accounts. Analyzing the content of Instagram posts can help identify the thoughts and feelings of the candidates in the face of the health crisis (Niknam et al, 2020).

## 2. Materials and Methods

This descriptive research delved into content analysis based on the monitoring of Instagram posts of the candidates most likely to be elected in four Latin American countries where presidential elections were held in 2021, Peru, Chile, Ecuador, and Honduras. The unit of analysis comprised images uploaded to Instagram along with accompanying captions. Stories, videos, or different carousels of images that accompany a publication were not included, only the first image that appeared on the post being taken into account.

The hypothesis put forward was that there was no consistency in the coronavirus prevention discourse with respect to the images and captions projected by the presidential candidates on their Instagram accounts.

We analyzed the candidates most likely to be elected according to polls, the parties to which they belonged, their Instagram accounts, and the num-

ber of followers starting from the date the pandemic was declared, March 11, 2020, through to December 31, 2020. Table 1 below summarizes the relevant candidate information.

**Table 1.** Candidates Analyzed

| Date and Source | Candidate | Party | Instagram | Followers |
|---|---|---|---|---|
| | | Ecuador | | |
| Election Date: 7/2/2021 | Andrés Arauz | Centro Democrático/ Movimiento Revolución Ciudadana (Left) | @ecuarauz | 15,900 |
| https://www.celag.org/ encuesta-ecuador- diciembre-2020 | Guillermo Lasso | Partido Adelante Ecuador Adelante (Right) | @guillermolasso | 117,000 |
| | Yaku Pérez | Partido Pachakutik (Left) | @yakuperezg | 56,100 |
| | | Chile | | |
| Election Date: 21/11/2021 | Daniel Jadue | Partido Comunista (Left) | @danieljaduejadue | 191,000 |
| https://www.cadem.cl/ encuestas/especial- electoral | Pamela Jiles | Partido Humanista (Left) | @pamelajilesdiputada | 529,000 |
| | Joaquin Lavín | Unión Demócrata Independiente (Right) | @lavinjoaquin | 122,000 |
| | | Honduras | | |
| Election Date: 28/11/2021 | Nasry Asfura | Partido Nacional (Right) | @papialaorden | 3,554 |
| http://www.cne.hn | Xiomara Castro | Libertad y Refundación (Left) | @XiomaraCastroZ | 611 |
| | Yani Rosenthal | Partido Liberal (Center) | @yanirosenthal | 1,327 |
| | | Peru | | |
| Election Date: 11/04/2021 | George Forsyth | Restauración Nacional (Center) | @george.forsyth | 309,000 |
| https://www.celag.org/ encuesta-peru- noviembre-2020 | Verónika Mendoza | Juntos por el Perú (Left) | @veromendoza_peru | 16,900 |
| | Hernando de Soto | Partido Avanza País (Right) | @hernandodsoto | 5,706 |

Source: Own elaboration.

An analysis sheet was constructed as an instrument for data collection on which all posts were recorded for each country and candidate from the date of the pandemic declaration until December 31, 2020. A total of 3,099 posts were analyzed including their images and the accompanying text.

The sheet identified four factors where the first refers to the relevance of the candidate on the Instagram social network, while the other three refer to the prevention of COVID-19 based on WHO recommendations:

1) Candidate Relevance
Data was collected on the age of the account, the number of followers, and the start of activity by the candidate. A count was made of the total number of posts, the total number of posts during the period analyzed, and the number of posts in which the candidate themselves appear.

2) Prevention: Mask Use
Analysis of the total number of posts where the candidate appeared in the period analyzed, counting the number of times they appeared alone and if they wore a mask, whether or not accompanied.

3) Prevention: Social Distancing
When the candidate appeared accompanied, we evaluated whether they wore a mask and if they complied with the recommendations on social distancing issued by the WHO. The analysis took into account whether the group comprised relatives (cohabitants) of the candidate and if it was in an open or closed space.

4) Political Discourse in the Face of COVID-19
Text and images were not combined simply by the addition or intersection of the component meanings. Bateman (2014) introduced the term meaning multiplication to refer to the creation of new meaning through the integration of images and text. Marsh and Domas (2003) created a taxonomy of text-image relationships to analyze how images and text interact. The authors developed 49 image-text relationships with the intention of developing a broad taxonomy applicable to all areas and types of documents, thereby generating a common language for the classification of image-text intent. We used the taxonomy developed by Kruk et al. (2019), based on the work of Bateman (2014) and Marsh and Domas (2003), to categorize the image-text meaning of Instagram posts. Kruk et al. (2019) analyzed a set of 1,299 Instagram posts to validate their taxonomy. The authors created three major taxonomies to categorize the sum of images and text on Instagram and emphasized the need for further research to expand and enrich the proposed classification. One of the taxonomies focused on speaker intent, while the other two (contextual and semiotic) captured different aspects of the relationship between image and caption.

A. Taxonomy of Speaker Intent:
The proposal of Kruk et al. (2019) on speaker intention was based on the illocutionary acts of Austin (1962) that refer to the intention of the message and the meaning of all the communicative elements involved. In other words, they refer to the final objective of the speaker when using the various modes of communication. After analyzing and grouping representative Instagram content, they created the following 8 categories designed to categorize the intention of the author on Instagram:

1. Advocative: Defend an idea, a movement.
2. Promotive: Promote events, products, organizations.
3. Exhibitionist: Create a self-image reflecting the person, status, etc.
4. Expressive: Express emotion, adhesion, or admiration towards an external entity or group.
5. Informative: Convey information about a topic or event using factual language.
6. Entertainment: Entertain through humor, art, memes.
7. Provocative/discriminatory: Direct attack on a subject or group.
8. Provocative/controversial: Shock.

B. Contextual Taxonomy

For the contextual taxonomy, the relationship between the literal meaning of the image and the text was classified. The 49 image-text relationships identified by Marsh and Domas (2003) were grouped into three superordinate categories, which distinguished images as minimally related to text, highly related to text, or related but going beyond text. According to Kruk et al. (2019), this classification frames the image only as subordinate to the text. The authors adapted the three major contextual relationships to make them symmetrical on Instagram:

1. Minimal relationship: The relationship between image and text is literal. The meaning between the two overlaps very little.
2. Close relationship: The literal meaning of the text and the image overlap considerably.
3. Transcendent relationship: The literal meaning of one of the modalities gathers and expands the meaning of the other.

Regarding this taxonomy, it is important to note that the categories "minimal" and "close" were placed on a continuous scale indicating a semantic overlap, while the category "transcendent" indicated an expansion of meaning which cannot be captured on a continuous scale.

C. Semiotic Taxonomy

Semiotic taxonomy captures the relationship between what is signified by its respective modality and its semiotics. The authors started with the distinctions made by Kloepfer (1977), modeled by Bateman (2014), and the parallel and non-parallel distinctions of Zhang, Hwa and Kovashka (2018), and composed three categories:

1. Parallel relationship: When image and text independently embody the same meaning. Zhang et al. (2018) used the term parallel when image and text are intended to convey the same message. For example, if the meaning of the image can be understood individually and the meaning of the text can be understood individually, the relationship is parallel.
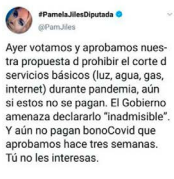
2.  Divergent relationship: When the image and the semiotic text move in different directions, creating a gap between the meanings suggested by image and text.
3.  Additive relationship: When image and semiotic text extend or modify each other.

Regarding this taxonomy, it is important to clarify that the semiotic classification is not always homologous with the contextual classification.

## 3. Results

The taxonomies described above made up one of the factors analyzed, "Political discourse: COVID-19 situation". The other factors were comprised of those referring to COVID-19 prevention according to the standards dictated by WHO (use of a mask, and social distancing), except the first one which analyzed the relevance of the candidates on Instagram. These factors were developed ad hoc in order to integrate them with the taxonomies of Kruk et al. (2019) (figure 1) and to incorporate them (Cárcamo-Morales, 2018) into the misinformation study. On presenting the results, we followed Bateman (2014) and employed tables (<http://bit.ly/3r3YCNW>), which were presented as a key technique when performing compositional cohesion analysis (Cárcamo-Morales, 2018), in this case of image and text modalities. Tables 2 through 5 below present the data analyzed for each country.

**Figure 1.** Examples of the taxonomies analyzed



Source: Own elaboration.

**Table 2.** Data Analyzed - Ecuador

| | | ECUADOR | | | | | |
|---|---|---|---|---|---|---|---|
| **Factor 1:** Relevance - Social Network Presence | **Name** | **Andres Arauz** | | **Guillermo Lasso** | | **Yaku Perez** | |
| | Number of followers | 21,000 | | 121,000 | | 60,500 | |
| | Date account established | 18/4/2015 | | 24/5/2012 | | 16/1/2019 | |
| | Total number of posts | 175 | | 1832 | | 1083 | |
| | Number of posts in the COVID-19 period | 124 | | 743 | | 339 | |
| | Percentage of total posts mentioning COVID-19 | 0.6% | | 7.6% | | 7.0% | |
| | Percentage of posts in the COVID-19 period mentioning COVID-19 | 0.8% | | 18.7% | | 22.4% | |
| | Number of posts/videos with candidate appearance | 114 | 91.9% | 454 | 61.1% | 212 | 62.5% |
| | Number of posts/videos without candidate appearance (text, advert, etc) | 10 | 8.1% | 289 | 38.9% | 127 | 37.5% |
| **Factor 2:** Prevention - Mask Use | Candidate with mask | 69 | 60.5% | 117 | 25.8% | 97 | 45.8% |
| | Candidate without mask | 43 | 37.7% | 327 | 72.0% | 113 | 53.3% |
| | Candidate use of mask unidentified | 2 | 1.8% | 10 | 2.2% | 2 | 0.9% |
| **Factor 3:** Prevention Social Distancing | Candidate alone | 14 | 12.3% | 154 | 33.9% | 58 | 27.4% |
| | Candidate accompanied by others | 100 | 87.7% | 300 | 66.1% | 154 | 72.6% |
| **Factor 4:** Political Discourse: COVID-19 Situation | Number of captions mentioning COVID-19 | 1 | | 139 | | 76 | |
| | Number of captions with official COVID-19 information | | | 1 | 0.7% | 5 | 6.6% |
| | Number of captions with unofficial COVID-19 information | 1 | 100.0% | 138 | 99.3% | 71 | 93.4% |
| | Number of hashtags related with COVID-19 | | | 122 | | | |

Source: Own elaboration.

**Table 3.** Data Analyzed - Chile

| | | CHILE | | | | | |
|---|---|---|---|---|---|---|---|
| Factor 1: Relevance - Social Network Presence | Name | Daniel Jadue | | Joaquín Lavín | | Pamela Jiles | |
| | Number of followers | 194,000 | | 124,000 | | 529,000 | |
| | Date account established | 4/12/2017 | | 20/3/2014 | | 2/6/2018 | |
| | Total number of posts | 554 | | 1,677 | | 1,601 | |
| | Number of posts in the COVID-19 period | 256 | | 433 | | 743 | |
| | Percentage of total posts mentioning COVID-19 | 5.4% | | 8.6% | | 3.3% | |
| | Percentage of posts in the COVID-19 period mentioning COVID-19 | 11.7% | | 33.3% | | 7.1% | |
| | Number of posts/videos with candidate appearance | 89 | 34.8% | 236 | 54.5% | 185 | 24.9% |
| | Number of posts/videos without candidate appearance (text, advert, etc) | 167 | 65.2% | 197 | 45.5% | 558 | 75.1% |
| Factor 2: Prevention - Mask Use | Candidate with mask | 18 | 20.2% | 149 | 63.1% | 120 | 64.9% |
| | Candidate without mask | 70 | 78.7% | 87 | 36.9% | 64 | 34.6% |
| | Candidate use of mask unidentified | 1 | 1.1% | | | 1 | 0.5% |
| Factor 3: Prevention Social Distancing | Candidate alone | 64 | 71.9% | 81 | 34.3% | 54 | 29.2% |
| | Candidate accompanied by others | 25 | 28.1% | 155 | 65.7% | 131 | 70.8% |
| Factor 4: Political Discourse: COVID-19 Situation | Number of captions mentioning COVID-19 | 30 | | 144 | | 53 | |
| | Number of captions with official COVID-19 information | 16 | 53.3% | 58 | 40.3% | 5 | 9.4% |
| | Number of captions with unofficial COVID-19 information | 14 | 46.7% | 86 | 59.7% | 48 | 90.6% |
| | Number of hashtags related with COVID-19 | 4 | | 7 | | 9 | |

Source: Own elaboration.

**Table 4.** Data Analyzed - Honduras

| | | HONDURAS | | | | | |
|---|---|---|---|---|---|---|---|
| Factor 1: Relevance - Social Network Presence | Name | Nasry Asfura | | Xiomara Castro | | Yani Rosenthal | |
| | Number of followers | 4,083 | | 708 | | 1,610 | |
| | Date account established | 16/7/2020 | | 28/10/2020 | | 7/10/2020 | |
| | Total number of posts | 148 | | 10 | | 40 | |
| | Number of posts in the COVID-19 period | 136 | | 10 | | 30 | |
| | Percentage of total posts mentioning COVID-19 | 1.4% | | 0.0% | | 0.0% | |
| | Percentage of posts in the COVID-19 period mentioning COVID-19 | 1.5% | | 0.0% | | 0.0% | |
| | Number of posts/videos with candidate appearance | 66 | 48.5% | 10 | 100.0% | 14 | 46.7% |
| | Number of posts/videos without candidate appearance (text, advert, etc) | 70 | 51.5% | | | 16 | 53.3% |
| Factor 2: Prevention - Mask Use | Candidate with mask | 19 | 28.8% | 7 | 70.0% | 4 | 28.6% |
| | Candidate without mask | 47 | 71.2% | 2 | 20.0% | 10 | 71.4% |
| | Candidate use of mask unidentified | | | 1 | 10.0% | | |
| Factor 3: Prevention Social Distancing | Candidate alone | 21 | 31.8% | 5 | 50.0% | 7 | 50.0% |
| | Candidate accompanied by others | 45 | 68.2% | 5 | 50.0% | 7 | 50.0% |
| Factor 4: Political Discourse: COVID-19 Situation | Number of captions mentioning COVID-19 | 2 | | | | | |
| | Number of captions with official COVID-19 information | | | | | | |
| | Number of captions with unofficial COVID-19 information | 2 | 100.0% | | | | |
| | Number of hashtags related with COVID-19 | | | | | | |

Source: Own elaboration.

**Table 5.** Data Analyzed - Peru

| | | PERU | | | | | |
|---|---|---|---|---|---|---|---|
| Factor 1: Relevance - Social Network Presence | Name | George Forsyth | | Hernando de Soto | | Verónika Mendoza | |
| | Number of followers | 317,000 | | 6,586 | | 17,900 | |
| | Date account established | 4/11/2004 | | 29/11/2016 | | 9/7/2017 | |
| | Total number of posts | 493 | | 23 | | 395 | |
| | Number of posts in the COVID-19 period | 148 | | 15 | | 122 | |
| | Percentage of total posts mentioning COVID-19 | 3.9% | | 0.0% | | 2.5% | |
| | Percentage of posts in the COVID-19 period mentioning COVID-19 | 12.8% | | 0.0% | | 8.2% | |
| | Number of posts/videos with candidate appearance | 93 | 62.8% | 5 | 33.3% | 43 | 35.2% |
| | Number of posts/videos without candidate appearance (text, advert, etc) | 55 | 37.2% | 10 | 66.7% | 79 | 64.8% |
| Factor 2: Prevention - Mask Use | Candidate with mask | 71 | 76.3% | 2 | 40.0% | 17 | 39.5% |
| | Candidate without mask | 22 | 23.7% | 3 | 60.0% | 26 | 60.5% |
| | Candidate use of mask unidentified | | | | | | |
| Factor 3: Prevention Social Distancing | Candidate alone | 35 | 37.6% | 2 | 40.0% | 26 | 60.5% |
| | Candidate accompanied by others | 58 | 62.4% | 3 | 60.0% | 17 | 39.5% |
| Factor 4: Political Discourse: COVID-19 Situation | Number of captions mentioning COVID-19 | 19 | | | | 10 | |
| | Number of captions with official COVID-19 information | 5 | 26.3% | | | | |
| | Number of captions with unofficial COVID-19 information | 14 | 73.7% | | | 10 | 100.0% |
| | Number of hashtags related with COVID-19 | | | | | | |

Source: Own elaboration.

## 3.1. Relevance

Despite the ongoing scourge of the pandemic, the candidates analyzed showed inconsistent and lopsided behavior with regards to following WHO recommendations. The effects of COVID-19, such as information on infection rates or prevention, did not stand out in their Instagram posts, despite it being a pre-electoral period. In most cases they were young, recently created accounts, as was the case of the candidates in Honduras. They appeared to have been opened for the campaign and none of the three showed their more personal side. Rather, they gave the impression of being managed by the respective communication teams of the candidates. Older and more personal accounts, through which it was possible to find out about the interests, hobbies, and families of the candidates, belonged to the candidates of Ecuador, Peru, and Chile.

In each country, one candidate stood out for their activity on Instagram: Lasso in Ecuador, Jiles in Chile, Forsyth in Peru, and Rosenthal in Honduras. However, despite being the most active candidates, none stood out for their content related to COVID-19. Lasso dedicated 18.7% of his posts, Jiles 7.1%, Forsyth 12.8%, and Rosenthal did not dedicate a single line to the pandemic. Castro (Honduras), De Soto (Peru), and Arauz (Ecuador) with 0.8%, presented total or near total silence on the virus. The candidates who spoke the most about the pandemic and preventive measures were Pérez in Ecuador (22.4%), Lavin in Chile as prefect and mayor (33.3%), Mendoza in Peru (23.3%), and minimally Asfura in Honduras as mayor of the Central District (1.4%).

## 3.2. Prevention

WHO advised preventive behaviors and measures focused on the use of masks and social distancing, and above all avoiding large crowds at public events.

In Ecuador, the candidate who most adopted the recommendation to wear a mask was Arauz, who wore one in 60.5% of the posts analyzed, especially when accompanied (97.1%), despite not taking too much care with social distancing (28.0%). Arauz was followed by Pérez, who wore a mask in 45.8% of his posts, although not always when accompanied (he only wore one on 32.1% of these occasions), but he was more careful with social distancing (66.2%). Lasso was the candidate who most neglected the measures and recommendations of WHO, given that he only wore a mask in 25.8% of the posts, rarely when surrounded by people (25.8%), and he was careless with social distancing (33.3%).

The candidates in Chile followed the recommendations inconsistently. In the use of masks, both Jiles and Lavín showed a similar rate of use (64.9% and 63.1%, respectively), in addition to maintaining social distancing on more than 60.0% of occasions. Regarding the non-use of a mask with peo-

ple, Lavín stood out with only 14.0% of the analyzed sample. However, of that 14.0% it would be possible to subtract 5.8% of the publications which were made in a family environment.

In Peru, Forsyth used a mask for most of his appearances (76.3%), especially when accompanied in both open and closed spaces, and also maintained social distancing (72.4%). Mendoza had a lower rate of mask use (39.5%), but when accompanied she wore it on all occasions. Both greet others according to the new normal, i.e., with a fist bump or an elbow bump. De Soto wore a mask in 60.0% of his publications, but there were only five.

In Honduras, the incidence of the virus was lower than in the other countries analyzed. Both Asfura and Rosenthal showed a low rate of mask use in their publications (29.0%, respectively) and did not maintain social distancing on 70.0% of occasions. In contrast, Castro almost always wore a face mask (70.0%), including in his profile picture, but on no occasion does he maintain social distancing from third parties.

### 3.3. Discourse in the Face of COVID-19

In the analysis of the taxonomy of speaker intention related to COVID-19, most of the messages of the candidates were informative and expressive, especially from those already holding public office. In the case of Pérez (Ecuador), 53.9% were messages with an informative intention taxonomy, which was not surprising given the function of the candidate as a provincial prefect. The same happened with Forsyth (Peru) who, as mayor of La Victoria, visited several groups to distribute masks and, at the same time, to inform about the risks of the pandemic as well as the importance of preventive measures. The case of Lavín was similar, with 144 publications related to COVID-19 with informative intention (64.6%) or to promote events (26.4%). In the case of Jadue, mayor of a populous district of Santiago de Chile, he published 4.3 times fewer than his peer Lavín and his main intention was promotive (43.3%) and to a lesser extent informative (33.3%).

Informative publications were also in the majority for candidates such as Lasso and Pérez (Ecuador), Mendoza (Peru), and Jiles and Jadue (Chile). The latter also stood out in expressive publications, similar to the candidates from Ecuador Lasso (48.9%) and Pérez (30.3%), and the Peruvian Mendoza (20.0%). However, the publications of Chile were more advocative, especially Jiles (39.6%).

Cases without statistical significance were Arauz (Ecuador), whose only post on Instagram corresponded to an informative message, and Asfura, the only candidate in Honduras who mentioned the pandemic, once with an informative intention and the other time with an expressive intention.

In the contextual taxonomy, differences were found between candidates from the same country, as in the cases of Chile and Ecuador. In Chile, the posts of Jiles focused on minimal (62.3%) and close (35.8%) relationships and those of Lavín stood out for significant relationships (62.5%). In Ecua-

dor, the posts of Pérez demonstrated the most significant relationships, with 51.3%, as opposed to the messages of Lasso, where their significance represented only 20.1%. In Peru, the candidates stood out for posts with a significant relationship (Forsyth with 52.6% and Mendoza with 50.0%) and a close relationship (Forsyth with 42.1% and Mendoza with 40.0%). Finally, in Honduras, the only two posts by Asfura had a close relationship, highlighted by the absence of images or comments.

Regarding the semiotic taxonomy, most of the information corresponded to the additive and parallel categories for all of the candidates from the four countries analyzed. The candidates from Ecuador stood out in the parallel category with Lasso (69.8%) and Pérez (48.7%), as well as the candidate Jiles from Chile (62.3%). In comparison, Forsyth from Peru had equal percentages for parallel and additive information (47.4%), comparable to his opponent Mendoza. Additive publications predominated in Chile with Lavín (87.5%) and Jadue (43.3%), as well as with Pérez in Ecuador (47.4%). Divergent publications had little relevance across all candidates.

## 4. Conclusions

According to their Instagram accounts, candidates seemed more concerned about the usual problems of the countries (economy, minorities, farm workers, weather disasters, corruption, new constitution, etc.) than reporting on preventive measures related to the pandemic. Several publications showed actions contradictory to WHO recommendations. Although mask use increased throughout 2020, its use was not a consistent practice among the candidates analyzed. For example, one might have thought the coronavirus had not reached Honduras given that only one candidate, Asfura, talked about COVID-19 in two publications, but appeared without a mask in 71.2% of the posts and without maintaining minimum social distance (73.3%).

The omission of COVID-19 from the various publications, together with the poor compliance by some candidates with the health recommendations, projected a misinformative discourse that could potentially confuse the public through imprecise and vague information, even more so when their multimodal messages were posted in situations that were clearly described as having a high risk of contagion. Candidates could generate distrust and confusion because there was no correlation between the discourse on the pandemic and the compliance with or reinforcement of WHO recommendations. In Ecuador for example, Lasso, the candidate with the most followers (121,000), with the most activity in the study period (743 posts), and with the most publications on the coronavirus (18.7%), appeared without a mask in 72.0% of the images and without maintaining social distancing (66.7%).

The behavior of the candidates in the face of COVID-19 was not consistent in all publications. Our hypothesis was confirmed: there was a lack of consistency in the discourse of coronavirus prevention with respect to the

images and texts projected by the candidates on their Instagram accounts. Also, the majority of publications referring to the coronavirus did not refer to prevention/education, but rather to the promotion of the image of the candidate. The misinformation detected referred to imprecise, vague, and sometimes even contradictory information (Karlova and Lee, 2011; Karlova and Fisher, 2013). The multimodal discourse of candidates in the face of the pandemic was mostly misinformative regarding COVID-19, as it could generate confusion and mistrust in the recipients. As the results showed, there were variations among the countries analyzed and among the candidates. Moreover, the misinformative effect lay in the modality of the image. It was in the photographs the candidates shared on Instagram that contradictions were observed, e.g., whether or not the candidates talked about the pandemic, or if the prevention measures against COVID-19 recommended by WHO were not always followed. Therefore, despite not detecting intentionality on the part of the issuer (which would be considered disinformation), the misinformation was accidental or innocent (Buckland, 1991) and could potentially generate distrust and confusion in the receiver.

In addition to being a serious health problem, the pandemic was also a political issue as government decisions impacted discourse and the public images of the candidates. In this respect, the speeches, recommendations, or relative positions regarding the pandemic served to inform the voters and showed how the candidates assumed relevance and presence from various positions, such as supportive, critical-denunciating, managing, or passive-indifferent. In this sense, the configuration of a message did not respond to a taxonomy of absolute intention, as other intentions could always be interpreted behind a message and associated discourse, such as the case where the multimodal discourse on the pandemic served as a broader means of communication than a mere reference to the pandemic.

Beyond the classification of the taxonomy of the speaker's intention, four message styles could be identified in the posts of the candidates related to the pandemic, which were understood and contextualized in the particular context of each country. In Ecuador, in general, a supportive-managing communicative style was observed, i.e., they tended to connect with the emotions of the users (expressive category), demonstrated concern for the situation of those affected, and, via Instagram, demonstrated concrete actions to alleviate and improve the conditions of those affected (informative category). In Chile, the communication style was critical-denunciating. This was evidenced in the allusions to the "ineffective" management of the pandemic by the opponents of the current Government (provocative category). The style was also advocative, arguing and demonstrating how the management of the candidate could potentially better handle the situation (advocative taxonomy). Honduras showed a passive-indifferent communicative style, where politicians had a passive or directly null attitude both in the management of the pandemic by the Government and in their own actions. In Peru, the predominant style in the posts was management-advocative, especially those

politicians who were running for the presidency from their current positions (mayors or governors). The discourse was oriented towards addressing the pandemic from the action of the person involved, demonstrating how they were managing and/or helping to address the problems of their jurisdiction (informative category).

In this aspect, note that the predominance of posts with minimal relation (in the contextual taxonomy) and parallel (in the semiotic taxonomy) indicated little recognition of the textual modality in Instagram, instead entrusting the delivery of meaning to images. However, it was striking how some candidates employed the textual semiotic resource to take their posts further, while others only reinforced an image through the textual format.

Likewise, the results of the analysis showed that a minimal contextual taxonomy relationship was not necessarily negative. In the triangulation of results, such as fewer posts and communications about the pandemic with more social activity while accompanied by other people (social distancing and greater use of masks), note that Arauz (Ecuador), Jiles (Chile), Forsyth (Peru), and Castro (Honduras) reinforced their messages visually. They represented the incorporation of the measures into their behavior that went beyond the formal discourse (texts), i.e., they did not talk about the measures, but they put them into practice in precautionary situations, particularly the use of masks. This behavior might contribute to the assumption of protective measures as a permanent and daily fact. Therefore, although the text and image had a minimal relationship, as a taxonomic evaluation they were the ones who most respected the rules in a proselytizing context.

The misinformative discourse could not be considered as a fulfilled discourse, at least not in a clear-cut way. The analysis showed that the results can vary in the sense that the shaping of the discourse of the candidate had many elements in its composition and in the end not all of them could be fulfilled. The misinformative behavior of the candidates varied in relation to time, such as the context of the communicative situation and the relationship of the candidate with different people. In this sense, there were messages which in their multimodal construction were completely clear and coherent (image, text, situation, and all the precautionary measures), but in others they were only half fulfilled or even varied in similar situations.

Of the total number of posts analyzed, only 1.0% in Honduras, 10.0% in Peru, 14.0% in Ecuador, and 17.3% in Chile were related to the management of the pandemic. That is to say, in the speeches of the candidates the importance given to the pandemic issue did not reach 20.0%. Moreover, in most cases it was simply an informative message, especially in the case of candidates who already held public office. The conclusion was that politicians focused a great deal on their campaigns and little on the pandemic. The omission of COVID-19 from most publications projected a misinformative discourse that may have potentially confused the public.

## Bibliographical references

AUSTIN, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

BATEMAN, J. (2014). *Text and image: A critical introduction to the visual/verbal divide*. New York: Routledge.

BEDNAR, P. and WELCH, C. (2008). "Bias, misinformation and the paradox of neutrality". *Informing Science*, 11, 85-106.

BEVELANDER, P. and WODAK, R. (2019). *Europe at the crossroads: Confronting populist, nationalist and global challenges*. Lund: Nordic Academic Press.

BUCKLAND, M. K. (1991). "Information as a thing". *Journal of the American Society for Information Science*, 42 (5), 351-360.

CÁRCAMO-MORALES, B. (2018). "El análisis del discurso multimodal: una comparación de propuestas metodológicas". *Forma y Función*, 31(2), 145-174.
<https://doi.org/10.15446/fyf.v31n2.74660>

CASTILLO-ESPARCIA, A.; FERNÁNDEZ-SOUTO, A. and PUENTES-RIVERA, I. (2020). "Comunicación política y Covid-19. Estrategias del Gobierno de España". *Profesional de la información*, 29 (4), e290419.
<https://doi.org/10.3145/epi.2020.jul.19>

FALLIS, D. (2009). *A conceptual analysis of disinformation*. Paper presented at the iConference, Chapel Hill, NC.

FOX, C. J. (1983). *Information and misinformation: An investigation of the notions of information, misinformation, informing, and misinforming*. Westport, CT: Greenwood.

HALLIDAY, A. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.

HJØRLAND, B. (2007). "Information: Objective or subjective/situational?". *Journal of the American Society for Information Science*, 58 (10), 1448-1456.

KARLOVA, N. A. and FISHER, K. E. (2013). "A social diffusion model of misinformation and disinformation for understanding human information behaviour". *Information Research*. Retrieved from <http://informationr.net/ir/18-1/paper573.html>

KARLOVA, N. A. and LEE, J. H. (2011). "Notes from the Underground City of Disinformation: A Conceptual Investigation". *Proceedings of the American Society for Information Science and Technology*, 48(1), 1-9.
<https://doi.org/10.1002/meet.2011.14504801133>

KLOEPFER, R. (1977). "Komplementaritat von Sprache und Bild, Am Beispiel von Comic, Karikatur, und Reklame". In: R. POSNER and H.-P. REINECKE (Eds.). *Zeichenprozesse, Semiotische Forschung in den Einzelwissenschaften*. Athenaum, 129-145.

KRUK, J.; LUBIN, J.; SIKKA, K.; LIN, X.; JURAFSKY, D. and DIVAKARAN, A. (2019). "Integrating Text and Image: Determining Multimodal Document Intent in Instagram Posts". *ArXiv Preprint AsXiv:1904.09073*. Retrieved from <https://arxiv.org/abs/1904.09073>.

LALANCETTE, M. and RAYNAULD, V. (2017). "The power of political image: Justin Trudeau, Instagram, and celebrity politics". *American Behavioral Scientist*, 1-37.
<https://www.doi.org/10.1177/0002764217744838>

LARSSON, A. O. (2017). "Skiing all the way to the polls". *Convergence: The International Journal of Research into New Media Technologies*, 25 (5-6), 1096-1110.
<https://www.doi.org/10.1177/1354856517741132>

Lees-Marshment, J. (2001). "The product, sales and market-oriented party. How labour learnt to market the product, not just the presentation". *European Journal of Marketing*, 35 (10), 1074-1084.
<https://doi.org/10.1108/EUM0000000005959>

Librero, E. (2020). *White paper: Instagram en Latinoamérica & España*. IEDGE Business School. Retrieved from <https://www.iedge.eu/instagram-2020-en-latinoamerica-y-espana>.

López-Rabadán, P.; López-Meri, A. and Doménech-Fabregat, H. (2016). "La imagen política en Twitter. Usos y estrategias de los partidos políticos españoles". *Index.comunicación*, 6 (1), 165-195. Retrieved from <http://journals.sfu.ca/indexcomunicacion/index.php/indexcomunicacion/article/view/270>.

Losee, R. M. (1997). "A discipline independent definition of information". *Journal of the American Society for Information Science*, 48 (3), 254-269.
<https://dl.acm.org/doi/abs/10.5555/249607.249615>

Lozano-Recalde, C. (2021). "Comunicación política en Instagram durante la emergencia sanitaria por el COVID 19. El caso de Otto Sonnenholzner en Ecuador". *ComHumanitas: Revista Científica de Comunicación*, 11 (3), 75-95.
<https://doi.org/10.31207/rch.v11i3.259>

Manfredi-Sánchez, J.; Amado-Suárez, A. and Waisbord, S. (2021). "Presidential Twitter in the face of COVID-19: Between populism and pop politics". *Comunicar,* 66, 83-94.
<https://doi.org/10.3916/C66-2021-07>

Marland, A. (2012). "Political photography, journalism, and framing in the digital age: The management of visual media by the Prime Minister of Canada". *The International Journal of Press/Politics*, 17 (2), 214-233.
<https://doi.org/10.1177/1940161211433838>

Marsh, E. E. and Domas White, M. (2003). "A taxonomy of relationships between images and text". *Journal of Documentation*, 59 (6), 647-672.

Menéndez, S. (2012). "Multimodalidad y estrategias discursivas: un abordaje metodológico". *Asociación Latinoamericana de Estudios del Discurso*, 12 (1), 57-73.
<https://doi.org/fdmt>

Moffitt, B. and Tormey, S. (2014). "Rethinking populism: Politics mediatisation and political style". *Political Studies*, 62 (2), 381-397.
<https://doi.org/10.1111/1467-9248.12032>

Muñoz, C. L. and Towner, T. L. (2017). "The image is the message: Instagram marketing and the 2016 presidential primary season". *Journal of Political Marketing*, 16 (3-4), 290-318.
<https://www.doi.org/10.1080/15377857.2017.1334254>

Niknam, F.; Samadbeik, M.; Fatehi, F.; Shirdel, M.; Rezazadeh, M. and Bastani, P. (2020). "COVID-19 on Instagram: A content analysis of selected accounts". *Health Policy and Technology*, 10 (1), 165-173.
<https://doi.org/10.1016/j.hlpt.2020.10.016>

O'halloran, K. (2012). "Multimodal Discourse Analysis". In: K. Hyland and B. Paltridge (Eds.). *Companion to Discours.* London and New York: Continuum, 120-137.

O'halloran, K. L.; Podlasov, A.; Chua, A.; Tisse, C.-L. and Lim, F. V. (2014). "Challenges and Solutions to Multimodal Analysis: Technology, Theory and Practice". In: Y. Fang and J. Webster (Eds.). *Developing Systemic Functional Linguistics: Theory and Application*. London, UK: Equinox, 271-297.

PARÉ, D. J. and BERGER, F. (2008). "Political marketing Canadian style? The Conservative Party and the 2006 feder-al election". *Canadian Journal of Communication*, 33, 39-63.
<https://doi.org/10.22230/cjc.2008v33n1a1928>

PARODI, G. (2010). "Multisemiosis y lingüística de corpus: artefactos (multi)semióticos en los textos de seis disciplinas en el corpus pucv-2010". *Revista de Lingüística Teórica y Aplicada*, 48 (2), 33-70.
<https://doi.org/10.4067/S0718-48832010000200003>

PASQUALI, M. (2021). Las muertes por coronavirus en América Latina. *Statista*. Retrieved from <https://es.statista.com/grafico/21572/muertes-por-covid-19-cada-100000-habitantes-en-latinoamerica/>.

QUEVEDO-REDONDO, R. and PORTALÉS-OLIVA, M. (2017). "Imagen y comunicación política en Instagram. Celebrificación de los candidatos a la presidencia del Gobierno". *El Profesional de la Información*, 26 (5), 916.
<https://www.doi.org/10.3145/epi.2017.sep.13>

RUIZ-DEL-OLMO, F. J. and BUSTOS-DÍAZ, J. (2016). "Del tweet a la fotografía, la evolución de la comunicación política en Twitter hacia la imagen. El caso del debate del estado de la nación en España (2015)". *Revista Latina de Comunicación Social*, 71, 108-123.
<https://doi.org/10.4185/RLCS-2016-1086>

SAMPIETRO, A. and SÁNCHEZ-CASTILLO, S. (2020). "Building a political image on Instagram: A study of the personal profile of Santiago Abascal (Vox) in 2018". *Communication and Society*, 33 (1), 169-184.
<https://doi.org/10.15581/003.33.1.169-184>

STAHL, B. (2006). "On the difference or equality of information, misinformation, and disinformation: A critical research perspective". *Informing Science Journal*, 9, 83-96.

VAN DIJK, T. (2006). "Discurso y manipulación: Discusión teórica y algunas aplicaciones". *Revista Signos*, 39 (60), 49-74.
<http://dx.doi.org/10.4067/S0718-09342006000100003>

VAN LEEUWEN, T. (2005). *Introducing Social Semiotics*. London: Routledge.

ZHOU, L.; BURGOON, J.; NUNAMAKER, J. and TWITCHELL, D. (2004). "Automated linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communication: A empirical investigation". *Group Decision Negotiation*, 13 (1), 81-106.

ZHOU, L. and ZHANG, D. (2007). "An ontology-supported misinformation model: Toward a digital misinformation library". *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37 (5), 804-813.

ZHANG, M.; HWA, R. and KOVASHKA, A. (2018). "Equal but not the same: Understanding the implicit relationship between persuasive images and text". *arXiv preprint arXiv:1807.08205.*

# Multimodal disinformation about otherness on the internet. The spread of racist, xenophobic and Islamophobic fake news in 2020*

José Gamir-Ríos
Universitat de València
jose.gamir@uv.es

Raquel Tarullo
Instituto de Política y Gobierno. CITNoBA (CONICET. UNNOBA. UNSAdA)
raqueltarullo@gmail.com

Miguel Ibáñez-Cuquerella
Universitat de València
miguel.ibanez.cuquerella@gmail.com

## Abstract

This work studies the use of disinformation to construct an image of otherness through the internet. We applied a content analysis methodology to the 161 racist, xenophobic or Islamophobic fake news pieces that were discredited in 2020 by the four Spanish information verification media entities accredited by the International Fact-Checking Network: *Maldita.es*, *Newtral*, *Efe Verifica* and *Verificat*. The results show that the most commonly used formats were image and video, that disinformation was most often based on taking information out of context and deception, and that the source could not be identified. The most shared characteristics associated otherness with receiving aid, violence and illegal immigration. The most commonly used images were photographs, which mostly showed people in a general manner (not individually). Despite this, disinformation was

---

not generated by manipulating images, but by inserting text over images. The use of supposed screenshots to create fictitious references or take truthful screenshots out of context was also notable.

**Keywords:** disinformation; social networks; otherness; racism; xenophobia; Islamophobia

**Resum.** *La desinformació multimodal sobre l'alteritat a Internet. Difusió de boles racistes, xenòfobes i islamòfobes el 2020*

Aquest treball estudia la construcció de la imatge de l'alteritat a través de la desinformació a Internet. Per fer-ho, aplica una metodologia d'anàlisi de contingut a les 161 boles de caràcter racista, xenòfob o islamòfob desmentides el 2020 pels quatre mitjans espanyols de verificació d'informació acreditats per la International Fact-Checking Network: *Maldita.es*, *Newtral*, *Efe Verifica* i *Verificat*. Els resultats evidencien que els formats més emprats van ser la imatge i el vídeo, i que les desinformacions es van basar majoritàriament en la descontextualització i l'engany, sense identificar-ne la font. Els atributs més transmesos van associar l'alteritat a la recepció d'ajudes, la violència i la immigració irregular. Les imatges més utilitzades van ser les fotografies, que majoritàriament van representar les persones de manera no individualitzada; no obstant això, la desinformació no es va produir a través de la manipulació visual, sinó de la inserció de text. També va ser notable la difusió de suposades captures de pantalla, aprofitades per crear referents ficticis o descontextualitzar-ne uns altres que existien realment.

**Paraules clau:** desinformació; xarxes socials; alteritat; racisme; xenofòbia; islamofòbia

**Resumen.** *La desinformación multimodal sobre la otredad en Internet. Difusión de bulos racistas, xenófobos e islamófobos en 2020*

Este trabajo estudia la construcción de la imagen de la otredad a través de la desinformación en Internet. Para ello, aplica una metodología de análisis de contenido a los 161 bulos de carácter racista, xenófobo o islamófobo desmentidos en 2020 por los cuatro medios españoles de verificación de información acreditados por la International Fact-Checking Network: *Maldita.es*, *Newtral*, *Efe Verifica* y *Verificat*. Los resultados evidencian que los formatos más empleados fueron la imagen y el vídeo, y que las desinformaciones se basaron mayoritariamente en la descontextualización y el engaño, sin identificar su fuente. Los atributos más transmitidos asociaron la otredad a la recepción de ayudas, la violencia y la inmigración irregular. Las imágenes más empleadas fueron las fotografías, que mayoritariamente representaron a las personas de manera no individualizada; sin embargo, la desinformación no se produjo a través de la manipulación visual, sino de la inserción de texto. También fue notable la difusión de supuestas capturas de pantalla, aprovechadas para crear referentes ficticios o descontextualizar otros realmente existentes.

**Palabras clave:** desinformación; redes sociales; otredad; racismo; xenofobia; islamofobia

## 1. Introduction

The spread of disinformation has increased with the popularisation of using social networks as a source for news and due to the growing suspicion of traditional media as intermediaries and interpreters, opening "the floodgates to

information and misinformation, truth and lies, scientific and non-scientific knowledge, facts and fiction" (Waisbord, 2018: 5). Thus, specific audiences with particular interests are reached (Mourão and Robertson, 2019: 2) in a context of confirmation bias (Wason, 1960) and cognitive dissonance (Festinger, 1962). This leads to people only looking for information that confirms their previous views, and ignoring data that question their prejudices (López-López, Castro Martínez and Oñate, 2020; Salaverría et al., 2020).

Consequently, disinformation is used to rally and unify people with related worldviews (us), and through appealing to negative emotions to polarise others (them) (Fernández, Revilla and Andaluz, 2020). It distorts images that people use to construct their opinion on issues about which they have no knowledge, and which they view as dangerous and threatening to their personal and collective safety (Engesser et al., 2017a). Populist ideology, which usually finds greater support in polarised contexts, uses these emotions to construct an otherness based on characteristics that define immigrants and ethnic, religious and sexual minorities as enemies of the people (Reinemann et al., 2017: 2), resulting in hate speech (Sanguinetti et al., 2018).

Academic research on disinformation has grown in recent years and has found that false content on immigration and racism is one of the most frequently studied categories, along with those related to politics (Guallar et al., 2020). However, this growing popularity has not resulted in an abundance of bibliography on the subject, in which the recent studies by Fernández et al. (2020), Molina-Cañabate and Magallón-Rosa (2019) and Vorobyeva et al. (2020) are noteworthy.

There has, however, been a variety of studies exploring how the Spanish press and television have represented and covered immigration and ethnic minorities (Igartua et al., 2005, 2007, 2011, 2012; Muñiz et al., 2006; 2008). Their conclusions indicate that the image the media issues of migrants and ethnic minorities is intertwined with socioeconomic threats, cultural dangers, aberrations, violence and crime (Marcos Ramos et al., 2014). In addition, studies that analysed the representation of immigration in Spanish fiction television programmes also associated immigrants with negative characteristics – aggressive, conflictive, disloyal, intolerant, illiterate, unemployed and unstable (Marcos Ramos et al., 2014; Lacalle, 2008; Ruiz-Collantes et al., 2006). These studies found that those stereotyped and skewed characteristics were used as attributes to construct an image of immigrants and ethnic minorities and contributed to the growth of xenophobia in Spain (Igartua et al., 2008, 2009).

This article uses a multimodal design to study the formal and discursive mechanisms used on the internet to construct an image of otherness. Its objective is to analyse the manipulation strategies used in fake news pieces that contain racist, xenophobic and Islamophobic data, as we believe disinformation plays a fundamental role in the construction of the image of migrants and persons belonging to ethnic or religious minorities.

## 2. Multimodal cheapfakes

The definition of disinformation is content that can lead to an incorrect view of the political world and that serves to deceive or "intentionally alter the perception of large groups of people or societies and influence their behaviour (political, economic, ideological, etc.)" (Del-Fresno-García, 2019: 3). This includes false content – originating from websites, consisting of unreal stories and attributing authority to the sources that are issuing the deceptive information (Salaverría et al., 2020) – but also rumours and incorrect or hyperpartisan information (Tucker et al., 2018: 3).

Information disorders (Del-Fresno-García, 2019) pollute the truth (Corner, 2017: 1105) and their objective is to "cause harm, obtain benefits and influence people's behaviour, although some ends are usually hidden and remain invisible" (Guallar et al., 2020: 597). Their level of deception may be influenced by bias or sensationalism, which can strain the truth without directly including the disinformation (Mourão and Robertson, 2019: 2). Populist messages and hate speech often adopt these tactics (Engesser et al., 2017a; Engesser et al., 2017b).

Salaverría et al. (2020) use the term 'fake news pieces' to refer to content that seems true but whose intention is to deceive the public, and content that is shared by any means of communication or digital platform (p. 4) – a definition that we will use for this study. Likewise, they produced a scale of severity, measured according to two categories: deceit and intention. On this basis, the higher the level of deceit and intention to spread the fake news pieces, the more serious it was (Salaverría et al., 2020).

Although disinformation was present before the popularisation of social networks as sources of information (Newman et al., 2020), its spread – and the proliferation of characteristics consistent with previous prejudices – has grown exponentially, encouraging the creation of communities of similar opinions (López-López et al., 2020). Previous studies have found that messages with disinformative content are shared more quickly than those with true content (Tarullo and Amado, 2020) and that there is less delay in sharing messages that fit individual preferences (Calvo and Aruguete, 2018).

The circulation of multimodal content (information in visual and text format) through social networks has become more complex for several reasons (Zhou, Wu and Zafarani, 2020). First, the methods used to detect disinformation identify textual information and the combination of text and image (Zafarani et al., 2019), but they have been unable to examine the relationship between the two (Jin et al., 2017), despite the fact that the inclusion of text is a form of image manipulation (Zhukova, 2019). Moreover, the combination of image and text is significant in political communication due to the fact that it is purposefully used to draw the attention by appealing to emotions (Corner, 2017). In addition, digital technologies facilitate the manipulation of visual content by offering tools that can completely transform material, creating disinformation narratives (Tandoc, Lim and Ling,

2018). In this respect, although deepfakes – that require quite advanced technological skills to falsify content – are seen as a principal problem in the intricate network of information disorders (Del-Fresno-García, 2019), recent studies have found that cheapfakes actually cause the most concern as they have gained the most ground in disinformative digital flows (Paris and Donovan, 2019). A cheapfake is "a media news piece that has been crudely manipulated, edited, mislabeled, or taken out of context to spread disinformation" (Schick, 2020) and does not require complex technological skills to produce or spread.

This makes the work carried out by verification platforms identifying disinformation absolutely crucial (Molina-Cañabate and Magallón-Rosa, 2019). Based on the above, and inspired by a recent article by Salaverría et al. (2020), this methodologically designed study has used the following research questions as a reference:

— RQ1. What formats were used for the dissemination of fake news pieces containing racist, xenophobic or Islamophobic information in Spain during 2020?
— RQ2. What sources did these fake news pieces use and to what degree of deception?
— RQ3. What characteristics did they transmit and with what frequency?
— RQ4. What format did the shared images have and what role did the text contained in them play?
— RQ5. In the case of fake news pieces that used photographs, what was their content?
— RQ6. In the case of fake news pieces that used other types of images, what was their reference and to what degree of truth were they reproduced?

## 3. Method

This study was limited to 2020 and used a content analysis methodology (Krippendorff, 1990).

### 3.1. Corpus

The research corpus consisted of fake news pieces that contained racist, xenophobic or Islamophobic information discredited in 2020 by the four Spanish information verification media accredited by the International Fact-Checking Network (IFCN), a network that in January 2021 included 75 media entities from all over the world specialised in verifying information. Specifically:

— *Maldita.es.* Digital media company founded by journalists Clara Jiménez and Julio Montes and edited by the non-profit organisation Maldita Contra la Desinformación.

— *Newtral.* Digital media company that has been edited by the Newtral Media Audiovisual production company since 2018, owned by the journalist Ana Pastor.
— *Efe Verifica.* The Efe Agency fact-checking department, since 2019.
— *Verificat.* Digital media company founded in 2019 by journalists Lorenzo Marini and Alba Tobell and edited by the non-profit organisation Associació Verificat.

One of the group's researchers collected and added the fake news pieces and refutations published by these media between 1 January and 31 December 2020 to a database. The total sample collected was 2,997 texts, of which 8.4% (n=253) referred to fake news pieces that construct otherness through hatred for ethnic, origin or religious reasons, a condition for inclusion that was observed by two of the author researchers collecting the samples, who resolved any differences jointly. After discarding the publications that contained compilations or repeatedly alluded to the same fake news piece, the final corpus comprised 161 units of analysis.

### 3.2. Analysis variables

The content analysis was carried out by applying a coding manual, as shown in Tables 1 and 2, which considers the study of six variables common to the entire corpus (n=161) and nine that are specific: three for fake news pieces spread through images (n=76), four for the cases in which these images were photographs (n=37) and two when they were screenshots of documents, posters, social network posts or media websites (n=27). The analysis considered both the headline and the text contained in the refutation and the fake news piece as well as the multimedia materials included in the disinformation material – links, audio files, images and videos – thus ensuring that the study carried out was fully multimodal.

The common variables of source and type of fake news piece are inspired by the recent research carried out by Salaverría et al. (2020) on disinformation linked to the COVID-19 pandemic in Spain. The former categorises the source as anonymous when it is not reflected in the fake news piece; fictitious, when it has been invented by the fake news piece itself; an impersonated entity when the entity exists, but has not delivered the message; and a real entity, when the entity exists and has issued the message. The latter distinguishes between: 1) joke, spreading false information with mocking or satirical intent; 2) exaggeration, a message that exaggerates a certain aspect of reality to the point that it crosses into deceit; 3) taken out of context, fake news piece that distorts the conditions in which information was produced; and 4) deceptive, disinformation that completely falsifies reality to give the wrong idea or impression (Salaverría et al., 2020: 10). The other variables and categories were created by the authors.

**Table 1.** Common variables and categories of analysis

| Variables | Categories | Reliability |
|---|---|---|
| 1. Format | 1) Simple text. 2) Link. 3) Audio file. 4) Image. 5) Video. | $\alpha k = 0.857$ |
| 2. Addition | [only if v1≠1] 1) Text. 2) No addition. | agreement = 100% |
| 3. Source | 1) Anonymous. 2) Fictitious. 3) Impersonated political/institutional entity. 4) Impersonated media entity. 5) Real political/institutional entity. 6) Real media entity. 7) Real citizen. | $\alpha k = 0.854$ |
| 4. Type | 1) Joke. 2) Exaggeration. 3) Taken out of context. 4) Deceptive. | $\alpha k = 0.747$ |
| 5. Characterisation | 1) Illegal immigration. 2) Benefits and favourable treatment. 3) Lack of Western values and/or civility. 4) Transmission of diseases. 5) Vandalism / crime / violence. 6) Terrorism. 7) Intention of territorial expansion. 8) Others / Does not apply. | $\alpha k = 0.821$ |
| 6. Recurrence | 1) Frequent. 2) Infrequent. | agreement = 90% |

Source: Authors' elaboration.

**Table 2.** Specific variables and categories of analysis

| Variables | Categories | Reliability |
|---|---|---|
| 7. Type of image | [only if v1=4] 1) Photograph. 2) Screenshot of a document, poster, social network post or media website. 3) Image-text composition. | $\alpha k = 1$ |
| 8. Text on the fake news piece | [only if v7=1 or 3] 1) Yes. 2) No. | agreement = 100% |
| 9. Function of the text | [only if v8 =1] 1) Locate the image. 2) Put the image into context. 3) Verbatim quote. | $\alpha k = 0.899$ |
| 10. Number of photographs | [only if v7=1] 1, 2, 3, 4, 5, 6. | $\alpha k = 1$ |
| 11. Function of the variety | [only if v10>1] 1) Comparison. 2) Reinforcement. | agreement = 100% |
| 12. Content of the photograph | [only if v7=1] 1) Contains people. 2) Does not contain people. | agreement = 100% |
| 13. Main characters in the photograph | [only if v12=1] 1) One person. 2) A group of people. 3) Crowd. | $\alpha k = 0.856$ |
| 14. Type of screenshot | [only if v7=2] 1) Document. 2) A social network post. 3) Media entity website. 4) Poster. | $\alpha k = 0.871$ |
| 15. Truthfulness of the screenshot | [only if v7=2] 1) Exists. 2) Does not exist. | agreement = 85.2% |

Source: Authors' elaboration.

The coding of the common variables was carried out by one researcher. To validate its reliability, a test was performed in which that researcher and an author researcher separately analysed a sample made up of 12.5% of the fake news pieces (n=20). In the case of the specific variables, the entire corpus

was coded by two researchers, also separately from each other. In both cases, the agreement percentages obtained in the dichotomous variables and Krippendorff's alpha coefficients obtained in the categorical variables, shown in Tables 1 and 2, allowed us to validate the reliability of the research (Igartua, 2006).

## 4. Results

### 4.1. Format of the fake news pieces

The main format taken by the disinformation was static image (47.2%, n=76), audiovisual (31.1%, n=50), plain text (10.6%, n=17), link to media stories (6.2%, n=10) and audio files (5%, n=8). Among the fake news pieces that did not use exclusively text (n=144), only 18.7% (n=27) were shared in their original format (images, videos, links and audio files), while 81.3% (n=117) were shared with additional texts written on them or shared by the users themselves.

The cross-analysis of both variables (Table 3) showed that 96% of the videos (n=48), 78.9% of the images (n=60) and 60% of the links (n=6) were shared with additional text; the proportion is only inverse in the case of audio files, with which text anchoring was associated in 37.5% of the cases (n=3).

**Table 3.** Formats of the fake news pieces and text additions

| Format / additional text | With text | No text | Total | |
|---|---|---|---|---|
| | | | Number of fake news pieces | Percentage |
| Simple text | Not applicable | Not applicable | 17 | 10.6% |
| Link | 6 | 4 | 10 | 6.2% |
| Audio | 3 | 5 | 8 | 5.0% |
| Image | 60 | 16 | 76 | 47.2% |
| Video | 48 | 2 | 50 | 31.1% |
| **Total** | 117 (81.3%) | 27 (18.7%) | 161 | 100% |

Source: Authors' elaboration.

### 4.2. Fake news sources and types

Some 50.3% of the disinformation pieces did not mention a source (n=81), 20.4% impersonated the source (n=36) and another 20.4% used real sources (n=36). In both these last two cases, the source came mainly from a political-institutional scope. Only 5% of fake news pieces invented the source (n=8). The most popular types of fake news pieces were deceptive (41.8%, n=77) and taken out of context (41.6%, n=67), while exaggerations (7.5%, n=12) and jokes (3.1%, n=5) were a minority.

The cross-analysis of both variables (Table 4) confirmed that the fake news pieces that did not identify a source and those that used real sources, mostly used a method of taking the information out of context, and those that impersonated sources opted for the deceptive method.

**Table 4.** Fake news sources and types

| Source / type of fake news | Joke | Exaggeration | Taken out of context | Deceptive | Total | |
|---|---|---|---|---|---|---|
| | | | | | Number of fake news pieces | Percentage |
| Anonymous | 2 | 7 | 43 | 29 | 81 | 50.3% |
| Fictional | 0 | 0 | 0 | 8 | 8 | 5.0% |
| Impersonated political entity | 2 | 0 | 4 | 27 | 33 | 20.5% |
| Impersonated media entity | 0 | 0 | 0 | 3 | 3 | 1.9% |
| Real political entity | 0 | 2 | 13 | 4 | 19 | 11.8% |
| Real media entity | 0 | 3 | 3 | 5 | 11 | 6.8% |
| Real citizen | 1 | 0 | 4 | 1 | 6 | 3.7% |
| Total | 5 (3.1%) | 12 (7.5%) | 67 (41.6%) | 77 (41.8%) | 161 | 100% |

Source: Authors' elaboration.

### 4.3. Characteristics and frequency of the fake news pieces

The characteristics most associated with otherness were a supposed favourable treatment and receiving aid (35.4%, n=57), vandalism and crime (21.1%, n=34) and illegal immigration (12.4%, n=20). The least present characteristics were a lack of Western values (7.5%, n=12), disease transmission (6.2%, n=10), terrorism (5.6%, n=9) and intention of territorial expansion (4.3%, n=7).

46.6% of the fake news pieces analysed were frequent (n=75), compared to 53.4% that were not (n=86). This condition was established with the refutations published by the four media entities that the sample comprised, and a repeated presence of the fake news piece in one or more media platforms on different dates.

As shown by Table 5, the characteristics that were more frequent than the average of 46.6% were those that characterised people of different origin, ethnicity or religion with a lack of Western values (75% frequency, n=9) and those that associated them with terrorism (66%, n=6).

**Table 5.** Characteristics associated with otherness in fake news pieces and their frequency

| Characteristic / frequency | Frequent | Infrequent | Total | |
|---|---|---|---|---|
| | | | Number of fake news pieces | Percentage |
| Illegal immigration | 9 | 11 | 20 | 12.4% |
| Supposedly favourable treatment | 26 | 31 | 57 | 35.4% |
| Lack of Western values | 9 | 3 | 12 | 7.5% |
| Transmission of diseases | 3 | 7 | 10 | 6.2% |
| Vandalism and crime | 12 | 22 | 34 | 21.1% |
| Terrorism | 6 | 3 | 9 | 5.6% |
| Intention of territorial expansion | 3 | 4 | 7 | 4.3% |
| Other / Not applicable | 7 | 5 | 12 | 7.5% |
| Total | 75 (46.6%) | 86 (53.4%) | 161 | 100.0% |

Source: Authors' elaboration.

## 4.4. Format of the images

Photographs made up 48.7% of the images (n=37); 35.5% were screenshots of documents, posters or social network posts / media entity websites (n=27); and 15.8% were image-text compositions created for social networks (n=12). This category includes both the layout of text on a coloured background and the combination of text and photography as long as the image covers less than half of the composition.

In turn, the study of the presence of text over the images that allow this anchoring (photographs and compositions, n=49) shows that it was only present in 42.9% of them (n=21). This analysis excluded screenshots, as their function is to produce or reproduce references to text, which we will explore in section 4.6.

The cross-analysis of both variables (Table 6) shows that only 24.3% of the photographs were shared with text printed over them. Among the functions of the text printed over the photographs and in the image-text compositions, the addition of contextual explanations stands out (66%, n=14), while adding verbatim quotes (23.8%, n=5) and locating the images (9.5%, n=2) were a minority.

**Table 6.** Format of images and presence of text in the fake news piece itself

| Format of images / text included in the fake news piece itself | No text | Text | | | Total | |
|---|---|---|---|---|---|---|
| | | Place | Context | Quote | Number of fake news pieces | Percentage |
| Photograph | 28 | 2 | 7 | 0 | 37 | 48.7% |
| Screenshot | Not applicable | Not applicable | Not applicable | Not applicable | 27 | 35.5% |
| Image-text composition | 0 | 0 | 7 | 5 | 12 | 15.8% |
| Total | 28 | 2 | 14 | 5 | 76 | 100.0% |

Source: Authors' elaboration.

## 4.5. Content and elements of manipulation of the photographs

Regarding the content of the photographs (n=37), 73% showed people in different situations (n=27), compared to 27% that reproduced objects or interior or exterior spaces (n=10). Of the photographs that included people, only 18.6% (n=5) showed a single person; in most cases it was impossible to make out who was carrying out the action, as they mostly showed a group of people (40.7%, n=11) or a crowd (40.7%, n=11).

In turn, 67.6% of the fake news pieces that reproduced photographs showed a single image (n=25), compared to 32.4% that contained two or more images (n=12). When it occurred, this diversity was more oriented to reinforcing the message (n=10) than to comparison (n=2).

Finally, 86.5% of the fake news pieces that included photographs reproduced the images in those photographs without any editing (n=32), compared to 13.5% that did edit the images (n=5) using resources such as reframing, substitutions or highlighting.

## 4.6. Content and truth of the screenshots

Regarding the content of the screenshots (n=27), 51.9% referenced an official document or communication (n=14), 22.2% showed a social network post (n=6), 18.5% showed a poster (n=5) and only 7.4% referenced a news item. This type of publication included a true reference 59.3% of the time (n=16), while in 40.7% of cases it had a reference that did not exist (n=11).

The cross-analysis of both variables (Table 7) shows that the screenshots of documents or posters mostly included true references, while the screenshots of news items were always false. In screenshots of social network posts, there was a balance between true and false references.

**Table 7.** Format of the screenshot and truth of the reference

| Format of the screenshot/ truth of the reference | True | False | Total | |
|---|---|---|---|---|
| | | | Number of fake news pieces | Percentage |
| Document | 10 | 4 | 14 | 51.9% |
| Screenshot of social network posts | 3 | 3 | 6 | 22.2% |
| Screenshot of news item | 0 | 2 | 2 | 7.4% |
| Poster | 3 | 2 | 5 | 18.5% |
| Total | 16 (59.3%) | 11 (40.7%) | 27 | 100.0% |

Source: Authors' elaboration.

## 5. Discussion and conclusions

This analysis of the racist, xenophobic and Islamophobic fake news pieces shared in Spain throughout 2020 has allowed us to draw conclusions on the

multimodal manipulations used to construct disinformation and a deceptive representation of otherness on the internet.

Unlike what was found in previous studies, which concluded that fake news pieces operated mainly textually (Salaverría et al., 2020), this research study confirms the predominance of visual and audiovisual formats. Moreover, we saw that disinformation is not generated through the manipulation of these resources, with image editing techniques, but through the addition of texts, which shows the importance of multimodality when constructing deceptive content. Similarly, fake news pieces are not only constructed through exaggeration – which contributes to a stereotyped representation of others, as has already been confirmed by previous studies on the image of immigration in Spanish press and television (Igartua et al, 2005, 2007, 2011, 2012; Lacalle, 2008; Marcos Ramos et al., 2014; Muñiz et al., 2006, 2008; Ruiz-Collantes et al., 2006) – but often by taking the image out of context and showing it in a deceptive manner. Thus, fake news pieces not only contribute to multimodal disinformation about otherness, but also facilitate the generation of hate speech (Fernández et al., 2020; Sanguinetti et al., 2018) by being shared on social networks, resulting in the formation of communities with like-minded opinions, through an appeal to negative emotions towards the unknown, which is presented as dangerous (Engesser et al., 2017a; López-López et al., 2020).

This study confirms the representation of otherness as an entity that receives state aid and allegedly favourable treatment, but also one that vandalises, commits crime and is violent, accentuated by showing minorities as a group and not as individuals. Moreover, the fake news pieces that transmit characteristics representing an otherness that is contrary to Western values (us) and supports terrorist acts (them), that is, an enemy of the people (Reinemann et al., 2017: 2), are frequent.

Although previous studies have found that fake news pieces need credible sources to appear plausible (Salaverría et al., 2020), this study did not find that these fake news pieces used known sources to provide credibility, at least not in a generalised manner. Although most of the sources identified came from a political/institutional scope, the sources that were most predominant were anonymous and, by making the sources anonymous, the authors of the fake news pieces avoided being penalised for committing hate crimes. The credibility authors sought to give the pieces was found in the format of the images. Screenshots mostly include documents and posters that were originally true, with additional text that takes the image out of context or deceives, objectives that the research by Salaverría et al. (2020) classified as very serious if they are also accompanied by the negative intentions of whoever creates and shares them.

To this effect, there is a mechanism that we could call "fake news personalisation": users add text to images to take them out of context or deceive, thus creating multimodal disinformation about otherness. Although digital technologies provide tools for content manipulation (Tandoc et al., 2018),

this study finds that those who create and share multimodal disinformation about otherness are extremely unprofessional. Far from intending to provide credibility and bring opinions together, their unprofessional manner reveals their intention to embolden similar positions, providing deceptive reasons to communities whose opinions are already formed (us) in an environment where confirmation bias (Wason, 1960) and cognitive dissonance (Festinger, 1962) act to reaffirm previous opinions and prejudices (López-López et al., 2020; Salaverría et al., 2020). Therefore, this study shows that you do not have to have a subtle design to deceive – as you do with deepfakes, which require advanced technological skills (Paris and Donovan, 2019) – as proven by the popularity of cheapfakes, which are used to generate coarse and crude disinformation multimodal content to promote hate speech (Sanguinetti et al., 2018).

This study has some limitations. The first is that only the fake news pieces discredited by the four Spanish media entities accredited by the International Fact-Checking Network have been included in the sample. Despite this, we believe that the various criteria for collecting disinformation on these platforms, as well as the numerous analysis units studied, have provided the sample with sufficient representativeness. Second, the texts included in the disinformation materials have been analysed, serving as an anchor, but – given the impossibility of compiling them – none of the accompanying texts posted by users has been analysed. However, we believe that the present study – which has analysed both the texts contained in the disinformation materials and the multimedia materials associated with them, especially the images, and the relationships between both elements – has taken a fully multimodal approach. A more in-depth analysis of the discursive construction of the additional texts (Fernández et al., 2020) remains for future research, as well as an analysis of the audiovisual manipulation elements used in videos.

## Bibliographical references

CALVO, E. and ARUGUETE, N. (2018). "#Tarifazo. Medios tradicionales y fusión de agenda en redes sociales". *Intermediaciones de la comunicación*, 13 (1), 189-213.
<https://doi.org/10.18861/ic.2018.13.1.2831>

CORNER, J. (2017). "Fake news, post-truth and media–political change". *Media, Culture & Society*, 39 (7), 1100-1107.
<https://doi.org/10.1177/0163443717726743>

DEL-FRESNO-GARCÍA, M. (2019). "Desórdenes informativos: sobreexpuestos e infrainformados en la era de la posverdad". *El profesional de la información*, 28 (3), 1-11.
<https://doi.org/10.3145/epi.2019.may.02>

ENGESSER, S.; ERNST, N.; ESSER, F. and BÜCHEL, F. (2017a). "Populism and social media: How politicians spread a fragmented ideology". *Information, Communication & Society*, 20 (8), 1109-1126.
<https://doi.org/10.1080/1369118X.2016.1207697>

ENGESSER, S.; FAWZI, N. and LARRSON, A. (2017b). "Populist online communication: introduction to the special issue". *Information, Communication & Society*, 20 (9), 1279-1292
<https://doi.org/10.1080/1369118X.2017.1328525>

FERNÁNDEZ FERNÁNDEZ, A.; REVILLA GUIJARRO, A. and ANDALUZ ANTÓN, L. (2020). "Análisis de la caracterización discursiva de los relatos migratorios en Twitter. El caso Aquarius". *Revista Latina de Comunicación Social*, 77, 1-18.
<https://doi.org/10.4185/RLCS-2020-1446>

FESTINGER, L. (1962). "Cognitive Dissonance". *Scientific American*, 207 (4), 93-107.
<https://doi.org/10.1038/scientificamerican1062-93>

GUALLAR, J.; CODINA, L.; FREIXA, P. and PÉREZ-MONTORO, M. (2020). "Desinformación, bulos, curación y verificación. Revisión de estudios en Iberoamérica 2017-2020". *Telos: Revista de Estudios Interdisciplinarios en Ciencias Sociales*, 22 (3), 595-613.
<https://www.doi.org/10.36390/telos223.09>

IGARTUA, J. J. (2006). *Métodos cuantitativos de investigación en comunicación.* Barcelona: Bosch.

IGARTUA, J. J.; BARRIOS, I. and ORTEGA, F. (2012). "Analysis of the Image of Immigration in Prime Time Television Fiction". *Comunicación y sociedad*, 25 (2), 5-28.

IGARTUA, J. J. and CHENG, L. (2009). "Moderating effect of group cue while processing news on immigration. Is framing effect a heuristic process?". *Journal of Communication*, 59 (4), 726-749.
<https://doi.org/10.1111/j.1460-2466.2009.01454.x>

IGARTUA, J. J.; CHENG, L.; MORAL, F.; FERNÁNDEZ, I.; FRUTOS, F. J.; GÓMEZ ISLA, J. and OTERO, J. A. (2008). "Encuadrar la inmigración en las noticias y sus efectos socio-cognitivos". *Palabra Clave*, 1 (1), 87-107.

IGARTUA, J. J.; MORAL-TORANZO, F. and FERNÁNDEZ, I. (2011). "Cognitive, attitudinal and emotional effects of the news frame and group cues on processing news about immigration". *Journal of Media Psychology*, 23, 174-185.
<https://doi.org/10.1027/1864-1105/a000050>

IGARTUA, J. J.; MUÑIZ, C. and CHENG, L. (2005). "La inmigración en la prensa española. Aportaciones empíricas y metodológicas desde la teoría del encuadre noticioso". *Migraciones*, 17, 143-181.

IGARTUA, J. J.; MUÑIZ, C.; OTERO, J. A. and DE LA FUENTE, M. (2007). "El tratamiento informativo de la inmigración en los medios de comunicación españoles. Un análisis de contenido desde la Teoría del Framing". *Estudios sobre el Mensaje Periodístico*, 13, 91-110.

JIN, Z.; CAO, J.; GUO, H.; ZHANG, Y. and LUO, J. (2017). "Multimodal fusion with recurrent neural networks for rumor detection on microblogs". In *Proceedings of the 2017 ACM on Multimedia Conference. ACM*, 795-816

KRIPPENDORFF, K. (1990). *Metodología de análisis de contenido.* Barcelona: Paidós.

LACALLE, C. (2008). *El discurso televisivo sobre la inmigración. Ficción y construcción de identidad.* Barcelona: Ediciones Omega.

LÓPEZ-LÓPEZ, P. C.; CASTRO MARTÍNEZ, P. and OÑATE, P. (2020). "Agenda melding y teorías de la comunicación: la construcción de la imagen de los actores políticos en las redes sociales". *Cuadernos del Centro de Estudios en Diseño y Comunicación*, 112, 21-34.

MARCOS RAMOS, M.; IGARTÚA PEROSANZ, J. J.; FRUTOS ESTEBAN, F. J.; BARRIOS VICENTE, I. M.; ORTEGA MOHEDANO, F. and PIÑEIRO NAVAL, V. (2014). "La representación de los personajes inmigrantes en los programas de ficción". *Vivat Academia*, 127, 43-71.
<https://doi.org/10.15178/va.2014.127.43-71>

MOLINA-CAÑABATE, J. P. and MAGALLÓN ROSA, R. (2019). "Procedimientos para verificar y desmontar informaciones falsas basadas en el discurso del odio. El caso de Maldita Migración". *RAEIC, Revista de la Asociación Española de Investigación de la Comunicación*, 6 (12), 95-122.
<https://doi.org/10.24137/raeic.6.12.5>

MOURÃO, R. R. and ROBERTSON, C. T. (2019). "Fake News as Discursive Integration: An Analysis of Sites That Publish False, Misleading, Hyperpartisan and Sensational Information". *Journalism Studies*, 20 (14), 2077-2095.
<https://doi.org/10.1080/1461670X.2019.1566871>

MUÑIZ, C.; IGARTÚA, J. J. and OTERO, J. A. (2006). "Imágenes de la inmigración a través de la fotografía de prensa. Un análisis de contenido". *Comunicación y Sociedad*, 19 (1), 103-128. Retrieved from <https://hdl.handle.net/10171/8302>.

MUÑIZ, C.; IGARTÚA, J. J.; DE LA FUENTE JUAN, M. and OTERO PARRA, J. A. (2008). "Imágenes periodísticas de la inmigración: aportaciones metodológicas al estudio de la comunicación visual". *Anàlisi: Quaderns de Comunicació i Cultura*, 37, 31-48.

NEWMAN, N.; FLETCHER, R.; SCHULZ, A.; ANDI, S. and KLEIS NIELSEN, R. (2020). *Reuters Institute Digital News Report 2020* [online]. [Date consulted: 03/01/2020] Retrieved from <https://www.digitalnewsreport.org/>.

PARIS, B. and DONOVAN, J. (2019). *Deepfakes and cheapfakes. The Manipulation of Audio and Visual Evidence* [online]. [Date consulted: 03/01/2020] Retrieved from <https://datasociety.net/library/deepfakes-and-cheap-fakes/>.

REINEMANN, C.; AALBERG, T.; ESSER, F.; STRÖMBACK, J. and DE VREESE, C. H. (2017). "Populist political communication: Toward a model of its causes, forms, and effects". In: AALBERG, T.; ESSER, F.; REINEMANN, C.; STRÖMBACK, J. and DE VREESE, C. H (Eds.). *Populist political communication in Europe.* London: Routledge, 12-25.

RUIZ-COLLANTES, X.; FERRÉS, J.; OBRADORS, M.; PUJADAS, E. and PÉREZ, O. (2006). "La imagen pública de la inmigración en las series de televisión españolas". *Política y cultura*, 26, 93-108.

SALAVERRÍA, R.; BUSLÓN, N.; LÓPEZ-PAN, F.; LEÓN, B.; LÓPEZ-GOÑI, I. and ERVITI, M.-C. (2020). "Desinformación en tiempos de pandemia: tipología de los bulos sobre la Covid-19". *El profesional de la información*, 29 (3), 1-15.
<https://doi.org/10.3145/epi.2020.may.15>

SANGUINETTI, M.; POLETTO, F.; BOSCO, C.; PATTI, V. and STRANISCI, M. (2018). "An Italian Twitter Corpus of Hate Speech against Immigrants". In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018).

SCHICK, N. (2020). "Don't underestimate the cheapfake". *MIT Technology Review* [online]. [Date consulted: 07/01/20] Retrieved from <https://www.technology review.com/2020/12/22/1015442/cheapfakes-more-political-damage-2020-election-than-deepfakes/>.

TANDOC, E. C.; LIM, Z. W. and LING, R. (2018). "Defining 'fake news'. A typology of scholarly definitions". *Digital Journalism*, 6 (2), 137-153.
<https://doi.org/10.1080/21670811.2017.1360143>

TARULLO R. and AMADO, A. (2020). "Brasil II: Bolsonaro: sus verdades y mentiras en la elección presidencial brasileña de 2018". In: *Fakecracia*. Buenos Aires: Biblos, 185-204.

TUCKER, J. A.; GUESS, A.; BARBERÁ, P.; VACCARI, C.; SIEGEL, A.; SANOVICH, S.; STUKAL, D. and NYHAN, B. (2018). "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature". *SSRN*.
<https://doi.org/10.2139/ssrn.3144139>

VOROBYEVA, T.; MOURATIDIS, K.; DIAMANTOPOULOS, F.-N.; GIANNOPOULOS, P.; TAVLARIDOU, K.; TIMAMOPOULOS, C.; PERISTERAS, V.; MAGNISALIS, I. and SHAH, S. I. H. (2020). "A fake news classification framework: application on inmigration cases". *Communication Today*, 11 (2), 118-130.

WAISBORD, S. (2018). "Truth is What Happens to News. On journalism, fake news, and post-truth". *Journalism Studies*, 19 (13), 1866-1878.
<https://doi.org/10.1080/1461670X.2018.1492881>

WASON, P. C. (1960). "On the failure to eliminate hypotheses in a conceptual task". *Quarterly Journal of Experimental Psychology*, 12 (3), 129-140.
<https://doi.org/10.1080/17470216008416717>

ZAFARANI, R.; ZHOU, X.; SHU, K. and LIU, H. (2019). "Fake News Research: Theories, Detection Strategies, and Open Problems". In *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3207-3208.
<https://doi.org/10.1145/3292500.3332287>

ZHOU, X.; WU, J. and ZAFARANI, R. (2020). "SAFE: Similarity-aware multi-modal fake news detection". In *Proceedings of the 24th Pacific-Asia Conference on Knowledge Discovery and Data Mining*.

ZHUKOVA, E. (2019). "Image substitutes and visual fake history: historical images of atrocity of the Ukrainian famine 1932-1933 on social media". *Visual Communication*.
<https://doi.org/10.1177/1470357219888673>

# Beyond challenges and viral dance moves: TikTok as a vehicle for disinformation and fact-checking in Spain, Portugal, Brazil, and the USA

Nadia Alonso-López
Universitat Politècnica de València
naallo1@har.upv.es

Pavel Sidorenko-Bautista
Universidad Francisco de Vitoria
pavel.sidorenko@ufv.es

Fábio Giacomelli
Universidad da Beira Interior
fabio@ubi.pt

## Abstract

This paper analyses the narrative of disinformation disseminated through the social network TikTok, a network which is popular at a global level and whose users are mainly young or very young. To do so, a study was carried out on the content of publications on TikTok in four countries with different idiosyncrasies and national realities: Spain, Portugal, Brazil, and the United States. Interviews were also conducted with fact-checking agencies on the potential for misinformation and fact-checking potential on this social network. The results suggest that due to its characteristics as a fresh, visual network with easily shareable and viral content, TikTok is a network that facilitates the spread of disinformation, but which in turn is a tool for debunking hoaxes beyond the range of action of conventional media.

**Keywords:** disinformation; TikTok; fake news; fact-checking

**Resum.** *Més enllà de reptes i balls virals: TikTok com a vehicle per al discurs desinformatiu i la verificació d'informació a Espanya, Portugal, el Brasil i els Estats Units*

Aquesta recerca analitza la narrativa de la desinformació a través de la xarxa social TikTok, una xarxa que és tendència a escala global i els usuaris majoritaris de la qual són joves o molt joves. Per dur-ho a terme es fa un estudi dels continguts de publicacions a TikTok de quatre països amb idiosincràsies i realitats nacionals diferents com són Espanya, Portugal, el Brasil i els Estats Units. Així mateix, es fan entrevistes amb agències de verificació de fets sobre el potencial desinformatiu i la verificació de la informació en aquesta xarxa social. Els resultats apunten al fet que, per les seves característiques de xarxa visual, fresca i amb continguts fàcilment compartibles i viralitzables, TikTok és una xarxa que facilita l'expansió de la desinformació, però, al seu torn, és una eina per desmentir boles fora del rang d'acció dels mitjans convencionals.

**Paraules clau:** desinformació; TikTok; notícies falses; mecanismes de verificació

**Resumen.** *Más allá de retos y bailes virales: TikTok como vehículo para el discurso desinformativo y la verificación de información en España, Portugal, Brasil y Estados Unidos*

Esta investigación analiza la narrativa de la desinformación a través de la red social TikTok, una red que es tendencia a escala global y cuyos usuarios mayoritarios son jóvenes o muy jóvenes. Para ello se realiza un estudio de los contenidos de publicaciones en TikTok de cuatro países con idiosincrasias y realidades nacionales diferentes como son España, Portugal, Brasil y Estados Unidos. Asimismo, se realizan entrevistas con agencias de verificación de hechos sobre el potencial desinformativo y la verificación de información en esta red social. Los resultados apuntan a que, por sus características de red visual, fresca y con contenidos fácilmente compartibles y viralizables, TikTok es una red que facilita la expansión de la desinformación, pero, a su vez, es una herramienta para desmentir bulos fuera del rango de acción de medios convencionales.

**Palabras clave:** desinformación; TikTok; noticias falsas; mecanismos de verificación

## 1. Introduction

### 1.1. The spread of disinformation

Disinformation is a multifaceted phenomenon with diverse aspects and different motives, such as spoofing, causing personal or corporate damage or influencing electoral processes (Guallar et al. 2020). People began to talk about fake news after the 2016 US presidential election campaign that resulted in the victory of Donald Trump (Allcott and Gentzkow, 2017; López, Vives and Badell, 2018), and this event, together with the Brexit process in the United Kingdom and the rise of Bolsonaro to the presidency of Brazil are considered among the most critical moments in terms of the spread of fake news (Cabezuelo and Manfredi, 2019; Da Silva and Dourado, 2019; Pérez, Meso and Mendiguren, 2020).

While disinformation represents a broader concept and is defined as false, inaccurate or misleading content that deliberately seeks to cause harm or benefit (European Commission, 2018), fake news refers to information with the

appearance of real news created with the aim of propagating a falsehood and misleading readers in order to obtain some kind of benefit, whether political or otherwise (Amorós, 2018). Salaverría et al (2020) defend the use of the term "hoax" to refer to false content created intentionally and disseminated massively through the network for the aforementioned reasons.

The spread of false content is not a recent phenomenon (Salas, 2019) and is related to the so-called Post-Truth Era (Keyes, 2004) characterised by social networks and the current hyperconnectivity in the interests of shorter, more immediate, instant and highly visual content (Mujika, García and Gibaja, 2020).

Users appropriate Internet content and produce their own messages based on it (Marzal and Casero, 2017). These users play a more active role, proposing multidisciplinary and multichannel dialogues while demanding new ways to express themselves (Li, Xiaohui and Zhengwu, 2019).

This scenario has fostered the spread of false content primarily through digital media, broadening the spectrum beyond the information disseminated by mass media (Ceron, De-Lima-Santos and Quiles, 2021; Salaverría, 2020; Sánchez and Magallón, 2020). These publications reinforce the beliefs of a community with a certain ideological bias, particularly in relation to politics, where dissemination is faster in comparison with other issues (Orbegozo, Morales-i-Gras and Larrondo, 2020), and are reaffirmed insofar as they refer to a reality that conforms to what people want to hear, rather than a reality based on evidence; i.e. so-called wishful thinking (Cassam, 2019; MacKenzie and Bhatt, 2020).

It is even possible to see today, for example, how through Facebook, YouTube, Twitter or more private social networks such as WhatsApp, democratic and electoral systems can be significantly undermined as a result of this (Bharali and Goswami, 2018; Grinberg et al, 2019).

This asymmetrical context and the increasing spread of false or speculative content has prompted journalists, news media and news agencies to create verification channels through which users can check whether a news item is true or false (Vázquez, Vizoso and López, 2019; Terol and Alonso, 2020; García-Ortega and García-Avilés, 2021), as well as different governments around the world to legislate in this regard (Meneses, 2018) and even social networks such as Twitter, Facebook and Snapchat themselves to apply mechanisms such as that implemented specifically at the beginning of 2021 with the digital veto against Donald Trump, while on other more recent platforms such as TikTok, for the time being, it is the algorithm that acts as arbiter and censor.

## 1.2. The social network TikTok

TikTok is a social network created from the foundations of Music.ly, a predecessor also owned by the Chinese company ByteDance, in which a message code has been developed around entertainment with large quantities of visual

stimuli, which involves a great deal of dynamism, speed of production, consumption of content and creative demand (Li et al., 2019; Shuai, Yuzhen and Yifang, 2019). In short, it relies on communication consisting of playbacks, funny stories and "mind-soothing" content (Yu-Liang, Chun-Chin and Shu-Ming, 2019).

It is a platform that has exceeded 800 million users worldwide, positioning itself as the most downloaded mobile app in the first quarter of 2020 (315 million installations worldwide) (Ditrendia, 2020). Its users are mostly 'Generation Z' (Rapkin, 2017; Shuai et al., 2019) although it is true that over the course of 2020 it reported a significant increase in Millennials and even 'Generation X' users as a consequence of the first strict lockdowns resulting from the COVID-19 pandemic.

Its nature as a digital trend has prompted broadcasters of all kinds, including the media and journalists (Sidorenko, Herranz and Cantero, 2020; Vázquez, Negreira and López, 2020) to innovate by quickly adapting to such a particular message code and heterogeneous audience.

However, it is not precisely the latter interlocutors who enjoy the greatest popularity and, therefore, the greatest audience. Since most consumers are still very young people, it can be considered a priori that the community of users of this social network is not always the best interlocutor for certain topics, which involves superficiality in some cases, or conversely, misrepresentation of certain data.

However, according to Van den Bremer and Siebelink (2020), TikTok users have a more critical view and participation in the content they consume through this platform only when they are genuinely interested in the subject, with a very common approach being to view the comment box to check it.

This is in line with Tan (2013) on YouTube usage habits, although they argue that this is still a superficial mechanism as there is no evidence of any attempt to check the information through other digital channels outside the social network. However, most of these users are concerned about their image and personal brand, so they often think carefully about what they are going to post, a practice that had already been emerging in young audience segments on other social networks (Vogel and Rose, 2016; Omar and Dequan, 2020).

TikTok has represented a major digital channel for the dissemination of information related, for example, to the COVID-19 pandemic (Basch, Hillyer and Jaimie, 2020), not only through the actions of official profiles such as that of the World Health Organisation, but also by individual users who, through specific hashtags, have allowed specialists to get a closer look at users' habits in relation to the pandemic, with the impressions and disinformation risks that any social network entails.

Likewise, as on other social networks, TikTok attempts to control the flow of speculative and biased videos (Ballesteros, 2020). However, as has been happening on Facebook, Twitter, Instagram and Snapchat, to refer to the social networks with the largest number of users worldwide, on TikTok

there is a significant flow of content that encourages disinformation unintentionally, but, above all, also intentionally.

Referring to more specific and recent contexts, during the US presidential elections in 2020, a large amount of content which was considered false went viral on this social network, which has been removed by the platform itself, as reported by The Guardian newspaper on 06/11/20 (see <http://bit.ly/3qprqQE>). There is even evidence of the creation of sensationalist media that also engages in disinformation, as in the case of the @notimundo profile (see <http://bit.ly/3oRzvNt>).

## 2. Methodology

Given that this is a global digital trend and that it largely involves young and very young audiences, the main aim of this paper is to determine whether there is inaccurate, speculative or false content on TikTok, and how it is constructed narratively, especially during the year 2020, a complex period as a consequence of COVID-19 and various global political episodes that have generated negative effects in the field of information, such as "infoxication" (López, 2020) due to an excess of content referring especially to the delicate health situation, as well as an "infodemic" (Papapicco, 2020) as a consequence of the flood – in the form of a parallel pandemic – of false news and related hoaxes.

Specifically, the questions posed in this regard are: What is the type of profile that creates hoaxes or misinforms through TikTok? How does the user community react to misinformation content on TikTok? Are they aware of it? What type of narrative describes the hoax or inaccurate content? Are there common or general characteristics?

Therefore, the initial hypotheses are:

H1:　TikTok is a platform prone to misinformation, as are other social networks with a history of misinformation.

H2:　The user community easily responds positively to disinformation content.

H3:　False and inaccurate posts come from fake profiles dedicated to disseminating disinformation.

H4:　Fake and inaccurate content gets high levels of engagement which gives it more visibility in the social network's algorithm.

For the purposes of the research, we propose an analysis of the content of publications on this social network in four countries (two European and two American) with different idiosyncrasies and national identities: Spain, Portugal, Brazil and the United States.

This is a preliminary and exploratory survey based on content selected by the researchers after an extensive review of hashtags with high levels of engagement in relation to current affairs or which expressly refer to disinfor-

mation: #vivaespaña (90.8 million views), #Trump2020 (18.48 thousand million views), #usa (31 thousand million views), #covidfake (424,800 views) #eleicoes2020 (211.6 million views), #bolsonaro (1.1 billion views), #antoniocosta (2.3 million views) and #marcelorebelodesousa (6.6 million views).

The particular case of #usa was taken into consideration due to the fact that many contents associated with other tags such as #Trump2020 or #covidfake also included it, which aroused suspicion in the content analysis.

The survey was carried out up to the 31st of December 2020, taking as a reference 10 profiles per country, for a total of 40 contents, listed in an assessment table by country with the following study variables: username, profile type, number of followers, 'likes' of the profile, narrative characteristics of the content, number of comments resulting from the content, number of times the content was shared, 'likes'. A brief description of each content will also be included under the results heading.

The selection of the posts was determined by the thematic proposal, rather than by any specific engagement indicator. This has thus allowed for a fairly wide range of possibilities that will reveal a variety of expressions of interaction between the user community and false or misleading content.

The types of user profile were classified as follows:

— (IU) individual user
— (PF) public figure
— NGO
— (PO) political organisation
— (FP) fake profile

In addition, the narrative features were established as:

— (S) selfie
— (UN) user does not appear
— (CH) challenge
— (MU) music
— (TX) text

In this respect, a selfie (S) refers to a video in which the person records him/herself with the front camera of the mobile device. User does not appear (UN) refers to accounts in which, from the profile picture to all content, the user remains anonymous and does not reveal his or her identity. Challenges (CH) refer to challenges and impersonations that are part of the intrinsic dynamics of the TikTok user community. Music (MU) is an important variable for the visualisation of the content by the platform's algorithm and is often used to emphasise an idea or image. Finally, text (TX) refers to any messages that reinforce or attempt to enforce or influence the understanding of specific images.

The measurement of the number of times the content examined was shared is limited only to the record provided by TikTok, which is partial data because this social network has the special feature of offering the possibility of sharing its content through any social network and digital platform, such as WhatsApp, Facebook or Instagram, which makes it impossible to carry out such measurements in each external medium, given the replicating effect that many of these have.

In order to obtain a professional assessment of this phenomenon and the special nature of the TikTok social network, 'fact-checkers' were consulted in the form of a structured interview about the potential work they carry out through this channel, publishing or monitoring the flow of content, their opinion on the informative qualities of the platform and the potential complexity of the platform, both in terms of its audience and the ease with which its content can be transferred to other social networks. Thus, after several attempts to contact the most representative verification media in the countries surveyed, Maldita.es (Spain) and Lupa (Brazil) finally participated, as well as Cotejo (Venezuela), the latter as an observer from a country which, although not part of the sample, is working in a national context in which there are many complaints about disinformation in the absence of impartial conventional media and a constant persecution of journalists.

The questions asked were about the reasons for having an active profile on TikTok, whether they consider a specific audience segment as part of this initiative, whether they monitor or have considered monitoring TikTok content, whether they consider TikTok to be a complex social network for the verification of fake news and hoaxes, and what they think about the flow of disinformation through this social network.

## 3. Results

There is no standard profile to define the typical "digital disinformer" on TikTok. In Table 5, based on the data obtained from Tables 1, 2, 3 and 4, it is possible to see that in Europe and America, at first glance, disinformation comes from individual users (IU), who typically post selfies (S) or simply videos from other digital platforms, where they are obviously unrecognisable (UN) (see Table 6). However, in any of the most common cases, texts are an important resource for reaffirming the biased idea that the user is trying to put across through multimedia content.

**Table 1.** Disinformation content produced in Spain on TikTok up to the 31st of December 2020

| No. | User | | Regarding the profile | | | Content: characteristics and engagement | | |
| | | Type | Followers | *'likes'* | Narrative | No. of comments | Shared (No. of times) | *'likes'* |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | @viva_espnaaa <https://vm.tiktok.com/ZSpD36H3/> | FP | 34200 | 505500 | UN | 1632 | 4543 | 76700 |
| 2 | @rafaadiaaz <https://vm.tiktok.com/ZSpD3dVg/> | IU | 385 | 8652 | UN-MU-TX | 621 | 116 | 7770 |
| 3 | @callemoslarepublica_1 <https://vm.tiktok.com/ZSpDnYWt/> | IU | 3655 | 93200 | S-CH-MU-TX | 562 | 162 | 7352 |
| 4 | @gerardoo.vera <https://vm.tiktok.com/ZSpUHPkP/> | IU | 377000 | 7.8 millions | S | 90 | 23 | 5467 |
| 5 | @gonzalovy <https://vm.tiktok.com/ZSpUSNy9/> | IU | 343 | 10100 | S-MU-TX | 714 | 219 | 3755 |
| 6 | @abascalpresidente1 <https://vm.tiktok.com/ZSp5XEBy/> | FP | 1661 | 45600 | UN | 108 | 196 | 1724 |
| 7 | @espanaviva <https://vm.tiktok.com/ZSp54NUE/> | FP | 4220 | 80900 | UN | 95 | 151 | 1432 |
| 8 | @fortjosu <https://vm.tiktok.com/ZSp53sHV/> | IU | 104000 | 826200 | UN-MU | 5922 | 32100 | 45200 |
| 9 | @anitaaa.98 <https://vm.tiktok.com/ZSp4LFmd/> | IU | 9025 | 211200 | MU-TX | 185 | 93 | 1033 |
| 10 | @xjuan_.dem <https://vm.tiktok.com/ZSp4LSqY/> | IU | 1305 | 20400 | S-MU-TX | 239 | 53 | 2863 |

Source: Own compilation.

Of the four countries surveyed, the second profile associated with disinformation content was the so-called "fake profile" (FP), such as cases 1, 6 and 7 (Table 1) 2, 6, 7, 8 and 9 (Table 2) 1, 3 and 7 (Table 3) and 3, 9 and 10 (Table 4), in which it can be seen that there is no specific publication theme, with their feeds being full of unconnected videos, where it is common to see humorous content interspersed with other political or "technical-scientific" content.

Cases such as number 1 in Table 1, whose description states that its feed consists of humour and politics, illustrates this situation. This particular feed has high levels of engagement, which results in greater promotion of the content by the algorithm.

**Table 2.** Disinformation content produced in Portugal on TikTok up to the 31st of December 2020

| No. | User | Type | Followers | 'likes' | Narrative | No. of comments | Shared (No. of times) | 'likes' |
|-----|------|------|-----------|---------|-----------|-----------------|-----------------------|---------|
| | | | Regarding the profile | | Content: characteristics and engagement | | | |
| 1 | @tiago…cenas <https://vm.tiktok.com/ZSsUHt6f/> | IU | 10500 | 1203000 | S | 23 | 1 | 105 |
| 2 | @teixi_cl0ud1 <https://vm.tiktok.com/ZSsyN2sB/> | FP | 12600 | 3799000 | UN-MU-TX | 157 | 119 | 3333 |
| 3 | @mafaldaatcosta <https://vm.tiktok.com/ZSsyNoRF/> | IU | 14300 | 140000 | S-MU-TX | 80 | 78 | 2173 |
| 4 | @portugualviral <https://vm.tiktok.com/ZSsUbxx2/> | PF | 85900 | 1500000 | UN-TX | 124 | 231 | 1132 |
| 5 | @portugalinhe <https://vm.tiktok.com/ZSsUExu7/> | PF | 44900 | 998600 | UN-TX | 0 | 6 | 267 |
| 6 | @ntvoficial <https://vm.tiktok.com/ZSsyBWvX/> | FP | 8060 | 62500 | UN | 5 | 10 | 114 |
| 7 | @memes_tuga69 <https://vm.tiktok.com/ZSsymuH5/> | FP | 10400 | 150900 | UN | 2 | 5 | 107 |
| 8 | @manellmiranda <https://vm.tiktok.com/ZSsySHqK/> | FP | 469 | 6817 | UN-TX | 34 | 217 | 4489 |
| 9 | @sabugaswow <https://vm.tiktok.com/ZSthU7dx/> | FP | 104 | 1145 | UN | 87 | 466 | 1088 |
| 10 | @fabiopakete <https://vm.tiktok.com/ZSsykfQd/> | IU | 4782 | 9601 | UN | 0 | 25 | 67 |

Source: Own compilation.

In Portugal and Brazil, disinformation content is more focused on politics, in some cases referring to the issue of the COVID-19 vaccination. There are even 'Deepfake' videos where the face of an individual is replaced, in this case, by that of the president of Portugal (cases 8 and 10 Table 2), attributing words or acts to him that have not really happened, misrepresenting his actions to very young, less informed audiences, either voluntarily or involuntarily, as Casero-Ripollés (2020) points out.

**Table 3.** Disinformation content produced in Brazil on TikTok up to the 31st of December 2020

| No. | User | Type | Followers | *'likes'* | Narrative | No. of comments | Shared (No. of times) | *'likes'* |
|-----|------|------|-----------|-----------|-----------|-----------------|-----------------------|-----------|
| | | **Regarding the profile** | | | **Content: characteristics and engagement** | | | |
| 1 | @cavaleiroandante <https://vm.tiktok.com/ZSsNYrBm/> | FP | 663 | 12500 | UN-TX | 305 | 3327 | 9169 |
| 2 | @germias946 <https://vm.tiktok.com/ZSsNBQsa/> | IU | 100400 | 1400000 | UN | 9102 | 213900 | 281700 |
| 3 | @cristao_puritano <https://vm.tiktok.com/ZSsNNHDU/> | FP | 17400 | 57700 | UN | 209 | 367 | 7063 |
| 4 | @fechadocombolsonaro2 <https://vm.tiktok.com/ZSsNeogp/> | IU | 18300 | 235100 | UN-TX | 236 | 767 | 5029 |
| 5 | @leandroaviolao <https://vm.tiktok.com/ZSsyCLJ3/> | IU | 142 | 4163 | S | 58 | 732 | 970 |
| 6 | @gustavodavi21 <https://vm.tiktok.com/ZSstWDLV8/> | IU | 2886 | 59900 | UN | 36 | 156 | 668 |
| 7 | @luiz_ricardo7 <https://vm.tiktok.com/ZSsUhVt4/> | FP | 9356 | 291900 | UN-TX | 394 | 1004 | 30900 |
| 8 | @pauloferreira <https://vm.tiktok.com/ZSsUgbJr/> | IU | 2237 | 20600 | UN-TX | 272 | 540 | 3223 |
| 9 | @planetjemil <https://vm.tiktok.com/ZStWp5xw> | IU | 5561 | 153700 | UN-TX | 2 | 3 | 92 |
| 10 | @leandroguerra95 <https://vm.tiktok.com/ZSsyjH9c/> | IU | 3746 | 86400 | UN-TX | 143 | 172 | 660 |

Source: Own compilation.

**Table 4.** Disinformation content produced in the United States on TikTok up to the 31st of December 2020

| No. | User | Type | Followers | 'likes' | Narrative | No. of comments | Shared (No. of times) | 'likes' |
|-----|------|------|-----------|---------|-----------|-----------------|-----------------------|---------|
| | | | Regarding the profile | | Content: characteristics and engagement | | | |
| 1 | @mel_d2ktk <https://vm.tiktok.com/ZSpQweno/> | IU | 1827 | 248400 | UN-TX | disabled | 6038 | 248000 |
| 2 | @toni_boloni <https://vm.tiktok.com/ZSpQWEJk/> | IU | 3910 | 114100 | UN-TX | 1343 | 5432 | 80200 |
| 3 | @aplhadavis0005 <https://vm.tiktok.com/ZSpC57vJ/> | FP | 8409 | 68800 | UN-TX | 3255 | 42900 | 68800 |
| 4 | @its_danzy <https://vm.tiktok.com/ZSpCYP9L/> | IU | 437700 | 19500 | S-TX | 310 | 110 | 10400 |
| 5 | @diarioexxtra <https://vm.tiktok.com/ZSpCDCrk/> | PF | 138000 | 1.4 millions | UN-TX | 2629 | 3615 | 198900 |
| 6 | @ashleyblake128 <https://vm.tiktok.com/ZSpCrxXQ/> | IU | 30500 | 166500 | S-TX | 1162 | 2592 | 191000 |
| 7 | @alexnino_ <https://vm.tiktok.com/ZSpCm4Rg/> | IU | 302800 | 9.9 millions | S-TX | 495 | 1787 | 18200 |
| 8 | @fonzieethenolaguy <https://vm.tiktok.com/ZSpCEmsL/> | IU | 572 | 4313 | UN-TX | 174 | 341 | 1676 |
| 9 | @wakeup0666 <https://vm.tiktok.com/ZSpXJvyx/> | FP | 1642 | 3566 | UN-MU-TX | 8 | 15 | 106 |
| 10 | @2reddotreaxts <https://vm.tiktok.com/ZSpX1caq/> | FP | 551 | 2913 | UN-MU-TX | 6 | 15 | 97 |

Source: Own compilation.

Perhaps the most sensitive cases have been seen in the United States (Table 4), where politics is mixed – often in a forced manner – with the health situation, with conspiracy theories of various kinds or with the supposed existence of superior races that are the ones who really manage the future of the planet, as in the case of the 'reptilian' theory (case 5 Table 4).

The first content in Table 4 is noteworthy in that it refers to a particular situation of racial hatred, with comments disabled, which has prevented the user community from expressing themselves accordingly. The video shows alleged footage from a security camera at the entrance to a private home, where a delivery man apparently exclaims that the person he is about to deliver to is a racist, so he decides to throw the package at the door.

The user who posted this video indicates with complementary text that the subject's reaction is due to the pro-Donald Trump flags in the person's garden (which are not visible in the footage).

Although the video may or may not be authentic, it is difficult to determine whether the voice that can be heard is really that of the delivery person in question or a fabrication. Also, as the flags in question are not visible, it could be a video taken out of context and biased by the accompanying text.

**Table 5.** Type of disinformation user profile by country (review up to the 31st of December 2020)

| Country | IU | PF | NGO | PO | FP |
|---|---|---|---|---|---|
| Spain | 7 | 0 | 0 | 0 | 3 |
| Portugal | 3 | 2 | 0 | 0 | 5 |
| Brazil | 7 | 0 | 0 | 0 | 3 |
| USA | 6 | 1 | 0 | 0 | 3 |
| Total | 24 | 3 | 0 | 0 | 13 |

Source: Own compilation.

In the sample surveyed, there was no evidence of disinformation content by political parties or non-governmental organisations, although some videos from parties such as VOX in Spain, or actions carried out through other digital platforms have led to the creation of fake accounts or the potential radicalisation of some users, who publish as if they were intermediaries of these ideas, which are ultimately false and end up fuelling hatred and social antagonism, as can be seen in the comments of cases 1 (Table 1), 7 (Table 3) or 5 (Table 4), for example.

**Table 6.** Most used type of narrative by disinformation profile (review up to the 31st of December 2020)

| Profile type | *Selfie* (S) | User does not appear (UN) | *Challenge* (CH) | Music (MU) | Text (TX) |
|---|---|---|---|---|---|
| IU | 9 | 12 | 1 | 7 | 16 |
| PF | 0 | 3 | 0 | 0 | 3 |
| NGO | 0 | 0 | 0 | 0 | 0 |
| PO | 0 | 0 | 0 | 0 | 0 |
| FP | 0 | 14 | 0 | 3 | 7 |
| Total | 9 | 29 | 1 | 10 | 26 |

Source: Own compilation.

According to the data shown in Table 6, the narrative practice most used in content classified as disinformation on TikTok is the publication of a video fragment taken out of context and coming from another digital platform, accompanied by some text that influences the observation and interpretation of the content by the user community.

## 4. Discussion and conclusions

From the sampling carried out, the content of the four countries analysed can be considered to fulfil the first hypothesis put forward, since it is indeed easy to spread disinformation through TikTok, with "individual users" who are not related – at least not publicly – to specific political parties or organisations, using generic hashtags on most occasions and challenges unrelated to the proposed content, but with high popularity rates.

The attempt to use high-visibility hashtags to try to increase the reach of content is noteworthy, although this has begun to be penalised by the algorithm, as stated by TikTok Spain's own staff (see <http://bit.ly/2LZgbPD>).

Individual profiles are the most likely to create this type of false content, which confirms the third hypothesis, since the false or inaccurate information comes from fake profiles dedicated to spreading disinformation, but, on the other hand, with a significant presence of individual users who also publish other types of content, leading other people to give them greater credibility.

It is worth noting that users do not always react positively to these posts, this being evident on occasions involving content with a political and health-related tone, where the user community has reacted negatively. Although this does not necessarily prevent them spreading and going viral, it does generate contradiction with them, as in cases 9 (Table 1) 1, 5, 7 and 10 (Table 2) 9 (Table 3) and, 9 and 10 (Table 4), and this does not allow the second hypothesis to be confirmed.

It is necessary to distinguish between users who show a clear political tendency, in some cases bordering on fanaticism, and express this in their posts, and disinformation content of a political, health, environmental, etc. nature, whose discourse may be more effective as the users themselves are not aware that it is fake content, and the spread of disinformation is not possible if there are no recipients to confirm and spread this fake content.

With regard to the fourth hypothesis on engagement and the operability of the TikTok algorithm, it was observed that fake content as such is not always the type of content that obtains the highest visibility. However, the algorithm is indeed conditioned by posts in which the fake content is presented in a humorous way, even if it deals with political issues.

A noteworthy example of this is case 1 (Table 1) entitled "Pedro Sánchez resignation", which is not satirical content, but the sender is a profile with high levels of engagement in general terms. The large number of followers it has (34,200) gives it a large audience, which broadens the possibilities of exposure, as these recipients are then free to extract the content from the platform and divert it to other digital channels according to their interests. The strategy of interspersing biased political content with other humorous non-political posts has allowed the profile to engage an audience that is large in number and potentially receptive to its message.

**Figure 1.** Video of political interest on TikTok profile defined as satirical (Video 1, Table 1)



Source: TikTok profile of @viva_espnaaa.

While in Brazil the disinformation and fake content refers to the context of politics and health (political criticism is not directed at the national government but rather at regional and local authorities), in Portugal it is more difficult to determine the specific reasons for spreading disinformation, although a tendency towards irony and humour is detected in the dissemination of fake content, also with a special emphasis on politics and, above all, on the figure of the president of the government and the prime minister.

**Figure 2.** Video from Brazil in which a local government is accused of mismanagement of public roads, something which is later refuted as it is neither the place nor the work in question (Video 7, Table 3)



Source: TikTok profile of @luiz_ricardo7.

However, this is not – for the moment – a scenario of serious political polarisation, although it is true that in 2020 there has been greater media exposure of André Ventura, representative of Chega, a far-right party, the only one in this country to have a profile on TikTok, but still with very low levels of engagement. Similarly, it cannot be ruled out that in time this scenario will change, and political content will increase and become more relevant on TikTok.

Conversely, in Spain, growing political polarisation does have a more visible impact than other topics on fake content and disinformation work on this social network. Most of the cases reviewed here (9 out of 10) refer to issues that have to do with discourse against the government, political representatives or the parties themselves, both in their economic and health management, specifically in relation to COVID-19.

Example 9 (Table 1) is even more striking because the person who publishes it is a girl who engages in posting the State budgets applied during the mandate of the different government presidents of the democratic era. Her final argument is that the socialist Pedro Sánchez bases his administration on the budgets of the previous conservative-led administration, alleging the socialist's inability to approve new national budgets. She reinforces this explanation with the phrase "STOP FAKE NEWS", although this may seem a contradiction (see Figures 3 and 4).

**Figures 3 and 4.** Political TikTok post with handling of data without source references (Video 9, Table 1)
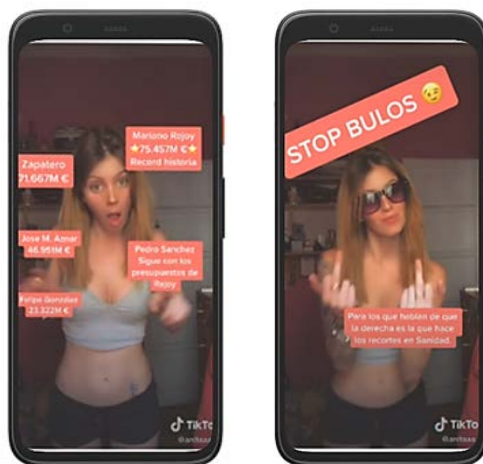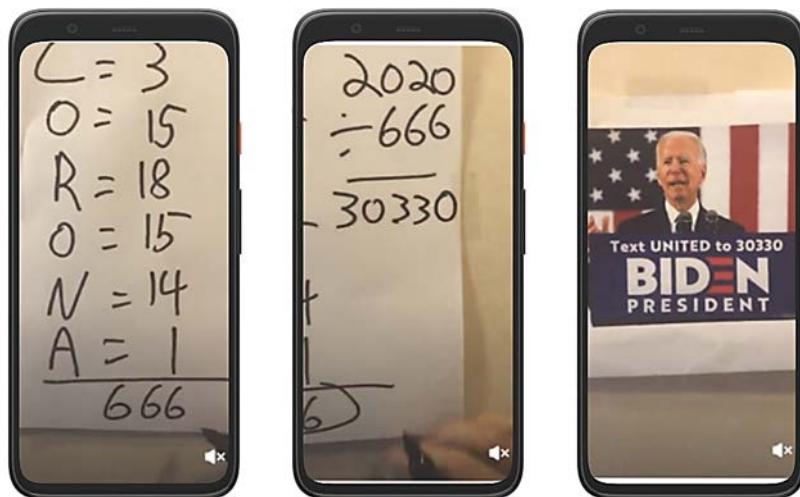


Source: TikTok profile of @anitaa.98.

Finally, in the United States, the 2020 presidential elections led to one of the most complex scenarios in terms of the flow of politically charged disinformation content on TikTok, which in many cases inevitably became inter-

twined with the health situation resulting from COVID-19, as Pablo Hernández of Maldita.es states, and as can be seen in the cases listed in Table 4.

This disinformation scenario becomes more complex when hybridised with messages that support conspiracy theories of various kinds, as in cases 3, 4 and 5 in Table 4, for example.

**Figures 5, 6 and 7.** Political TikTok post linking the Covid-19 virus to the number of the Beast and Joe Biden's US 2020 campaign (Video 3, Table 4)



Source: TikTok profile of @aplhadavis0005.

As Table 6 shows, at present, disinformation on TikTok is narratively determined by individual users, most of them fake profiles, who insert videos and photos from other digital platforms, taken out of context and reinforced with biased texts.

In view of this situation, several journalists and specialists have now seen the need to apply verification criteria and participate in TikTok. Thus, Agencia Lupa (Brazil) and Maldita.es (Spain) justify their profiles on this channel as being a consequence of the rapid expansion of the social network during the first strict lockdowns imposed as a result of COVID-19, due to the fact that it is a "network used mainly by young people who may be more vulnerable to fake news and because it is a trending platform".

Maldita.es has been monitoring the social network since 2019 and generates educational content because, according to Pablo Hernández (Maldita.es), "it is necessary to act by creating TikTok format denials and providing truthful information through the social network itself". Lupa, meanwhile, tends to intervene on the basis of fake content sent to them by users themselves. Both fact-checkers consider it important to create content in the message code that the platform itself imposes, in order to achieve a better reach.

From Venezuela, Karla Torres de Cotejo warns that although TikTok is a very popular social network among young people, this does not imply that it is a medium prone to a greater flow of disinformation and fake content. TikTok is not a platform where users go specifically to seek information – quite the contrary. However, on the way through the main feed, they do come across information and, therefore, also disinformation and fake content.

The content consumption proposed in TikTok is not designed to promote profiles but specific content. Users construct their "news" and "reality" with what the algorithm proposes in the feed, which is not necessarily made up of the accounts followed. This algorithm volatilises a large amount of content, which is likely to escalate rapidly, making the platform vulnerable to disinformation, which is currently incipient, but which, as it continues to grow, could become a more serious problem.

According to the latest State of Mobile Devices 2021 report from the App Annie consultancy (see <http://bit.ly/3nNrTdv>), people tend to spend more time on TikTok (20 hours in 2020) than on Facebook (16 hours in 2020), and more time connected to mobile devices than to conventional TV: in the US almost 4 hours per day and in Brazil almost 5 hours per day, for example.

Curiously, in contrast, due to its status as a trending network, the current convergence of so many audience segments and the fact that it is a platform where the tone of the message is fresh and dynamic, it seems that TikTok is seeking to gradually and organically establish itself as a more expeditious and accessible way to disprove fake news outside the range of action of conventional media, as in the case of the report of the "fake snow" in Madrid in January 2021 (see <http://bit.ly/39EoLvC>), even providing the possibility of the emergence of new journalistic and/or information proposals such as @holajulen <http://bit.ly/2XNyKsM>, @veronicafumanal <http://bit.ly/3sxMeHx> and @ac2alityespanol <http://bit.ly/3sx60mb>.

This paper has attempted to advance the differentiation of the potential disinformation narrative on TikTok with regard to other social digital platforms, with the aim of enriching the study of this type of statement and thus adding a new resource to the respective analysis in order to mitigate the scourge of lies and speculation in today's digital society.

## Bibliographical references

ALLCOTT, H. and GENTZKOW, M. (2017). "Social media and fake news in the 2016 election". *Journal of Economic Perspectives*, 31 (2), 211-236.
<https://doi.org/10.1257/jep.31.2.211>

AMORÓS GARCIA, M. (2018). *Fake News. La verdad de las noticias falsas*. Barcelona: Plataforma Editorial.

BALLESTEROS-HERENCIA, C.A. (2020). "La propagación digital del coronavirus: Midiendo el *engagement* del entretenimiento en la red social emergente TikTok". *Revista Española de Comunicación en Salud*, suplemento 1, 171-185.
<https://doi.org/10.20318/recs.2020.5459>

BHARALI, B. and GOSWAMI, A. L. (2018). "Fake News: credibility, cultivation syndrome and the New Age Media". *Media Watch*, 9 (1), 118-130.
<https://doi.org/10.15655/mw/2018/v9i1/49277>

BASCH, C. H.; HILLYER, G. C. and JAIMIE, C. (2020). "COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages". *International Journal of Adolescent Medicine and Health*, 1-3.
<https://doi.org/10.1515/ijamh-2020-0111>

CABEZUELO, L. F. and MANFREDI, J. L. (2019). "Posverdad, *fake-news* y agenda política en el discurso de Trump en Twitter". *Historia y comunicación social*, 24 (2), 449-483.

CASERO-RIPOLLÉS, A. (2020). "Impacto del Covid-19 en el sistema de medios. Consecuencias comunicativas y democráticas del consumo de noticias durante el brote". *Profesional de la Información*, 29 (2).
<https://doi.org/d8kd>

CASSAM, Q. (2019). *Cynicism, epistemic insouciance, and Boris Johnson* [online]. The Philosophers' Magazine. [Date consulted: 12/19/20]. Retrieved from <https://bit.ly/2QGyoUS>.

CERON, W.; DE-LIMA-SANTOS, M.F. and QUILES, M. G. (2021). "Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content". *Online Social Networks and Media*, 21.
<https://doi.org/10.1016/j.osnem.2020.100116>

DA SILVA G., W. and DOURADO, T. (2019). "Fake News, un fenómeno de comunicação política entre jornalismo, política e democracia". *Estudos em Jornalismo e Midia*, 16 (2), 33-45.
<https://doi.org/10.5007/1984-6924.2019v16n2p33>

DITRENDIA (2020). *Móviles en España y en el Mundo 2020* [online]. Digital Marketing Trends. [Date consulted: 01/10/2021]. Retrieved from <https://bit.ly/3iY6iwV>.

EUROPEAN COMMISSION (2018). *A multi-dimensional approach to disinformation. Report of the independent High-Level Group on Fake News and Online-Disinformation. Directorate-General for Communication Networks, Content and Technology* [online]. European Union. [Date consulted: 01/10/2021]. Retrieved from <https://bit.ly/3nAz2j2>.

GARCÍA-ORTEGA, A. and GARCÍA-AVILÉS, J. A. (2020). "Uso del diseño lúdico para combatir la desinformación: los newsgames como herramienta para la concienciación sobre los bulos". *Icono 14*, 19 (1), 179-204.
<https://doi.org/10.7195/ri14.v19i1.1598>

GUALLAR, J.; CODINA, Ll.; FREIXA, P. and PÉREZ-MONTORO, M. (2020). "Desinformación, bulos, curación y verificación. Revisión de estudios en Iberoamérica 2017-2020". *Telos: Revista de Estudios Interdisciplinarios en Ciencias Sociales*, 22 (3), 595-613.
<https://doi.org/10.36390/telos223.09>

GRINBERG, N.; JOSEPH, H.; FIREDLAND, L.; SWIRE-THOMPSON, B. and LAZER, D. (2019). "Fake News on Twitter during the 2016 US presidential election". *Science*, 363 (6425), 374-378.
<https://doi.org/10.1126/science.aau2706>

KEYES, R. (2004). *The Post-truth Era: Dishonesty and deception in contemporary life*. New York: St. Martins Pr.

LI, X.; XIAOHUI, Y. and ZHENGWU, Z. (2019). "Research on the causes of the "TikTok" app becoming popular and the existing problems". *Journal of Advanced Management Science*, 7 (2), 59-63.
<https://doi.org/10.18178/joams.7.2.59-63>

López, E. (2020). "Las pandemias en el siglo XXI: un desafío para la salud pública". *Actualización en Nutrición*, 21 (1), 1-3.

López-Borrull, A.; Vives-Gràcia, J. and Badell, J.I. (2018). "Fake news, ¿amenaza u oportunidad para los profesionales de la información y la documentación?". *Profesional de la información*, 27 (6), 1346-1356.
<https://doi.org/10.3145/epi.2018.nov.17>

Mackenzie, A. and Bhatt, I. (2020). "Lies, Bullshit and Fake News". *Postdigit Sci Educ*, 2.
<https://doi.org/10.1007/s42438-019-00085-4>

Marzal, J. and Casero, A. (2017). "El discurso publicitario: núcleo de la comunicación transmedia". *AdComunica*, 14, 11-19.
<https://doi.org/10.6035/2174-0992.2017.14.1>

Meneses, J. P. (2018). "Sobre a necessidade de conceptualizar o fenómeno das fake news. *OBS\* Observatorio*, Special Issue, 37-53.
<https://doi.org/10.15847/obsOBS12520181376>

Mujika-Alberdi, A.; García-Arrizabalaga, I. and Gibaja-Martins, J. J. (2020). "Un estudio sobre hábitos de consumo de información en la era digital". *Anàlisi: Quaderns de Comunicació i Cultura*, núm. extraordinari 2020, 31-46.
<https://doi.org/10.5565/rev/analisi.3227>

Omar, B. and Dequan, W. (2020). "Watch, Share or Create: The Influence of Personality Traits and User Motivation on TikTok Mobile Video Usage". *International Journal of Interactive Mobile Technologies*, 14 (4), 121-137.
<https://doi.org/10.3991/ijim.v14i04.12429>

Orbegozo-Terradillos, J.; Morales-i-Gras, J. and Larrondo-Ureta, A. (2020). "Desinformación en redes sociales: ¿compartimentos estancos o espacios dialécticos? El caso Luther King, Quim Torra y El Confidencial". *Revista Mediterránea de Comunicación*, 11 (2), 55-69.
<https://www.doi.org/10.14198/MEDCOM2020.11.2.2>

Papapicco, C. (2020). "Informative Contagion: The coronavirus (COVID-19) in Italian journalism". *Online Journal of Communication and Media Technologies*, 10 (3), 1-12.
<https://doi.org/10.29333/ojcmt/7938>

Pérez-Dasilva, J.A.; Meso-Ayerdi, K. and Mendiguren-Galdospín, T. (2020). "Fake news y coronavirus: detección de los principales actores y tendencias a través del análisis de las conversaciones en Twitter". *Profesional de la información*, 29 (3).
<https://doi.org/10.3145/epi.2020.may.08>

Rapkin, M. (2017). "The social media platform that has Gen Z obsessed". [online]. *Wall Street Journal* [Date consulted: 12/20/20]. Retrieved from <https://on.wsj.com/3t05dcs>.

Salas, A. C. (2019). "La primera fake news de la historia". *Historia y comunicación social*, 24 (2), 411-431.
<https://dx.doi.org/10.5209/hics.66268>

Salaverría, R. (2020). *Claves para hacer frente a la desinformación*. [en línea] Biko Insights #4 [Date consulted: 01/10/21]. Retrieved from <https://bit.ly/3nq529q>.

Salaverría, R.; Buslón, N.; López-Pan, F.; León, B.; López-Goñi, I. and Erviti, Mª.C. (2020). "Desinformación en tiempos de pandemia: tipología de los bulos sobre la Covid-19". *Profesional de la información*, 29 (3).
<https://doi.org/10.3145/epi.2020.may.15>

SÁNCHEZ-DUARTE, J.M. and MAGALLÓN-ROSA, R. (2020). "Infodemia y COVID-19. Evolución y viralización de informaciones falsas en España". *Revista Española de Comunicación en Salud*, suplemento 1, 31-41.
<https://doi.org/10.20318/recs.2020.5417>

SIDORENKO, P.; HERRANZ, J. and CANTERO, J. (2020). "Use of New Narratives for COVID19 reporting: from 360º videos to ephemeral TikTok videos in online media2". *Trípodos*, 47 (1), 105-122.

SHUAI, Y.; YUZHEN, Z. and YIFANG, M. (2019). "Analysis of the reasons and development of short video application – Taking Tik Tok as an example". *9th International Conference on Information and Social Science* (ICISS 2019). Francis Academic Press, 340-343.
<https://doi.org/10.25236/iciss.2019.062>

TAN, E. (2013). "Informal learning on YouTube: Exploring digital literacy in independent online learning". *Learning, Media and Technology*, 38, 463-477.

TEROL-BOLINCHES, R. and ALONSO-LÓPEZ, N. (2020). "La prensa española en la Era de la Posverdad: el compromiso de la verificación de datos para combatir las *Fake News*". *Revista Prisma Social*, 31, 304-327.

VÁZQUEZ-HERRERO, J.; NEGREIRA-REY, M. C. and LÓPEZ-GARCÍA, X. (2020). "Let's dance the news! How the news media are adapting to the logic of TikTok". *Journalism*, 1-19.
<https://doi.org/10.1177%2F1464884920969092>

VÁZQUEZ-HERRERO, J.; VIZOSO, Á. and LÓPEZ-GARCÍA, X. (2019). "Innovación tecnológica y comunicativa para combatir la desinformación: 135 experiencias para un cambio de rumbo". *Profesional de la información*, 28 (3).
<https://doi.org/10.3145/epi.2019.may.01>

VAN DEN BREMER, A. J. M. and SIEBELINK, R. (2020). *Critical Thinking on TikTok*. Unpublished Thesis. Utrecht University, Utrech.

VOGEL, E. A. and ROSE, J. P. (2016). "Self-reflection and interpersonal connection: Making the most of self-presentation on social media". *Translational Issues in Psychological Science*, 2, 294-302.
<https://doi.org/10.1037/tps0000076>

YU-LIANG, F. Y. L.; CHUN-CHIN, C. and SHU-MING, W. (2019). "Evaluation of charm factors of short video user experience using FAHP – a case study of TikTok app". *IOP Conference Series: Materials Science and Engineering*, 688, 1-5.
<https://doi.org/10.1088/1757-899X/688/5/055068>

# Audiovisual verification in the evolution of television newsrooms: Al Jazeera and the transition from satellite to the cloud

Lola Bañon Castellón

Universitat de València. Theory of Languages Department
lola.banon@uv.es

## Abstract

With the spread of the digital sphere and the proliferation of images from indirect sources that can be accessed by systems and users, verification routines have become essential to ensure media corporations' credibility. The advances in artificial intelligence which allow automated fact-checking (AFC) initiatives to be created help detect falsehoods, but they do not eliminate the need for human intervention. On the contrary, information professionals are necessary, and their functions increasingly include procedures such as mediating in videos and images. This study analyses the evolution of verification routines in audiovisual journalism and how new techniques have influenced the perception of trustworthiness and the reorganization of the television newsroom by allowing content from outside the media's own newsroom. To do so, it combines a method that examines the main literature on verification processes and compares it with the procedure used by Al Jazeera. In this regard, an exploration was conducted out based on participant observation in this international TV channel via interaction with journalists and heads of the corporation. The results indicate that advances in verification procedures make it possible to introduce visual material from the social media into the corporation's common news topics contributing to the transition from the traditional newsroom to the cloud structure and the inclusion of new audiences. These changes affect journalistic routines in a profession which has no longer been in-person for some time, in which correspondents coexist with journalists working in virtual mobility, seeking and receiving images in and from the social media.

**Keywords:** verification; fact-checking; Al Jazeera; fake news; television; Newsroom

**Resum.** *La verificació audiovisual en el procés d'evolució de les redaccions televisives: Al Jazeera i el trànsit del satèl·lit a l'estructura de núvol*

Amb l'extensió de la digitalitat i la multiplicació de les imatges de procedència indirecta a les quals poden accedir sistemes i usuaris, les rutines de verificació han esdevingut imprescindibles per assegurar la credibilitat de les corporacions mediàtiques. Els avenços en la intel·ligència artificial en la conformació d'iniciatives de comprovació automàtica (*automated fact-checking*, AFC) faciliten la detecció de falsedats però no anul·len la intervenció humana. Els professionals de la informació, per contra, són necessaris i inclouen de manera creixent entre les seves funcions procediments com la mediació en el tractament de vídeos i imatges. Aquest treball analitza l'evolució de les rutines de verificació del periodisme audiovisual i com les seves noves tècniques, en permetre l'entrada de continguts produïts a l'exterior de la redacció del mitjà, han influït en la percepció del concepte de fiabilitat i en la reorganització de la redacció televisiva. S'ha efectuat una exploració basada en l'observació participant en la cadena Al Jazeera en interacció amb periodistes i responsables de la corporació. Els resultats apunten al fet que els avenços en els processos de verificació possibiliten la incorporació de materials visuals dels mitjans socials a l'agenda comuna temàtica de la corporació i contribueixen a l'evolució de la transició de la redacció tradicional a l'estructura en núvol i a la incorporació de noves audiències. Els canvis afecten les rutines periodístiques en un col·lectiu que fa temps que va deixar de ser estrictament presencial i on les corresponsalies coexisteixen amb periodistes en mobilitat virtual buscant i rebent imatges en i des de les xarxes socials.

**Paraules clau:** verificació; *fact-checking*; Al Jazeera; notícies falses; televisió; redacció

**Resumen.** *La verificación audiovisual en el proceso de evolución de las redacciones televisivas: Al Jazeera y el tránsito del satélite a la estructura de nube*

Con la extensión de la digitalidad y la multiplicación de las imágenes de procedencia indirecta a las que pueden acceder sistemas y usuarios, las rutinas de verificación se han hecho imprescindibles para asegurar la credibilidad de las corporaciones mediáticas. Los avances en la inteligencia artificial en la conformación de iniciativas de comprobación automática (*automated fact-checking*, AFC) facilitan la detección de falsedades pero no anulan la intervención humana. Los profesionales de la información, por el contrario, son necesarios e incluyen de forma creciente entre sus funciones procedimientos como la mediación en el tratamiento de vídeos e imágenes. Este trabajo analiza la evolución de las rutinas de verificación del periodismo audiovisual y cómo sus nuevas técnicas, al permitir la entrada de contenidos producidos en el exterior de la redacción del medio, han influido en la percepción del concepto de fiabilidad y en la reorganización de la redacción televisiva. Se ha efectuado una exploración basada en la observación participando en la cadena Al Jazeera en interacción con periodistas y responsables de la corporación. Los resultados apuntan a que los avances en los procesos de verificación posibilitan la incorporación de materiales visuales de los medios sociales a la agenda común temática de la corporación y contribuyen a la evolución de la transición de la redacción tradicional a la estructura en nube y a la incorporación de nuevas audiencias. Los cambios afectan a las rutinas periodísticas en un colectivo que hace tiempo dejó de ser estrictamente presencial y en donde las corresponsalías coexisten con periodistas en movilidad virtual buscando y recibiendo imágenes en y desde las redes sociales.

**Palabras clave:** verificación; *fact-checking*; Al Jazeera; noticias falsas; televisión; redacción

## 1. Introduction

Checking information has been basic practice in journalism since its inception. In just a few years, however, the spread of the social media has changed the dynamics of how the large television corporations work. The growing presence of images from the digital sphere has led the entire newsroom to undertake verification tasks that require some degree of technological training.

Every minute, 500 hours of video made by users from all corners of the globe using the cameras on their mobile phones are uploaded to the Internet (Mohsin, 2020). This shift poses a challenge for newsrooms, which have to ensure the veracity of this material in order to preserve their credibility.

The digital sphere poses a scenario in which the reputation of journalism has been brought into play, since not only have the traditional media lost sole sovereignty of the discourse but the social media themselves have become a source with no stable interlocution protocols; instead, there are only fortuitous exchanges which lack solid knowledge or trust. Hence, there is a growing interest in finding new technological verification resources, and the appearance of professional, specialized fact-checkers are at the root of this new reporting style (Graves, 2012: 1). Indeed, over the last decade, fact-checking platforms have been set up in over 50 countries, 90% of them since 2010 (Fernández-García, 2017).

Artificial intelligence and advances in creating automated fact-checking initiatives (AFC) have not yet rendered communication professionals expendable in this process, although if there is any suspicion they help streamline the detection of lies, online rumors and other kinds of disinformation (Graves, 2019).

## 2. Theoretical context: Disinformation and editing on the Internet

In this convergence of media industries, the digital sphere and telecommunications, information flows at a speed that would have been unimaginable only a few years ago (Thussu, 2007: 43). The Internet has turned newsrooms into permanently connected networks with uninterrupted information flows that have to be updated with little processing time given the demand for news production. This evolution has changed journalistic work (Castells, Van der Haak and Parks, 2012). Verification neutralizes disinformation only if it has not yet reached its peak circulation, because once it spreads the rebuttal may not reach the audience (Rodriguez-Fernández, 2019).

TV audiences' migration to the Internet has turned them into not only information consumers but also content broadcasters, thus generating what is called networking journalism in which professionals and amateurs work together to get a story (Jarvis, 2006) through collaboration, with the journalists acting increasingly as moderators (Duffy, 2011: 9).

In this news structure determined by digital platforms, there are converging possibilities for connectivity and innovation with the proliferation and saturation of news flowing at hyperspeed (Beckett, 2020).

Any technological advance has a potential distorting effect, and there is no information that cannot be turned into disinformation (Virilio, 1995). The Internet has facilitated an intentional change in meanings, which is why not only the volume of information but also the veracity of the content must be considered (Hernon, 1995).

In this scenario, the definition of fake news comes from the lack of correspondence between statements and facts, but if it is also done intentionally, it becomes what common sense identifies as propagandistic lies. Disinformation, on the other hand, is more difficult to detect (Del-Fresno-García, 2019). It is a phenomenon that is known for its intentionality, with false sources or no identification of the sources (Rodriguez Andrés, 2018), and an abuse of power to gain some benefit (Van Dijk, 2006).

Hypertextual media are multiplying their possibilities for becoming active subjects of disinformation in a scenario in which an unprecedented volume of images and content is being put into circulation on the Internet. Machines are capable of generating products that are journalistic in appearance for the purpose of disinformation, a polysemic phenomenon that mutates quickly (Manfredi and Ufarte, 2020: 65).

## 3. Method, research questions and hypotheses

The method used is based on a review of the scientific literature on the evolution of fact-checking, the strategies used to deal with disinformation (Roozenbeek and Van der Linden, 2019) and the preservation of the media's ethical principles in this communication context (McBride and Rosenstiel, 2013).

Consideration of verification is fundamental in today's digital journalism (Adair, 2019), and websites devoted to this purpose have been proven to be instruments that restore credibility (López Pan and Rodriguez Rodriguez, 2020) and work towards democratic construction in a period of changes (Amazeen, 2017), while acknowledging the difficulties posed by the discursive ambiguities (Lim, 2018).

We conducted the research in this context with a procedure based on scientific production via a systematic review that enabled us to provide a snapshot of the matter (Codina, 2017). We also applied participant observation using the Mixed Methods Review perspective, as well as interactions with stakeholders (Grant and Booth, 2009), namely the professionals in the Al Jazeera corporation who were consulted directly.

This work was carried out on visits to Doha, which came about because I was on the international jury at Al Jazeera's Documentary Festival from 2012 to 2017. I have also had interactions seeking updates in 2020 through my job as a journalist with Ahmad Ashour, the senior content editor in Al Jazeera's digital unit, and Taysir Alouni, one of the corporation's most veteran journalists and a witness to how the channel has evolved since it was founded. The contacts were initially in-person, but in recent months the pandemic has meant that the work has been done via videoconferencing. I have also sought

the professional opinion of Esraa Rabaya, the head of strategy and audience development at Al Jazeera. In this case, the communication and information exchange was conducted by email and telephone. The other professionals that contributed their points of view have preferred to maintain their anonymity. My aim is not to analyze specific cases of verification but to examine the little-studied matter of how audiovisual journalistic verification routines have evolved, as well as the new techniques used. By allowing content produced outside the medium's newsroom to be used, this content has influenced the perception of trustworthiness and the reorganization of the television newsroom.

In keeping with the literature exploration and especially my participant observation of the verification process, specifically in the case of Al Jazeera, the following research questions were posed:

Q1. Does the inclusion of material from digital sources in the media's agenda lead to changes in organization and hierarchies in a large audiovisual newsroom?

Q2. Do verification processes broaden the topics on news programs' agendas?

Q3. Are new forms and aesthetics being created for reporting with the participation of materials from social media users?

Q4. By permitting materials from the social media to be on the agenda, is the new verification a key resource in locating new audiences?

From these perspectives, the intention is to analyze the influence of verification processes on the journalistic routines of Al Jazeera's professionals and their impact on the different platforms that make up the audiovisual media conglomerate.

Hence, the following hypotheses were put forward:

H1. The new verification processes are essential in enabling the social media to be included into a global cloud newsroom model.

H2. The process of confirming the veracity of sources gives rise to new structures in which there is direct, continual mingling of communication professionals with engineers and IT experts within the newsroom, giving rise to new hierarchies in the decisions on the topics on the agenda.

H3. The increased accessibility to such diverse, plural materials and sources made possible by the social media increases the flow of images that may be broadcast, which makes it essential to use human supervision and algorithmic technology from artificial intelligence.

## 4. Interaction and evolution of Al Jazeera's verification routines

The television corporation Al Jazeera was founded in 1996 with the aim of providing a different view of the news for Arabic peoples (Miles, 2005: 234), and it quickly gained a wide audience in the region (El-Nawawy and Iskander, 2002). Five years later, its website in Arabic was set up (Satti, 2020). Its broadcasting and website in English were launched in 2008, chal-

lenging the West's media hegemony in the region and becoming a point of reference for modernization in audiovisuals and communication in Arabic (Zayani, 2005:1).

It began at a time when the land-based model of television distribution, which could easily be subject to government intervention, was evolving towards satellite broadcasting with the potential for expansion.

Al Jazeera reached its international standing following the 9/11 attacks in 2001, when it became the first network to provide images of Osama Bin Laden. That material was also distributed to the major global television corporations through agencies. The matter of whether the images of the Al Qaeda leader were real and how they had been obtained even became a subject of debate in high-level international politics.

At that time, without today's digital technology, verification depended on human resources, which for Al Jazeera meant having a correspondent in Afghanistan, Taysir Alony, who was keenly aware of the political climate in the region, as well as image and sound technicians who could certify that the videos and audios had not been manipulated. At times, the process ended in a decision to delay the broadcast, while other times the materials were broadcast but with subtitles under the images with the phrases "Attributed to Al Qaeda" or "We have been unable to verify the authenticity".

Today, technology has evolved to make it possible to use Internet-based video editing and management systems in the newsroom, which allows metadata to be imported, registered, organized, and synchronized so different departments have access to the cloud where all the information is stored. This entails a radical shift in the newsroom's professional relationship with the material that it handles, going from a binary relationship (in which the journalist sends something via satellite) to a networked distribution in which audiovisual materials and data move around the Internet and are placed not in a specific repository but in a shared-access cloud.

### 4.1. New editorial structure: The digital unit

In 2006, Al Jazeera set up its digital unit after noticing the changes in technology and politics as well as in the attitude of a prosumer audience that was shifting from being passive observers to active collaborators.

The unit conducted an experiment by handing out cameras to activists in different regions in the Arab world, and in 2008, during Mubarak's rule, it was considering the idea of a report with material from Egyptian bloggers.

In December 2010, the Tunisian revolution erupted and the authorities of the Ben Ali regime closed down the Al Jazeera head offices. Even so, the corporation had no problems providing images from the main conflict sites because social media and mobile phones were quite widespread by then. The same happened days later in Egypt: thanks to the installation of a satellite, Al Jazeera broadcast the signal from Cairo's Tahir Square to the world.

**Figure 1.** Recording from social media made in Tunisia. A man celebrating the fall of Ben Ali in January 2011 (Al Jazeera)



Source: (23) شاهد أقوى فيديو من ثورة تونس: يحيا الشعب.. بن علي هرب - YouTube <https://www.youtube.com/watch?v=OKKvc4sxwfw>.

## 4.2. Distortion of user-generated content

When the Arab revolutions erupted in 2011, thousands of images were sent to Al Jazeera's headquarters, making it a vital source of information for the rest of the world. At decisive moments, the traffic on Al Jazeera's website rose by 1,000% and liveblogging by 2,000%. The traditional blog was overwhelmed, and Al Jazeera had to seek a platform capable of securely storing the traffic it was receiving from the social media. It hired a content management system (CMS), a framework for managing images and content that hosts users' forums and blogs (Bañon, 2016). User-generated content (UGC) is defined as images and content uploaded to the Internet by authors unrelated to the media (Van Dijk, 2009; Cheong and Morrison, 2013; Bahtar and Mazzini, 2016).

At that time, Al Jazeera was still seeking a model for guidelines to validate the images it received from activists, especially Egyptians and Tunisians, who were particularly active. It was a new experience for the entire world media

and mistakes were inevitably made in the midst of this historical acceleration. Nevertheless, a culture began to coalesce in which verification played an essential role, with the war in Syria marking a turning point. A specific unit with newsroom professionals was created, tasked primarily with verifying UGC, which became a news source virtually as important as agencies and the traditional newsroom (Marai, 2017: 22).

As the war in Syria dragged on, Al Jazeera became strongly committed to covering it, which brought about a new phase in the network's verification procedures. Training was provided to a hundred journalists to improve their skills dealing with UGC.

Even so, the network was unable to avoid mistakes. Recognizing and explaining them is part of its newsroom code of conduct. One example occurred in December 2015, when several channels, including Al Jazeera, broadcast a video going around YouTube in which some Canadian children were supposedly singing the song *Tala al Badr Aleina* to welcome refugees. The song comes from the Islamic tradition and tells of the times when the prophet Muhammad had to seek refuge in the city of Medina. At a time when there was a sense of solidarity towards Syrian refugees, the video received thousands of displays of solidarity around the world. However, after the broadcast it was proven that it had nothing to do with them. The network apologized and prepared a second piece explaining how the misinformation had occurred (Marai, 2007: 26).

**Figure 2.** Concert falsely attributed as a welcome for Syrian refugees in 2015 (Al Jazeera)



Source: Access to the video broadcast with false information about the performance: <https://www.youtube.com/watch?app=desktop&v=DSgoZNkA92>; access to the news item in which Al Jazeera apologizes and recognizes the mistake. Watch | Facebook <https://www.facebook.com/ajplusarabi/videos/989020491141496/>.

Another example of how the atmosphere can create the conditions for disinformation occurred following the terrorist attacks in Paris on 13 November 2015, which were committed simultaneously by gunmen and suicide bombers

in different places and caused 130 deaths. The Associated Press had recordings of the attack and identified a suicidal woman as Hasna Ait Boulahcen, without providing photos of her. But they began to circulate in the social media and Al Jazeera trusted one that several media had published. It turned out that the person in that photo was really Nabila Bakkatha. They located her in Morocco and interviewed her, admitting the mistake (El Katatney, 2017: 62).

**Figure 3.** Al Jazeera Plus broadcasts a video with statements by Nabila Bakkatha, a Moroccan woman, rectifying the news in which she had been identified as a suicide terrorist (Al Jazeera)
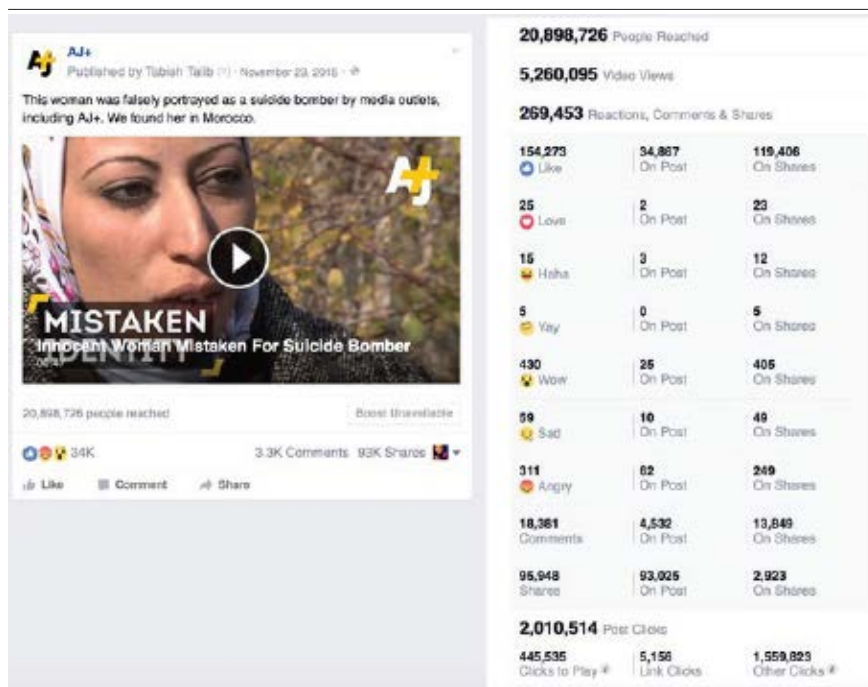


Source: Access to the news item in which Al Jazeera recognizes the mistake: <https://youtu.be/-xih_Q8OXpw>.

## 4.3. The effects of verification on the newsroom

The introduction of streamed images from the social media also entails a psychological leap in the concept of the newsroom, and in Al Jazeera there was resistance from some journalists (Khalifeh, 2017: 29).

In the midst of this digital context, the verification procedure always begins with checking routines similar to the journalistic procedures before the social media. Hence, before conducting technical checks, journalists conduct a prior exploration in an attempt to discover the origin of the source and ascertain whether it is reliable and therefore sent by someone they know or shared by a close colleague.

The level of falsehood in materials from UGC can appear in different forms:

— Totally authentic.
— True, but not recorded on the date it is broadcast.
— True, but recorded in a different place from the one stated.
— Manipulated, with distorted or added elements.
— False and created with the explicit intention to deceive (Ghazayel, 2017: 69).

Using this classification, the basic verification routine follows these basic steps:

— Checking Twitter lists, enabling the followers to be classified.
— Creating our own list to add accounts that interest us.
— Searching via Google for lists created by other organizations and people, creating networks.
— Using platforms like *Tweetdeck* that enable interaction with the platform.
— Joining a relevant thematic WhatsApp and Telegram list.
— Language check. Al Jazeera has a target audience in which Arabic is the majority language, with 420 million speakers and 30 dialects (Unesco, 2020). That implies consideration of the language, especially to confirm geographical location.
— And there is one basic precaution in the newsroom: downloading the video before beginning the verification process, since there is a risk it may be taken down.

## 4.4. Credibility tools and IT

Every day, digital journalism has more new resources to improve the verification procedure (Pellicer, 2019). Here are some of the accessible tools:

— Reverse image search, enabling images to be located if they are found elsewhere in the Internet. Google Image Search or Tineye are used with the key frames. The origin and context of the photo can then be ascertained.
— Comparing with other servers.
— Invid or Rev Eyecon can also conduct reverse searches for videos.
— ImgOps enables images with similar content to be compared and alterations to be detected.
— YouTube Data Viewer enables access to the main metadata.
— Fingerprinting Organizations with Collected Archives (FOCA) and Metashield Clean Up detect metadata and information that is not visible in documents that can be found on websites.
— Foller.me is a tool to find information about Twitter profiles.

— Pipl enables people's profiles to be found via what they post on the Internet.
— Anylizer analyses videos frame-by-frame, discovering irregularities if changes have been made.
— Google Earth, Wikimapia, Open Streets Map and Google Maps can be used to verify places via geolocation.

The list for detecting bots or fake profiles does not end there, nor do the techniques. For example, one can also look for the video's URL or its ID code in Twitter to trace the first person who started circulating it. It is also important to be able to demonstrate that the stated location is correct (Younes and Mackintosh, 2017: 45-46).

### 4.5. The purpose of the Al Jazeera+ online platform

By having professionally checked material, this study has been able to broaden the topics on the agenda by introducing material from the digital sphere while preserving credibility – an essential factor for any medium, and even more so in a region in serious political upheaval.

Hence, Al Jazeera's digital platform is an Internet space that works: the YouTube channel has racked up an average of three million views, and it has over a million followers on Twitter. In this vein, it has successfully managed a transition involving turning a simple news website into an experimental informative online platform (Bañon, 2016).

As a result of this evolution, in 2014 the corporation launched Al Jazeera Plus (AJ+), offering videos directly for Facebook, YouTube, Instagram, Twitter and Medium through mobile apps. It thus uses the social media to attract followers by encouraging them to share their stories, with formats based on very recent images or very short videos with little text. These changes and the entry into the digital world rendered it necessary to rethink internal procedures (Roettgers, 2013).

The experience of AJ+ increased the importance of verification in journalistic routines: the professionals are not always on the ground and the eyewitness reports come from indirect sources whose authorship is unknown at first glance, nor is there always information on the circumstances in which the images were recorded or the reasons for spreading them. One example of a verification process taken to the extreme of instantaneousness was the case of the Address Hotel fire in Dubai on New Year's Eve, 2015. Minutes before the traditional media reported the news, AJ+ used the Datamindr service to which it subscribed in order to verify the location of the tweets. They thus detected that three eyewitnesses were transmitting the information with *Periscope*.[1] The verification was completed by checking the location and comments from eyewitnesses in the social media and by securing permission to

---

1.   A Twitter app that enables live videos to be transmitted.

use images from one of them. Thus, AJ+ created a story in a matter of minutes (El Katatney, 2017).

**Figure 4.** The video of the Address Hotel fire was viewed three million times and shared 40,000 times in just one day, with an impact calculated at twenty million people



Source: Al Jazeera.

The observation that new formats have been incorporated into news spaces has thus been confirmed. One feature of AJ+'s posts is raw videos uploaded directly from mobile phones to the social media which are far from the quality standards of traditional broadcasts. They are products created for the digital sphere bearing the shared behavior and consumption of different social media in mind. This has enabled AJ+ to achieve audience growth, especially at times of humanitarian crises such as in Palestine or the exodus of refugees from the war in Syria. Such newsworthy situations have revealed that the most solid criteria to become viral are relevance, immediacy, and emotiveness (Rahimi, 2015). AJ+ averages half a million interactions a day, reaching 30 to 40 million people (El Katatney, 2017: 66).

## 5. Results: Examples of verification and impact in the newsroom

Al Jazeera's verification unit contributes 60 news items a day on average, which are added to the topics in the agenda after passing the pertinent checks. Here are examples in which it successfully detected fake news:[2]

---

2.  Data provided by Ahmed Mansour, Senior Editor for Digital Content at Al Jazeera.

1. Displacement of Syrian refugees after floods in northern Syria (20-01-21):
    Fake Link: <https://twitter.com/ahmadal_shame/status/1350887459366854656?s=19>
    Old story: <https://twitter.com/hamza_habboush/status/1205184394669051905?s=19>
    The verification unit discovered that the recording was from a video from 2019.

2. Photo of floods in Sudan (6-09-2020):
    Fake Link: <https://twitter.com/NadineKh/status/1302718409382264844>.
    Real Story: <https://www.flickr.com/photos/un_photo/4081320650>.
    The verification unit discovered the photograph was from Haiti, dated 2017.

3. Attack on Hassan Nasrallah's vehicle in the Lebanon (18-01-21):
    Fake story link: <https://twitter.com/ar28ar/status/1351405437111984129>
    Real story link: <https://www.facebook.com/100062033836280/videos/106490371428757>
    The verification unit discovered that although the date was correct, the image showed an explosion in the Syrian city of Idlib.

It is clear that using materials from the digital sphere entails a transition from the traditional, highly hierarchical newsroom to another model more in line with a horizontal rationale. This shift means that journalists have to work with a content management system, which enables them to make productions using accessible software and coexist with a system hosted in the cloud.

The nine members of Al Jazeera's verification unit are all under the age of 35 and have previously worked in the newsroom, preferably in posts in the international and sports sections.

The trend is towards a newsroom in which there is a shift from being producers of visual and news products to mediators and custodians of quality standards, with verification the top priority. This evolution will intensify as the subscription systems for distributing signals using broadband or IPTV connections improve,[3] thus enabling streaming media[4] with audiovisual content broadcast via the Internet. Guiding protocols are also increasingly being adopted to improve the security of the Internet as a mechanism for conveying the production processes (Hunter, 2021).

In the case of Al Jazeera, this transformation itself has given rise to new sources and a differentiated agenda of topics, although journalists specializing

3. Internet Protocol Television.
4. Streaming audio-visual content broadcast via the Internet.

in the tasks typical of each medium are still required even in newsrooms that have undergone integration processes (Salaverría and Negredo, 2008: 169).

In the traditional newsroom, organizing a live broadcast required hiring a satellite, finding a time zone when the content can be uploaded to it, and broadcasting it, not to mention a very significant economic investment. With the Make TV platform, Al Jazeera can download dozens of videos via Internet simultaneously. Thus, journalists from anywhere in the world can send news to a central clearinghouse, which decides whether the image should be stored or broadcast live, depending on the circumstances. One-way organization is thus giving way to work dynamics in the cloud, where the stream of images and data have to be managed. Satellites are still necessary for distribution, but the mobile networks have been playing an increasing role in channeling, exchanging, moving and organizing information. They generate and lend coherence to this movement and organization precisely in the cloud system.

At the core of this process are the verification routines, which provide recognition to the materials from the digital sphere. This intensifies the need for mediation from journalists, IT professionals and cross-disciplinary professionals. They are also highly specialized tasks that require a collaborative culture with the outside world, in which important entities become associates. For example, Al Jazeera is a member of the First Draft Partner, an entity encompassing the largest social media platforms and verification projects which are seeking to tackle these challenges together (Fernández-García, 2017).

Verification tasks guarantee credibility, but their use in a newsroom in the cloud where streams circulate poses four significant challenges: overcoming ethical tensions, putting proven facts into context, generating relevant information and observing public interest, and raising awareness about certain social issues. (Srisaracam, 2019)

## 6. Discussion and conclusions

This study has confirmed that including information and images originating from the digital sphere has made it necessary to broaden the agenda of topics and the number of sources (Q2). Since there is material coming from the circuit outside the newsroom, the verification process is crucial and has led to changes in the organization of the newsroom (Q1), with the creation of a group of journalists and technicians specializing in this task. The qualitative characteristics of these materials coming from the audience's mobile phones are far from the usual quality standards in the professional sphere, but they provide advantages such as immediacy and accessibility, which is leading to a new aesthetics of reporting (Q3) and contributing to the culture of TV viewers' participation. In this sense, management and verification makes it possible to use material from the social media with quality assurances. These social media have helped redirect audiences to the screen and locate new potential audiences (Q4) for television broadcasts or the channel's website.

In this study, therefore, Hypothesis 1 was confirmed and we can conclude that the verification processes and their improvement are key features in enabling traditional newsrooms to transition to a cloud structure, since they allow the veracity of visual material from the social media to be checked, thereby preserving a fundamental feature of quality journalism. This model is also shifting from hierarchically organized professional relationships towards more horizontal models, in which technological training is essential.

Hence, Hypothesis 2 is also confirmed in the sense that the new newsroom has news professionals coexisting with experts in artificial intelligence and IT. The new nature of the newsroom is no longer strictly in-person; correspondents coexist with journalists via virtual mobility, seeking and receiving images streamed from the social media. The cloud needs to organize the traffic, and managing it is leading to the establishment of new horizontal hierarchies within the newsroom. This should be the subject of subsequent studies, because the harmonization of the list of functions of the different information professionals should include excellence as a goal not only to find the best image or video but also to create the context to provide audiences with the best explanation via new formats generated in the social media. Technologies enable all of the material to be available to the group in the cloud newsroom, although verification protocols are the first step in this availability. Nevertheless, this availability is subject to usage protocols, for example prioritizing broadcasting in the network's news programs over others for strategic reasons.

Finally, Hypothesis 3 is also confirmed, as the use of algorithmic technology is essential in verifying the volume and diversity of materials generated on the Internet. However, at the same time, the need for verification does not exclude traditional techniques. A newsroom with a hybrid coexistence of professionals in no way diminishes the importance of journalism but instead makes it necessary. There is increasing evidence of the need to lend meaning and context to the information, which means that journalists keep and preserve their topic specialization.

Moreover, insofar as verification preserves the audience's trust, it allows topics and sources in the agenda that have traditionally been excluded, including new coverage. It is therefore a relevant and inevitable way to incorporate new audiences, thereby helping to grow audiences and rendering it possible to maintain big audiovisual corporations.

The additional challenge posed for journalism by the new cloud newsrooms is how to boost creativity and create protocols for internal cohesion to motivate the professionals to organize the creation of quality content that serves society.

## Acknowledgments

## Bibliographical references

ADAIR, B. (2019). "El fact checking es lo más importante del periodismo en la era digital". *El Mundo* [Consulted: 02/01/21]. Retrieved from <https://www.elmundo.es/television/2019/07/25/5d38750afc6c833d428b4625.html>.

AMAZEEN, M. (2017). "Journalistic interventions: The structural factors affecting the global emergence of fact-checking". *Journalism: Theory, Practice & Criticism*, 21.
<https://doi.org/10.1177%2F1464884917730217>

BAHTAR, A. and MAZZINI, M. (2016). "The Impact of User – Generated Content (UGC) on Product Reviews towards Online Purchasing - A Conceptual Framework". *Procedia Economics and Finance*, 37, 337-342.
<https://doi.org/10.1016/S2212-5671(16)30134-4>

BAÑON, L (2016). "Redes sociales y audiencias en televisión: el caso al Jazeera como referente para nuevos modelos de plataformas digitales informativas para la construcción de una nueva RTVV". *AdComunica*, 11, 191-194
<http://dx.doi.org/10.6035/2174-0992.2016.11.15>

BECKETT, C. (2020). *The Challenge of AI journalism: an historic moment?* [Consulted: 08/12/20]. Retrieved from <https://charliebeckett.medium.com/the-challenge-of-ai-journalism-an-historic-moment-7bbb8de10daa>.

CASTELLS, M.; VAN DER HAAK, B. and PARKS, M. (2012). "The Future of Journalism: Networked Journalism". *International Journal of Communication*, 6. [Consulted: 08/12/20]. Retrieved from <https://ijoc.org/index.php/ijoc/article/viewFile/1750/832&gt>.

CHEONG, H. and MORRISON, M. (2013). "Consumers' Reliance on Product Information and Recommendations Found in UGC", *Journal of Interactive Advertising*, 8 (2), 38-49.
<https://doi.org/10.1080/15252019.2008.10722141>

CODINA, L. (2017). *Revisiones sistematizadas y cómo llevarlas a cabo con garantías: Systematic reviews y SALSA Framework.* [Consulted: 18/01/20]. Retrieved from <https://www.lluiscodina.com/revision-sistematica-salsa-framework/>.

DEL-FRESNO-GARCIA, M. (2019). "Desórdenes informativos: sobreexpuestos e infrainformados en la era de la posverdad". *El Profesional de la Información*, 28 (3).
<https://doi.org/10.3145/epi.2019.may.02>

DUFFY, M. (2011). "Networked Journalism and Al-Jazeera English: How the Middle East Network Engages the Audience to Help Produce News". *Journal of Middle East Media*, 7. [Consulted: 18/01/20] Retrieved from <http://www.qu.edu.qa/file_source/qu/conference/jmem2017/Vol/7/En/Networked%20Journalism%20and%20Al-Jazeera%20English.pdf>.

EL KATATNEY, E. (2017). *Why Social Media Storytelling can take you closer to your audience.* Finding the Truth Amongst the Fakes, 55-67. Al Jazeera Media Institute, Doha.

EL-NAWAWY, M. and ISKANDER, A. (2002). *Al-Jazeera: How the free Arab news network scooped the world and changed the Middle East.* Cambridge, MA: Westview Press.

FERNÁNDEZ-GARCIA, N. (2017). "Fake news: una oportunidad para la alfabetización mediática". *Nueva sociedad*, 269. [Consulted: 20/10/20] Retrieved from <https://biblat.unam.mx/hevila/Nuevasociedad/2017/no269/8.pdf>.

GHAZAYEL, M. (2017). *Verification of User-Generated Video Sourced from Social Media.* Finding the Truth Amongst the Fakes, 68-88. Al Jazeera Media Institute, Doha.

GRANT, M. and BOOTH, A. (2009). "A typology of reviews: an analysis of 14 review types and associated methodologies". *Health Information and Libraries Journal*, 26, 91-108.
<https://doi.org/10.1111/j.1471-1842.2009.00848.x>

GRAVES, L. (2012). *Deciding What's True: Fact-Checking Journalism and the New Ecology of News.* [Consulted: 20/01/21] Retrieved from <https://core.ac.uk/download/pdf/161442732.pdf>.

— (2019). "Understanding the Promise and Limits of Automated Fact-Checking". *Reuters Institute* [Consulted: 18/01/20] <https://www.reutersagency.com/wp-content/uploads/2019/03/reuters-institute-graves-factsheet-180228.pdf>.

HERNON, P. (1995). Disinformation and Misinformation through the Internet: Findings of an Exploratory Study. *Government Information Quarterly*, 12 (2), 133-139.
<https://doi.org/10.1016/0740-624X(95)90052-7>

HUNTER, P. (2021). "Broadcasters seek new work life balance around remote working". *The Broadcast Bridge* [Consulted: 06/01/21] Retrieved from <https://www.thebroadcastbridge.com/content/entry/16344/broadcasters-seek-new-work-life-balance-around-remote-working>.

JARVIS, J. (2006). "Networked journalism". *Buzzmachine* [Consulted: 21/01/21] Retrieved from <https://buzzmachine.com/2006/07/05/networked-journalism/>.

KHALIFEH, M. (2017). *Social Newsgathering and News Verification in the Arab World.* Finding the Truth Amongst the Fakes, 29-44. Al Jazeera Media Institute, Doha.

LIM, C. (2018). "Checking how fact-checkers check". *Research & Politics*, July 2018.
<https://doi.org/10.1177/2053168018786848>

LÓPEZ PAN, F. and RODRIGUEZ RODRIGUEZ (2020). "El Fact Checking en España. Plataformas, prácticas y rasgos distintivos". *Estudios Sobre el Mensaje Periodístico*, 26 (3), 1045-1065.
<https://doi.org/10.5209/esmp.65246>

MCBRIDE, K. and ROSENSTIEL, T. (2013). *The new ethics of journalism: Principles for the 21st Century.* Los Angeles: Sage CQ Press.

MANFREDI, J.L. and UFARTE, M. J. (2020). "Inteligencia artificial y periodismo: una herramienta contra la desinformación". *Revista CIDOB d'Afers Internacionals*, 124, p. 49-72.
<https://doi.org/10.24241/rcai.2020.124.1.49>

MARAI, M. (2017). *In Newsrooms… Mistakes will always happen.* Finding the Truth Amongst the Fakes, 14-28. Al Jazeera Media Institute, Doha.

MILES, H. (2005). *Al-Jazeera, how Arab TV news challenged the world.* London: Abacus Publishers.

MOHSIN, M. (2020). *Estadísticas YouTube 2020-10 datos fascinantes de YouTube.* [Consulted: 11/01/21] Retrieved from <https://cl.oberlo.com/blog/estadisticas-youtube>.

PELLICER, M. (2019). *Herramientas contra las fake news* [Consulted: 19/12/20] Retrieved from <https://miquelpellicer.com/2019/02/herramientas-contra-las-fake-news/>.

RAHIMI, S. (2015). *How AJ+ reported from Baltimore using only mobile phones.* [Consulted: 15/12/20] Retrieved from <https://www.poynter.org/reporting-editing/2015/how-aj-reported-from-baltimore-using-only-mobile-phones/>.

Rodriguez Andrés, R. (2018). "Fundamentos del concepto de desinformación como práctica manipuladora en la comunicación política y las relaciones internacionales". *Historia y Comunicación Social*, 23 (1), 231-244.
<https://doi.org/10.5209/HICS.59843>

Rodriguez-Fernández, L. (2019). "Desinformación: retos profesionales para el sector de la comunicación". *El Profesional de la Información*, 28 (3).
<https://doi.org/10.3145/epi.2019.may.06>

Roettgers, J. (2013). *An exclusive behind-the-scenes look at AJ+, Al Jazeera's upcoming online news network.* [Consulted: 15/12/20] Retrieved from <https://gigaom.com/2013/11/22/an-exclusive-behind-the-scenes-look-at-aj-al-jazeeras-upcoming-online-news-network/>.

Roozenbeek, J. and Van Der Linden, S. (2019). "Fake news game confers psychological resistance against online misinformation". *Palgrave Communications*, 5 (65).
<https://doi.org/10.1057/s41599-019-0279-9>

Salaverría, R. and Negredo, S. (2008). *Periodismo integrado, convergencia de medios y reorganización de redacciones.* Barcelona: Editorial Sol 90.

Satti, M. (2020). "Al Jazeera Arabic and Al Jazeera English Websites: Agenda-Setting as a Means to Comparatively Analyze Online News Stories". *Communication & Society*, 33 (1)
<https://doi.org/10.15581/003.33.1.1-13>

Srisaracam, S. (2019). *Digital convergence in the newsroom: examining cross-media news production and quality journalism.* PhD Thesis Cardiff University. [Consulted: 15/01/21] Retrieved from <http://orca.cf.ac.uk/130199/>.

Thussu, D. (2007). *News as Entertainment: The Rise of Global Infotainment.* London: Sage Publications.

Unesco (2020). *World Arabic Language Day* [Consulted: 23/01/21] Retrieved from <https://en.unesco.org/commemorations/worldarabiclanguageday>.

Van Dijk, T. (2006). "Discurso y manipulación: discusión teórica y algunas aplicaciones". *Signos*, 39 (60), 49-74.
<https://doi.org/10.4067/S0718-09342006000100003>

— (2009). "Users Like You? Theorizing Agency in User-Generated Content". *Media, Culture & Society*, 31 (1), 41-58.
<https://doi.org/10.1177/0163443708098245>

Virilio, P. (1995). "Speed and Information: Cyberspace Alarm!". *CTheory*. [Consulted: 01/12/20] Retrieved from <https://journals.uvic.ca/index.php/ctheory/article/view/14657/5523>.

Younes, R. and Mackintosh, E. (2017). *Trusting What You See Online - It's Not Just About the Tools.* Finding the Truth Amongst the Fakes, 45-54. Al Jazeera Media Institute, Doha.

Zayani, M. (2005). *The Al Jazeera phenomenon. Critical Perspectives on new Arab media.* London: Pluto Press.

# *Deepfakes*: el próximo reto en la detección de noticias falsas*

Francisco José García-Ull
Universidad Europea de Valencia
Universidad Miguel Hernández. Instituto Mediterráneo de Estudios de Protocolo
franciscojose.garcia@universidadeuropea.es
fran.garcia@protocoloimep.com

## Resumen

Un *deepfake* o ultrafalso es un vídeo hiperrealista manipulado digitalmente para representar a personas que dicen o hacen cosas que en realidad nunca sucedieron. Estas representaciones sintéticas, generadas mediante técnicas computarizadas basadas en inteligencia artificial (IA), plantean serias amenazas para la privacidad, en un nuevo escenario en el que se incrementan los riesgos derivados de las suplantaciones de identidad. Con la sofisticación de las técnicas para el desarrollo de *deepfakes*, resulta cada vez más complicado detectar si las apariciones públicas o declaraciones de personajes influyentes responden a parámetros de realidad o, por el contrario, son resultado de representaciones ficticias. Este estudio tiene como objetivo plantear un estado de la cuestión a través del análisis de la actualidad académica y mediante una exhaustiva revisión bibliográfica. En el presente trabajo se busca dar respuesta a las cuestiones que planteamos a continuación, que entendemos de interés general, tanto en una vertiente económica y social como en diversas áreas de investigación: ¿qué son los *deepfakes*?, ¿quién los produce y qué tecnología los respalda?, ¿qué oportunidades plantean?, ¿qué riesgos se asocian a estos documentos multimedia?, ¿qué métodos existen para combatir estas falsificaciones? Y enmarcando el estudio en el ámbito de la teoría de la información: ¿se trata de una revolución o de una evolución de las *fake news*? Como sabemos, las noticias falsas influyen en la opinión pública y son efectivas a la hora de apelar a emociones y modificar comportamientos. Podemos asumir que estos nuevos textos audiovisuales serán tremendamente eficaces a la hora de

minar, más si cabe, la credibilidad de los medios digitales, así como de acelerar el ya evidente agotamiento del pensamiento crítico.

**Palabras clave:** *deepfakes*; noticias falsas; aprendizaje profundo; inteligencia artificial; desinformación

**Resum.** Deepfakes*: el pròxim repte en la detecció de notícies falses*

Un *deepfake* o hipertrucatge és un vídeo hiperrealista manipulat digitalment per representar persones que diuen o fan coses que en realitat mai no van succeir. Aquestes representacions sintètiques, generades mitjançant tècniques informatitzades basades en intel·ligència artificial (IA), plantegen serioses amenaces per a la privacitat, en un nou escenari en el qual s'incrementen els riscos derivats de les suplantacions d'identitat. Amb la sofisticació de les tècniques per al desenvolupament de *deepfakes*, resulta cada vegada més complicat detectar si les aparicions públiques o declaracions de personatges influents responen a paràmetres reals o, per contra, són resultat de representacions fictícies. Aquest estudi té com a objectiu plantejar un estat de la qüestió a través de l'anàlisi de l'actualitat acadèmica i mitjançant una exhaustiva revisió bibliogràfica. En el present treball es busca donar resposta a les qüestions que plantegem a continuació, que entenem d'interès general, tant en un vessant econòmic i social com en diverses àrees de recerca: què són els *deepfakes*?, qui els produeix i quina tecnologia els dona suport?, quines oportunitats plantegen?, quins riscos s'associen a aquests documents multimèdia?, quins mètodes existeixen per combatre aquestes falsificacions? I emmarcant l'estudi en l'àmbit de la teoria de la informació: es tracta d'una revolució o d'una evolució de les *fake news*? Com sabem, les notícies falses influeixen en l'opinió pública i són efectives a l'hora d'apel·lar emocions i modificar comportaments. Podem assumir que aquests nous textos audiovisuals seran tremendament eficaços a l'hora de minar, més si fos possible, la credibilitat dels mitjans digitals, així com d'accelerar el ja evident esgotament del pensament crític.

**Paraules clau:** *deepfakes*; notícies falses; aprenentatge profund; intel·ligència artificial; desinformació

**Abstract.** *DeepFakes: The Next Challenge in Fake News Detection*

A deepfake is a hyper-realistic video, digitally manipulated to represent people saying or doing things that never really happened. With the sophistication of techniques for developing these counterfeits, it is becoming increasingly difficult to detect whether public appearances or statements by influential people respond to parameters of reality or, on the contrary, are the result of fictitious representations. These synthetic documents, generated by computerized techniques based on Artificial Intelligence (AI), pose serious threats to privacy, in a new scenario in which the risks derived from identity theft are increasing. This study aims to advance the state of the art through the analysis of academic news and through an exhaustive literature review, seeking answers to the following questions, which we understand to be of general interest, from both an economic and a social perspective and in various areas of research. What are deepfakes? Who produces them and what technology supports them? What opportunities do they present? What risks are associated with them? What methods exist to combat them? And framing the study in terms of information theory: is this a revolution or an evolution of fake news? As we know, fake news influences public opinion and is effective in appealing to emotions and modifying behaviours. We can assume that these new audiovisual texts will be tremendously effective in undermining, even more if possible, the credibility of digital media, as well as accelerating the already evident exhaustion of critical thinking.

**Keywords:** deepfakes; fake news; deep learning; artificial intelligence; disinformation

## 1. Introducción

Las nuevas herramientas basadas en inteligencia artificial (IA) permiten la recreación de representaciones audiovisuales realistas originales que simulan la apariencia y el habla de los seres humanos. Estas representaciones sintéticas se conocen como *deepfake* (palabra compuesta en la que se combinan *deep learning*, o aprendizaje profundo, y *fake*, es decir, falso) y plantean serias amenazas para la privacidad, en un nuevo escenario en el que se incrementan los riesgos derivados de las suplantaciones de identidad. La tendencia demuestra una constatable proliferación de estas técnicas de representación virtual que, unidas a su fácil acceso y usabilidad, hacen posible generar de manera sencilla contenidos multimedia con una falsa apariencia de realidad. Los *deepfakes* pueden generarse a partir de imágenes, audios o vídeos.

Este escenario plantea serios retos que se incorporan a la ya compleja detección de noticias falsas. Los estudios recientes en el ámbito apuntan a que, en efecto, resultará cada vez más complicado detectar técnicamente si las apariciones públicas o declaraciones de personajes influyentes responden a parámetros de realidad o, por el contrario, son resultado de representaciones ficticias.

Los *deepfakes* son cada vez más realistas. Aunque hasta el momento podemos observar algunas características comunes en estas imágenes sintéticas (rostros sin gafas ni barba y que miran directamente a la cámara), el resultado es cada vez más sofisticado. De hecho, si bien los estudios para la identificación de estos vídeos manipulados han resultado eficaces atendiendo a indicadores biométricos (como el número de veces que el protagonista pestañea, diversas expresiones faciales o la composición de la dentadura), la mayoría de especialistas coinciden en que, a medida que se perfeccione la técnica, las diferencias con respecto al sujeto real (o *pristine*) serán cada vez más imperceptibles.

En este artículo explicamos cómo el uso de estas herramientas aplicadas al intercambio, reanimación y manipulación de rostros ha hecho ya saltar algunas alarmas. *Deepfakes* de personalidades como Barack Obama, Donald Trump o Mark Zuckerberg han conseguido acaparar repercusión mediática internacional, mientras que populares actrices o periodistas han sido las primeras víctimas de estas representaciones manipuladas en las que su rostro es insertado en contenidos de carácter pornográfico.

## 2. Marco teórico

En los últimos años, las noticias falsas se han convertido en un tema que amenaza el discurso público, la sociedad humana y la democracia (Borges, Martins y Calado, 2019; Mackenzie y Bhatt, 2018; Qayyum et al., 2019). En un escenario donde el caos reina en gran parte del ecosistema de información del que dependen las sociedades (Lin, 2019), la información falsa se propaga rápidamente a través de las redes sociales, donde puede impactar a millones de usuarios (Figueira y Oliveira, 2017). Las noticias falsas tienen un importante impacto en la construcción de la realidad por parte del receptor, hasta el punto

de que influyen en su percepción y toma de decisiones, aun a sabiendas de su origen deliberadamente ficticio (Keersmaecker y Roets, 2017). Actualmente, uno de cada cinco usuarios de internet recibe sus noticias a través de YouTube, solo superado por Facebook (Anderson, 2018). Este aumento en la popularidad del vídeo destaca la necesidad de herramientas para confirmar la autenticidad del contenido de los medios y las noticias, ya que las nuevas tecnologías permiten manipulaciones convincentes de vídeos o audios (Anderson, 2018). Dada la facilidad para obtener y difundir información errónea a través de las plataformas de redes sociales, tanto en forma de publicación como en los comentarios (Atasanova et al., 2019), cada vez es más difícil saber en qué confiar, lo que genera consecuencias perjudiciales para la toma de decisiones informadas (Borges et al., 2019; Britt et al., 2019). De hecho, hoy vivimos en lo que algunos autores identifican como un escenario de posverdad, que se caracteriza por la desinformación digital, el sesgo mediático (Hamborg et al., 2018), la generación de información falsa y la distorsión deliberada de la realidad, para manipular creencias y emociones e influir en la opinión pública y en actitudes sociales (Anderson, 2018; Qayyum et al., 2019).

Para identificar los *deepfakes*, debemos comprender las razones de su existencia y la tecnología que los respalda. Sin embargo, dado que los *deepfakes* aparecieron en internet en 2017, la literatura académica sobre el tema todavía es escasa. En este sentido, el presente estudio tiene como objetivo discutir qué son los *deepfakes* y quién los produce, cuáles son los beneficios y las amenazas que introducen, así como mostrar algunos ejemplos de *deepfakes* actuales, para señalar cómo identificarlos y, en su caso, combatirlos. Esta investigación analiza la actualidad académica sobre *deepfakes* y contribuye a la literatura entorno a las noticias falsas y a la desinformación.

Los recientes avances tecnológicos han facilitado la creación de vídeos hiperrealistas que utilizan intercambios de rostros y que dejan pocos rastros de manipulación (Chawla, 2019). Los *deepfakes* son el producto de aplicaciones de IA que fusionan, combinan, reemplazan y superponen imágenes y videoclips para crear vídeos falsos que parecen auténticos (Maras y Alexandrou, 2019). Así, es posible generar, por ejemplo, un vídeo humorístico, pornográfico o político de una persona que dice o hace algo sin su consentimiento (Day, 2019; Fletcher, 2018). El factor que cambia el juego de los *deepfakes* es el alcance, la escala y la sofisticación de la tecnología involucrada, ya que casi cualquier persona con una computadora puede fabricar vídeos falsos que son prácticamente indistinguibles de los medios auténticos (Fletcher, 2018). Si bien los primeros ejemplos de *deepfakes* se centraron en líderes políticos, actrices, comediantes y artistas con rostros entretejidos en vídeos porno (Hasan y Salah, 2019), en el futuro los *deepfakes* probablemente se usarán cada vez más para la pornografía de venganza —*revenge porn*—, el acoso, la evidencia de vídeos falsos en los tribunales, el sabotaje político, la propaganda terrorista, el chantaje, la manipulación del mercado y las noticias falsas (Maras y Alexandrou, 2019).

## 3. Metodología

Este estudio realiza una revisión bibliográfica, basada en la literatura académica emergente sobre *deepfakes*. En marzo de 2021 se recopilaron un total de 35 artículos académicos centrados en *deepfakes*. Todos estos estudios científicos han sido escritos en inglés y publicados entre 2018 y 2021. Fueron encontrados a través de la búsqueda de Google Scholar, utilizando las palabras clave *deepfake*, *deep fake* y las correspondientes formas plurales. Además, se realizaron búsquedas análogas en los buscadores de las principales editoriales académicas (Springer Nature, Wiley, ElSevier, ACS y Taylor & Francis). El corpus del estudio está compuesto por textos centrados en *deepfakes* que desarrollan sus investigaciones desde distintos prismas o áreas, entre los que se encuentran fundamentalmente las ciencias sociales (periodismo y teoría de la información) y los lenguajes y sistemas informáticos.

## 4. Resultados

### 4.1. *¿Qué son los* deepfakes*?*

Una combinación de *aprendizaje profundo* y *falso*, los *deepfakes* son vídeos hiperrealistas manipulados digitalmente para representar a personas que dicen y hacen cosas que en realidad nunca dijeron ni sucedieron. Los *deepfakes* se basan en redes neuronales que analizan grandes conjuntos de datos para aprender a imitar las expresiones faciales, los gestos y la voz de una persona. El proceso implica introducir imágenes de dos personas en un algoritmo de aprendizaje profundo para intercambiar caras (Rössler et al., 2018).

Los *deepfakes* salieron a la luz en 2017 cuando un usuario de Reddit publicó vídeos que mostraban a celebridades en situaciones sexuales comprometedoras. Se trata de montajes difíciles de detectar, ya que se basan en grabaciones reales que incluso pueden tener audio con sonido auténtico.

Los *deepfakes* están optimizados para ser compartidos fácilmente en redes sociales, donde las conspiraciones, los rumores y la información errónea se difunden fácilmente. Al mismo tiempo, el escenario de posverdad empuja a las personas a pensar que no pueden confiar en ninguna información a menos que provenga de sus redes sociales, incluidos familiares, amigos cercanos o parientes, y respalda las opiniones que ya tienen. De hecho, muchas personas están abiertas a cualquier cosa que confirme sus puntos de vista existentes, incluso si sospechan que puede ser falsa (Jang y Kim, 2018). Existen aplicaciones gratuitas para generar *deepfakes* de forma sencilla y fácilmente accesibles, que permiten a usuarios con pocas habilidades técnicas y sin ninguna experiencia artística editar vídeos, intercambiar caras, alterar expresiones y sintetizar el habla casi a la perfección.

En cuanto a la tecnología, los *deepfakes* son el producto de las redes generativas antagónicas (RGA), también conocidas como GAN en inglés. Son una clase de algoritmos de IA que se utilizan en el aprendizaje no supervisado, implementados por un sistema de dos redes neuronales que compiten mutua-

mente en una especie de juego de suma cero. Esta técnica, presentada por investigadores de la Universidad de Montreal (Goodfellow et al., 2014), puede generar fotografías que parecen auténticas a observadores humanos. Las RGA se basan en dos redes neuronales artificiales que trabajan juntas para crear medios de aspecto real. Estas dos redes llamadas *el generador* y *el discriminador* se entrenan en el mismo conjunto de datos de imágenes, vídeos o sonidos. Luego, el primero intenta crear nuevas muestras que sean lo suficientemente buenas para engañar a la segunda red, que trabaja para determinar si los nuevos medios que ve son reales. De esa manera, se impulsan entre sí para mejorar.

En el caso de los *deepfake*, el generador de las RGA crea nuevas imágenes basándose en una base de datos de fotografías o vídeos (a mayor número de

**Figura 1.** Edmond de Belamy. Primer retrato generado por una IA, subastado en Christie's en Nueva York, 2018



Fuente: Portrait of Edmond de Belamy. Wikipedia.

registros en la base de datos, mayor precisión). Una vez el algoritmo generador crea el nuevo contenido, el algoritmo discriminador realiza un cribado sobre el resultado generado, para delimitar si la imagen o vídeo reúne las características para ser un *deepfake* realista.

Una RGA puede observar miles de fotos de una persona y producir un nuevo retrato que se aproxime a esas fotos sin ser una copia exacta de ninguna de ellas. El resultado es una imagen, vídeo o audio original. En un futuro cercano, las RGA se capacitarán con menos información y podrán intercambiar cabezas, cuerpos enteros y voces. Aunque los *deepfakes* suelen requerir una gran cantidad de imágenes para crear una falsificación realista, los investigadores de la materia están desarrollando técnicas que permiten generar un vídeo falso a partir de una sola fotografía, por ejemplo, una autofoto.

**Figura 2.** Aritmética vectorial para rostros generados por RGA.



man with glasses    −    man without glasses    +    woman without glasses    =    woman with glasses
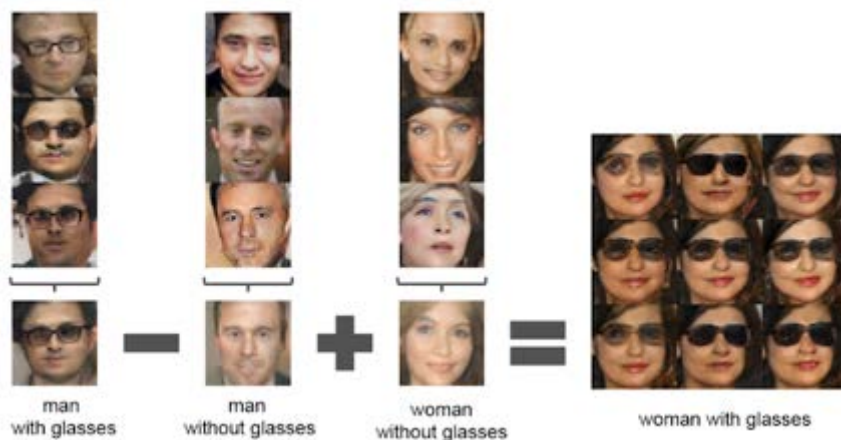
Fuente: Radford, Metz y Chintala (2015).

## 4.2. ¿*Quién produce* deepfakes?

Podemos distinguir entre al menos cuatro principales actores involucrados en la producción de *deepfakes*, a saber: comunidades de aficionados, actores políticos, delincuentes y estafadores y, por último, actores legítimos, como productores audiovisuales, creadores artísticos (Floridi, 2018) o agencias de publicidad.

En 2017 un usuario de Reddit presentó una colección de *deepfakes* pornográficos de celebridades. En solo unos meses, la comunidad de seguidores que se generó alrededor de *deepfakes* llegaría a los 90.000 miembros.

En estas comunidades, muchos aficionados se centran en *deepfakes* relacionados con la pornografía, mientras que otros colocan a actores famosos en películas en las que nunca aparecieron para producir efectos cómicos. En general, los seguidores tienden a ver los vídeos creados por IA como una

nueva forma de humor en línea y una contribución al desarrollo de dicha tecnología: más como un rompecabezas intelectual, que como una forma de engañar o amenazar a las personas. Sus *deepfakes* están destinados a ser entretenidos, divertidos o centrados en la sátira política, y pueden ayudar a ganar seguidores en las redes sociales. Algunos aficionados pueden estar buscando beneficios personales más concretos, como dar visibilidad al potencial de esta tecnología para exponer sus creaciones y así promocionar su talento.

Sin embargo, también es cierto que existe la amenaza de que actores políticos, *hacktivistas*, agitadores, inferencias de gobiernos extranjeros, etcétera puedan usar *deepfakes* en campañas de desinformación para manipular la opinión pública y debilitar la confianza en las instituciones de un país determinado. En este sentido, los *deepfakes* podrían interferir en unas elecciones o agitar a determinados segmentos para sembrar disturbios civiles.

Los estafadores también están utilizando cada vez más los *deepfakes* con el fin de llevar a cabo la manipulación del mercado y las acciones de una compañía, así como otros delitos financieros. Los delincuentes ya han utilizado audios falsos generados por IA para hacerse pasar por un ejecutivo en el teléfono que solicita una transferencia de efectivo urgente (otra vuelta de tuerca en el conocido fraude del CEO). Existe la tecnología para falsificar videollamadas a tiempo real utilizando imágenes de rostros ya disponibles públicamente en internet.

### 4.3. *Posibles amenazas de los* deepfakes

Con relación al objetivo que persiguen y de acuerdo con la bibliografía revisada, la comunidad académica coincide en que los *deepfakes* pueden significar una amenaza para distintos actores sociales, como el sistema político o el empresarial, así como para la ciudadanía en general. En este sentido, estas falsificaciones sintéticas ejercen presión sobre los periodistas que luchan por filtrar noticias reales de las falsas, amenazan la seguridad nacional al difundir propaganda que pueda alterar los resultados de unas elecciones, obstaculizan la confianza de los ciudadanos en la información transmitida por las autoridades y plantean problemas de ciberseguridad tanto para internautas como para organizaciones.

Es muy probable que la industria del periodismo tenga que enfrentarse a un problema masivo de confianza del consumidor debido a los *deepfakes*. Los *deepfakes* representan una amenaza mayor que las noticias falsas tradicionales porque son más difíciles de detectar. La tecnología permite la producción de vídeos de noticias aparentemente legítimos que ponen en riesgo la reputación de los periodistas y los medios de comunicación. Además, ganar la carrera para acceder a las imágenes de vídeo filmadas por el testigo de un incidente puede proporcionar una ventaja competitiva a un medio de comunicación, mientras que el peligro aumenta si las imágenes ofrecidas son falsas.

Los *deepfakes* pueden obstaculizar la alfabetización digital y la confianza de los ciudadanos en la información proporcionada por la autoridad. No obs-

tante, el aspecto más dañino de los *deepfakes* puede que no sea la desinformación *per se*, sino más bien la falta de confianza en las noticias, incluso en formato audiovisual, derivada de un contacto constante con la desinformación. Este aspecto refuerza el fenómeno denominado «apocalipsis de la información» o «apatía por la realidad» (Westerlund, 2019), que derivaría en un agotamiento del pensamiento crítico. Además, las personas pueden incluso descartar imágenes genuinas como falsas, simplemente porque se han arraigado en la noción de que cualquier cosa que no quieran creer debe ser falsa. En otras palabras, la mayor amenaza no reside en el hecho de que el receptor de la información sea engañado, sino en que la información misma pierda toda la credibilidad.

Los problemas de ciberseguridad constituyen otra amenaza impuesta por los *deepfakes*. El mundo empresarial ya ha expresado interés en protegerse contra los fraudes virales, ya que los *deepfakes* podrían usarse para manipular el mercado y las acciones, por ejemplo, al mostrar a un director ejecutivo formulando insultos racistas o misóginos, anunciando una fusión falsa, difundiendo declaraciones falsas sobre pérdidas financieras o retratando a responsables cometiendo prácticas delictivas. Además, los anuncios de productos o pornografía *deepfake* podrían utilizarse para sabotear una marca o como forma de chantaje y extorsión. Además, la tecnología *deepfake* permite la suplantación digital en tiempo real de un ejecutivo, por ejemplo, para pedir a un empleado que realice una transferencia de efectivo urgente o proporcione información confidencial. La tecnología *deepfake* puede incluso crear una identidad fraudulenta y, en vídeos de transmisión en directo, convertir el rostro de un adulto en el rostro de un niño o una persona más joven. Se trata de una posibilidad que genera especial preocupación, dado su potencial uso por parte de redes de pederastia.

## 4.4. *Ejemplos de* deepfakes

La mayoría de *deepfakes* de hoy en plataformas sociales como YouTube o Facebook pueden verse como obras artísticas o divertidas inofensivas que utilizan figuras públicas, pero también hay ejemplos del lado oscuro de los *deepfakes*, a saber, el porno de celebridades y de venganza (*revenge porn*), o los intentos de influencia política.

Muchos *deepfakes* se centran en celebridades, políticos y líderes corporativos porque internet está repleto de fotos y vídeos originales de ellos, lo que hace factible la construcción de grandes repositorios de imágenes, necesarios para entrenar un sistema de *deepfake* de IA. La mayoría de estos *deepfakes* son bromas y memes divertidos, con efecto cómico o satírico. Un *deepfake* puede mostrar, por ejemplo, a Sylvester Stallone actuando en películas que nunca ha protagonizado, como *Indiana Jones* o *Terminator 2*.

**Figura 3.** *Deepfake* en el que Sylvester Stallone protagoniza *Terminator 2*



Fuente: Ctrl Shift Face. YouTube.

Algunos ejemplos interesantes de *deepfakes* incluyen el Museo Dalí en Florida (Estados Unidos), que ha utilizado la tecnología basada en IA para crear montajes que devuelven a la vida a Salvador Dalí (Kwok y Koh, 2020), o el anuncio publicitario de una marca cervecera en el que se muestra una representación sintética de la cantante Lola Flores.

**Figura 4.** Izquierda: Una representación sintética de Salvador Dalí recibe a los visitantes del Museo Dalí en Florida. Derecha: Un *deepfake* de Lola Flores en el anuncio de la marca Cruzcampo



Fuente: Izquierda, The Dalí Museum. YouTube. Derecha, CruzcampoTV. YouTube.

Sin embargo, también están apareciendo cada vez más ejemplos de *deepfakes* dañinos. Los *deepfakes* permiten la pornografía de celebridades y de venganza, es decir, la pornografía involuntaria con imágenes de celebridades y no celebridades, que se comparte en las redes sociales sin su consentimiento. Así, rostros populares como el de la actriz Scarlett Johansson han aparecido en películas para adultos *deepfakes*, en las que sus caras se han superpuesto a las

de estrellas porno. El resultado es una cosificación de los cuerpos de las mujeres como algo para ser consumido visualmente, que además elude su consentimiento (Wagner y Blewer, 2019). En la escena política, un *deepfake* de 2018 creado por el cineasta de Hollywood Jordan Peele presentaba al expresidente estadounidense Barack Obama discutiendo los peligros de las noticias falsas y burlándose del presidente Donald Trump. Un *deepfake* podría incluso utilizarse para crear falsos recuerdos, dada la maleabilidad de la memoria humana, e impactar directamente en la toma de decisiones de individuos con cargos de alta responsabilidad (Liv y Greenbaum, 2020).

**Figura 5.** Polémico vídeo en el que los creadores denuncian mediante un *deepfake* de B. Obama los peligros de esta tecnología



Fuente: BuzzVideoFeed. YouTube.

En 2019, un vídeo alterado de la política estadounidense Nancy Pelosi se volvió viral y tuvo un alcance masivo. El vídeo se ralentizó para hacerla parecer intoxicada. En un vídeo *deepfake* de 2018, Donald Trump ofreció consejos a la gente de Bélgica sobre el cambio climático. El vídeo fue creado por un partido político belga —el Partido Socialista Flamenco (SP.A)— con el objetivo de generar un debate social. Hacia el final del vídeo, Trump dice: «Todos sabemos que el cambio climático es falso, al igual que este vídeo». Sin embargo, la última frase no se tradujo en los subtítulos holandeses. El vídeo provocó indignación por la intromisión del presidente estadounidense de un país extranjero en la política climática de Bélgica. En 2019, el Partido Demócrata de los Estados Unidos creó un *deepfake* de su propio presidente, Tom Pérez, para resaltar la amenaza potencial de los *deepfakes* para las elecciones de 2020.

Si bien estos son ejemplos de influencia política limitada, otros *deepfakes* pueden tener un impacto más duradero. En África Central, en 2018, un vídeo del presidente de Gabón, Ali Bongo, a quien se creía en mala salud o

muerto, fue citado como el detonante de un fallido golpe de estado por parte del ejército gabonés. Y en Malasia, un clip viral *deepfake* de la confesión de un hombre de haber tenido relaciones sexuales con un ministro del gabinete local provocó controversia política. También se han utilizado altos ejecutivos para crear *deepfakes*. En junio de 2019, un *deepfake* de alta calidad elaborado por dos artistas británicos con el CEO de Facebook, Mark Zuckerberg, acumuló millones de visitas. El vídeo muestra falsamente a Zuckerberg elogiando a Spectre, una organización ficticia malvada de la serie de James Bond. Con imágenes de noticias, IA y un actor de voz, el vídeo estaba destinado a mostrar cómo se pueden utilizar estas representaciones sintéticas para manipular la realidad.

## 4.5. *Métodos para combatir las falsificaciones profundas*

Los artículos revisados sugieren que hay cuatro formas de combatir las falsificaciones profundas: la legislación y la regulación; la concienciación de las políticas corporativas; la educación y capacitación, y la tecnología *antideepfakes*, que incluye detección de *deepfakes*, autenticación de contenido y prevención de *deepfakes*.

En la actualidad, los *deepfakes* no se tratan específicamente en las leyes civiles o penales, aunque los expertos legales han sugerido adaptar las leyes actuales para cubrir la difamación, la falsificación o la suplantación de identidad. En este sentido, la creciente sofisticación de las tecnologías de IA exige nuevos tipos de leyes y marcos regulatorios. Por ejemplo, los *deepfakes* plantean preocupaciones sobre la privacidad y los derechos de autor, ya que las representaciones visuales de las personas en los vídeos *deepfakes* no son copias exactas de ningún material existente, sino más bien nuevas representaciones generadas por IA. Por lo tanto, los reguladores deben navegar por un panorama legal difícil en torno a las leyes de propiedad y libertad de expresión para regular adecuadamente el uso de esta tecnología.

Por otro lado, el aumento del discurso del odio, las noticias falsas y la desinformación que contaminan las redes sociales (Aldwairi y Alwahedi, 2018) han llevado a algunas empresas a tomar más medidas, como suspender las cuentas de los usuarios e invertir en tecnología de detección más rápida. Reddit y Pornhub han prohibido la pornografía *deepfake* y otra pornografía no consentida, y actúan cuando los usuarios etiquetan y denuncian dicho material.

Facebook pretende prohibir estas falsas representaciones y prevenir su proliferación. Precisamente con este objetivo, la red social ha iniciado una colaboración con la agencia de noticias Reuters destinada a identificar y evitar el efecto viral que podrían causar ciertas imágenes sintéticas. Esta cooperación se enmarca en el programa Facebook Journalism Project, que tiene como objetivo la creación de herramientas de verificación de noticias para combatir la vigente desinformación en la red social. Los algoritmos de Instagram, por otro lado, no recomendarán que las personas vean contenido marcado como falso por los verificadores de hechos (Harrison, 2019; Constine, 2019).

También algunos de los principales medios de comunicación, como *The Wall Street Journal*, *The Washington Post* o Reuters, han tomado medidas para detectar, etiquetar y desmentir información falsa, especialmente *deepfakes* (Vizoso et al., 2021).

**Figura 6.** *Fact-Checker* de Instagram
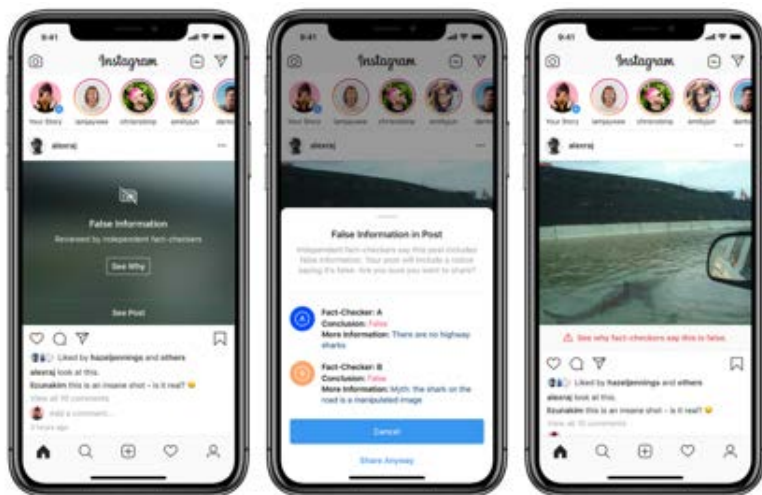


Fuente: Constine (2019).

La educación y la formación son cruciales para combatir los *deepfakes*. A pesar de la considerable cobertura mediática y las preocupaciones presentadas por las autoridades, el público aún no ha tenido en cuenta la amenaza de falsificaciones profundas. El estudio de las noticias falsas demuestra que el individuo tiende a pensar que la información falsa afectará con mayor probabilidad a terceras personas que a los miembros de su grupo (Jang y Kim, 2018). Esta sensación ficticia de seguridad crea un ideal caldo de cultivo para la proliferación de paparruchas, bulos y montajes. En este sentido, cobra fundamental importancia la necesidad de concienciar al público sobre el potencial de uso indebido de la IA. Mientras que las representaciones sintéticas brindan a los ciberdelincuentes nuevas herramientas para la ingeniería social, las empresas y organizaciones deben estar en alerta máxima y establecer planes de acción frente a esta latente amenaza. En este sentido, no es de extrañar que numerosos expertos recomienden que en las escuelas se enseñe el pensamiento crítico y la alfabetización digital (Dagdilelis, 2018), ya que estos rasgos contribuyen a la capacidad de los niños para detectar noticias falsas e interactuar de forma más respetuosa entre ellos en línea.

También es importante recordar que la calidad no es un indicador de la autenticidad de un vídeo. Además, las personas deben comprender que, a medida que se desarrolle la tecnología, se requerirán menos fotografías de

rostros reales para crear *deepfakes* y que nadie es inmune. Cualquiera que publique una sola selfi o un vídeo que capture 30 fotogramas por segundo en un sitio de redes sociales corre el riesgo de sufrir una falsificación profunda (Westerlund, 2019). Si bien el mejor método es mantener las fotos y los vídeos fuera de internet, incorporar al vídeo objetos en movimiento o grabar bajo ciertas condiciones de iluminación pueden brindar cierta protección. Las empresas, los gobiernos y las autoridades que utilizan tecnología de reconocimiento facial y almacenan grandes cantidades de datos con fines de seguridad y verificación son actores especialmente sensibles que deben prevenir posibles ataques informáticos y fugas de información.

La tecnología *antideepfake* proporciona quizás el conjunto de herramientas más variado para detectar *deepfakes*, autenticar contenido y evitar que el contenido se utilice para producir representaciones sintéticas.

En la actualidad, se desarrollan en paralelo distintos métodos para detectar *deepfakes* de manera automatizada. Si bien la tecnología basada en IA ya demuestra su eficacia en la detección de noticias falsas (Cybenko y Cybenko, 2018), la gran cantidad de material a analizar representa en el momento un importante reto tecnológico. Por ejemplo, los usuarios cargan 500 horas de contenido por minuto en YouTube. Twitter lucha con ocho millones de cuentas a la semana que intentan difundir contenido manipulado. *Deepfakes* publicados y compartidos en Twitter acumulan millones de visualizaciones (Pérez et al., 2021).

Esto crea enormes desafíos para que las tecnologías revisen todo el material publicado en poco tiempo. Además, los desarrolladores de *deepfakes* tienden a utilizar los resultados de las investigaciones de *deepfakes* publicadas para mejorar su tecnología y sortear nuevos sistemas de detección. Por ejemplo, los investigadores descubrieron que los primeros resultados de *deepfakes* no lograban imitar la velocidad a la que una persona parpadea (Li et al., 2018), mientras que los programas recientes han solucionado la falta de parpadeo o parpadeo antinatural después de que se publicaran los hallazgos.

Los expertos en investigación digital forense (Köhn et al., 2006) han sugerido indicadores sutiles para detectar *deepfakes* que incluyen una variedad de imperfecciones en el brillo y la distorsión de la cara; la ondulación en los movimientos de una persona; inconsistencias con el habla y los movimientos de los labios (Korshunov y Marcel, 2019); movimientos anormales de objetos fijos, como un pie de micrófono; inconsistencias en la iluminación, reflejos y sombras; bordes borrosos; ángulos y desenfoque de rasgos faciales; falta de respiración; dirección visual antinatural; frecuencia de parpadeo anormal (Li et al., 2018); falta de rasgos faciales, como un lunar conocido en una mejilla; suavidad y peso de la ropa y el cabello; piel demasiado suave; falta de detalles de cabello y dientes; desalineación en la simetría facial; inconsistencias en los niveles de píxeles, y comportamiento extraño de un individuo que hace algo inverosímil (Westerlund, 2019). Algunos investigadores han demostrado también la eficacia en el análisis computerizado mediante RGA de las principales expresiones faciales (sorpresa, enfado, felicidad, tristeza, miedo, repugnancia y neutralidad) en la detección de *deepfakes* (Anwar et al., 2019). Si

bien cada vez es más difícil para las personas distinguir entre un vídeo real y uno falso, la IA puede ser fundamental para detectar *deepfakes*.

Además, los investigadores que están desarrollando tecnologías RGA pueden diseñar y poner en práctica las salvaguardas adecuadas para que sus tecnologías sean más difíciles de usar indebidamente con fines de desinformación.

Uno de los principales retos tecnológicos actuales persigue una detección automatizada de *deepfakes* cada vez más precisa. En este sentido, las principales plataformas tecnológicas (Google, Facebook o Baidu) están promoviendo competiciones con importantes premios en metálico para animar a la comunidad de desarrolladores a implementar métodos cada vez más fiables en la detección de vídeos falsificados mediante IA. En 2020 se consiguió una precisión en la detección automatizada de *deepfakes* entorno al 80% de acierto. El proyecto «Advanced Deep Learning for Computer Vision», de la Universidad Técnica de Múnich, logró una precisión del 81% a partir de la base de datos Faceforensic++. Por otro lado, el desarrollador Selim Seferbekov resultó ganador del Deepfakes Detection Challenge, con una precisión del 82%.

**Figura 7.** Funcionamiento de *Faceforensics*++, herramienta para el reconocimiento automatizado de imágenes digitales manipuladas
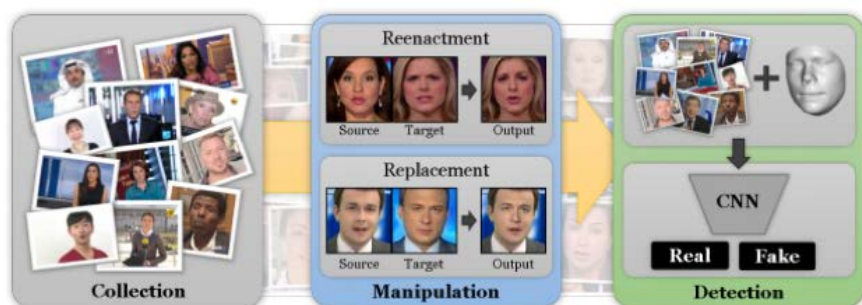


Fuente: Rössler et al. (2018).

## 5. Conclusiones

Del mismo modo que afrontamos los retos derivados de la ciberseguridad en general, el primer paso hacia una solución para esta desinformación es comprender el problema y su capacidad para afectarnos. Solo entonces es posible desarrollar e implementar soluciones técnicas que puedan resolver los desafíos. Dicho esto, ninguna de las soluciones tecnológicas puede eliminar por completo el riesgo de *deepfakes*, y el solucionismo tecnológico o la necesidad de buscar soluciones tecnológicas a cada problema, como señala Morozov (2013), pueden incluso desorientar la discusión de preguntas más existenciales sobre por qué existen *deepfakes* y qué otra amenaza puede imponer la IA (Westerlund, 2019). El desafío de la desinformación multimedia habilitada por IA es evolutivo, no revolucionario (Whyte, 2020). No obstante, pode-

mos prever que la democratización de las herramientas basadas en IA producirá en la información efectos estructurales y sistemáticos.

Por lo tanto, y a modo de resumen, asumimos que la estrategia más eficiente para combatir la propagación de los *deepfakes* implica una combinación de avances legales, educativos y tecnológicos y, por supuesto, la implicación de actores gubernamentales, académicos y científicos.

Como sabemos, las noticias falsas, que obedecen a lógicas virales cuando se transmiten mediante redes sociales y servicios de mensajería instantánea, influyen en la opinión pública y son efectivas a la hora de apelar a emociones y modificar comportamientos. En este contexto de posverdad, donde la información se encorseta en píldoras y el titular se prepara como cebo de clics, el incremento de estas técnicas multimedia tiene un efecto catalizador.

No podemos prevenir un escenario donde un *deepfake* decante la balanza en unas elecciones, provoque una crisis bursátil o sea el detonante de una revuelta. Sin embargo, podemos asumir que estas técnicas serán tremendamente eficaces a la hora de minar, más si cabe, la credibilidad de los medios digitales, así como de acelerar el ya evidente agotamiento del pensamiento crítico.

## Referencias bibliográficas

ALDWAIRI, M. y ALWAHEDI, A. (2018). «Detecting Fake News in Social Media Networks». *Procedia Computer Science*, 141, 215-222.
<https://doi.org/10.1016/j.procs.2018.10.171>

ANDERSON, K. E. (2018). «Getting acquainted with social networks and apps: combating fake news on social media». *Library HiTech News*, 35 (3), 1-6.
<https://doi.org/10.1108/LHTN-02-2018-0010>

ANWAR, S.; MILANOVA, M.; ANWER, M. y BANIHIRWE, A. (2019). «Perceptual Judgments to Detect Computer Generated Forged Faces in Social Media». En: SCHWENKER, F. y SCHERER. S. (eds.). *Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction*. MPRSS, 2018. *Lecture Notes in Computer Science*, 11.377. Springer, Cham.
<https://doi.org/10.1007/978-3-030-20984-1_4>

ATANASOVA, P.; NAKOV, P.; MÀRQUEZ, L.; BARRÓN-CEDEÑO, A.; KARADZHOV, G.; MIHAYLOVA, T.; MOHTARAMI, M. y GLASS, J. (2019). «Automatic Fact-Checking Using Context and Discourse Information». *Journal of Data and Information Quality*, 11 (3), art. n. 12.
<https://doi.org/10.1145/3297722>

BORGES, L.; MARTINS, B. y CALADO, P. (2019). «Combining Similarity Features and Deep Representation Learning for Stance Detection in the Context of Checking Fake News». *Journal of Data and Information Quality*, 11 (3), art. n.º 14.
<https://doi.org/10.1145/3287763>

BRITT, M. A.; ROUET, J.-F.; BLAUM, D. y MILLIS, K. (2019). «A Reasoned Approach to Dealing with Fake News». *Policy Insights from the Behavioral and Brain Sciences*, 6 (1), 94-101.
<https://doi.org/10.1177/2372732218814855>

CHAWLA, R. (2019). «Deepfakes: How a pervert shook the world». *International Journal of Advance Research and Development*, 4 (6), 4-8.
<http://doi.org/10.22215/timreview/1282>

CONSTINE, J. (2019). «Instagram hides false content behind warnings, except for politicians». *TechCrunch*. Recuperado de <https://techcrunch.com/2019/12/16/instagram-fact-checking>.

CYBENKO, A. K. y CYBENKO, G. (2018). «AI and Fake News». *IEEE Intelligent Systems*, 33 (5), 3-7.
<https://doi.org/10.1109/MIS.2018.2877280>

DAGDILELIS, V. (2018). «Preparing teachers for the use of digital technologies in their teaching practice». *Research in Social Sciences and Technology*, 3 (1), 109-121.
<http://doi.org/10.46303/ressat.03.01.7>

DAY, C. (2019). «The Future of Misinformation». *Computing in Science & Engineering*, 21 (1), 108-108.
<https://doi.org/10.1109/MCSE.2018.2874117>

FIGUEIRA, A. y OLIVEIRA, L. (2017). «The current state of fake news: challenges and opportunities». *Procedia Computer Science*, 121, 817-825.
<https://doi.org/10.1016/j.procs.2017.11.106>

FLETCHER, J. (2018). «Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance». *Theatre Journal*, 70 (4), 455-471. ProjectMUSE.
<https://doi.org/10.1353/tj.2018.0097>

FLORIDI, L. (2018). «Artificial Intelligence, Deepfakes and a Future of Ectypes». *Philosophy & Technology*, 31 (3), 317-321.
<https://doi.org/10.1007/s13347-018-0325-3>

GOODFELLOW, I. J.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A. y BENGIO, Y. (2014). «Generative Adversarial Networks». arXiv:1406.2661.

HAMBORG, F.; DONNAY, K. y GIPP, B. (2018). «Automated identification of media bias in news articles: an interdisciplinary literature review». *International Journal on Digital Libraries*, 20, 391-415.
<https://doi.org/10.1007/s00799-018-0261-y>

HARRISON, S. (2019). «Instagram Now Fact-Checks, but Who Will Do the Checking?». *Wired*. Recuperado de <https://www.wired.com/story/instagram-fact-checks-who-will-do-checking/>.

HASAN, H. R. y SALAH, K. (2019). «Combating Deepfake Videos Using Blockchain and Smart Contracts». *IEEE Access*, 7, 41.596-41.606.
<https://doi.org/10.1109/ACCESS.2019.2905689>

JANG, S. M. y KIM, J. K. (2018). «Third person effects of fake news: Fake news regulation and media literacy interventions». *Computers in Human Behavior*, 80, 295-302.
<https://doi.org/10.1016/j.chb.2017.11.034>

KEERSMAECKER, J. de y ROETS, A. (2017). «Fake news: Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions». *Intelligence*, 65, 107-110.
<https://doi.org/10.1016/j.intell.2017.10.005>

KÖHN, M.; OLIVIER, M. S. y ELOFF, J. H. (2006). «Framework for a Digital Forensic Investigation». *ISSA* 1-7.

KORSHUNOV, P. y MARCEL, S. (2019). «Vulnerability assessment and detection of deepfake videos». *International Conference on Biometrics (ICB)*, 1-6. IEEE.
<http://doi.org/10.1109/ICB45273.2019.8987375>

Kwok, A. O. y Koh, S. G. (2020). «Deepfake: a social construction of technology perspective». *Current Issues in Tourism*, 1-5.
<https://doi.org/10.1080/13683500.2020.1738357>

Li, Y.; Chang, M. C. y Lyu, S. (2018). «In ictu oculi: Exposing AI created fake videos by detecting eye blinking». *IEEE International Workshop on Information Forensics and Security* (WIFS), 1-7. IEEE.

Lin, H. (2019). «The existential threat from cyber-enabled information warfare». *Bulletin of the Atomic Scientists*, 75 (4), 187-196.
<https://doi.org/10.1080/00963402.2019.1629574>

Liv, N. y Greenbaum, D. (2020). «Deep Fakes and Memory Malleability: False Memories in the Service of Fake News». *AJOB Neuroscience*, 11 (2), 96-104.
<https://doi.org/10.1080/21507740.2020.1740351>

Mackenzie, A. y Bhatt, I. (2018). «Lies, Bullshit and Fake News: Some Epistemological Concerns». *Postdigital Science and Education*.
<https://doi.org/10.1007/s42438-018-0025-4>

Maras, M. H. y Alexandrou, A. (2019). «Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos». *International Journal of Evidence & Proof*, 23 (3), 255-262.
<https://doi.org/10.1177/1365712718807226>

Morozov, E. (2013). «To save everything, click here: The folly of technological solutionism». *Public Affairs*.

Pérez, J.; Meso, K. y Mendiguren, T. (2021). «Deepfakes on Twitter: Which Actors Control Their Spread?». *Media and Communication*, 9 (1), 301-312.
<http://dx.doi.org/10.17645/mac.v9i1.3433>

Qayyum, A.; Qadir, J.; Janjua, M. U. y Sher, F. (2019). «Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News». *IT Professional*, 21 (4), 16-24.
<https://doi.org/10.1109/MITP.2019.2910503>

Radford, A.; Metz, L. y Chintala, S. (2015). «Unsupervised representation learning with deep convolutional generative adversarial networks». *arXiv preprint* arXiv:1511.06434.

Rössler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C., Thies, J. y Niessner, M. (2018). «Faceforensics: A large-scale video dataset for forgery detection in human faces». *arXiv preprint* arXiv:1803.09179.

Vizoso, A.; Vaz-Álvarez, M. y López-García, X. (2021). «Fighting Deepfakes: Media and Internet Giants' Converging and Diverging Strategies Against Hi-Tech Misinformation». *Media and Communication*, 9 (1), 291-300.
<http://dx.doi.org/10.17645/mac.v9i1.3494>

Wagner, T. L. y Blewer, A. (2019). «The Word Real Is No Longer Real: Deepfakes, Gender, and the Challenges of AI-Altered Video». *Open Information Science*, 3 (1), 32-46.
<https://doi.org/10.1515/opis-2019-0003>

Westerlund, M. (2019). «The Emergence of Deepfake Technology: A Review». *Technology Innovation Management Review*, 9 (11), 39-52.
<http://doi.org/10.22215/timreview/1282>

Whyte, C. (2020). «Deepfake news: AI-enabled disinformation as a multi-level public policy challenge». *Journal of Cyber Policy*, 5 (2), 199-217.
<https://doi.org/10.1080/23738871.2020.1797135>