

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



Recommended citation: SÁNCHEZ-CASTILLO, S. and LÓPEZ-OLANO, C. (2021). “Vetting and verifying multimodal false information. A challenge for democratic societies”. *Anàlisi: Quaderns de Comunicació i Cultura*, 64, 1-7. DOI: <<https://doi.org/10.5565/rev/analisi.3461>>

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

* This monograph has been carried out in the framework of the consolidated R&D project entitled “Ecología de la desinformación: la construcción de las noticias falsas y su impacto en el espacio público” (AICO2020/224), project financed by Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital of the Generalitat Valenciana (2020-2021).

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*, *Neutral*, *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](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 médiatique 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>>.