COGITATIO

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 1, Pages 291–300
DOI: 10.17645/mac.v9i1.3494

Article

Fighting Deepfakes: Media and Internet Giants' Converging and Diverging Strategies Against Hi-Tech Misinformation

Ángel Vizoso *, Martín Vaz-Álvarez and Xosé López-García

Faculty of Communication Sciences, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain;
E-Mails: angel.vizoso@usc.es (A.V.), martin.vaz.alvarez@usc.es (M.V.-A), xose.lopez.garcia@usc.es (X.L.-G.)

* Corresponding author

Submitted: 22 July 2020 | Accepted: 15 October 2020 | Published: 3 March 2021

Abstract
Deepfakes, one of the most novel forms of misinformation, have become a real challenge in the communicative environment due to their spread through online news and social media spaces. Although fake news have existed for centuries, its circulation is now more harmful than ever before, thanks to the ease of its production and dissemination. At this juncture, technological development has led to the emergence of deepfakes, doctored videos, audios or photos that use artificial intelligence. Since its inception in 2017, the tools and algorithms that enable the modification of faces and sounds in audiovisual content have evolved to the point where there are mobile apps and web services that allow average users its manipulation. This research tries to show how three renowned media outlets—*The Wall Street Journal*, *The Washington Post*, and *Reuters*—and three of the biggest Internet-based companies—Google, Facebook, and Twitter—are dealing with the spread of this new form of fake news. Results show that identification of deepfakes is a common practice for both types of organizations. However, while the media is focused on training journalists for its detection, online platforms tended to fund research projects whose objective is to develop or improve media forensics tools.

Keywords
deepfake; Facebook; fact-checking; fake news; information verification; Google; misinformation; social media; Twitter

Issue
This article is part of the issue "Disinformation and Democracy: Media Strategies and Audience Attitudes" edited by Pere Masip (University Ramon Llull, Spain), Bella Palomo (University of Málaga, Spain) and Guillermo López (University of Valencia, Spain).

1. Introduction

The implementation of artificial intelligence in technologically mediated communicative processes in the networked society poses new challenges for journalistic verification. Simultaneously, it has enhanced different stages of news production systems. The effects of the technology that houses artificial intelligence are present both in the communicative flows and in a large part of the socialization dynamics. Hence, the threat introduced by the emergence of deepfakes, doctored videos by using artificial intelligence, arises as one of the most recent hazards for journalistic quality and news credibility. Although deepfakes are not only a concern for journalism, their existence has raised the uncertainty among users when trying to access news content. Likewise, the increasing sophistication of this form of fake news has put professionals on alert (Vaccari & Chadwick, 2020).

Misinformation has increased its relevance over the last few years, having now a major significance in the public agenda (Vargo, Guo, & Amazeen, 2018). In consequence, the number of projects and measures for counteracting this phenomenon has grown considerably. Example of this could be the *Action Plan Against Disinformation* developed by the European Commission (2018). Media and journalists are aware about how the

success of hoaxes undermines democracy and its reliability (Geham, 2017). Therefore, they try to react with actions that facilitate transparency and the fulfilment of their professional and ethical rules, like fact checking (Lowrey, 2017). This is an issue on which different lines of thinking have been opened. All of them try to counterbalance the result of misinformative political and social trends that became significative in 21st century societies (McNair, 2017) in a context where social media plays a central role as a space for the generation and dissemination of fake news and the consequences that this entails (Nelson & Taneja, 2018).

Techniques that guarantee the information verification's efficiency—one of the core elements of journalism since its consolidation as a communicative technique in the modern age (Kovach & Rosenstiel, 2014)—are looking inside technological innovation for tools with the ability to support professionals in their daily tasks. It is true that the norms followed for producing accurate informative pieces are in some cases unclear and nuanced (Shapiro, Brin, Bédard-Brûlé, & Mychajlowycz, 2013). Nonetheless, journalism should not retain antiquated verification techniques, but should rather update them to computational methods in order to evaluate dubious information (Ciampaglia et al., 2015). There are currently revamped verification systems with fact-checking techniques. Those may contribute to the elaboration of news pieces that, after the application of a complex group of cultural, structural, and technological relations would show the legitimation of news in the digital age (Carlson, 2017, p. 13). Although a high level of mistrust remains, some techniques used in these information verification services are able to build a bigger reliance by the users (Brandtzaeg & Følstad, 2017).

Furthermore, this scenario has seen the emergence of new proposals for renewed professional practices and profiles (Palomo & Masip, 2020). This could be the case of constructive journalism, whose objective is regaining the lost trust of the media (McIntyre & Gyldensted, 2017). This is a journalistic movement that explores new paths. However, it will take time to see if these new approaches fit in the frame of emerging journalism in the Third Millennium, with a clear commitment to social service, transparency and accuracy.

2. Literature Review

2.1. Misinformation Through the Ages

Falsehood, fantasy and fake news have walked along with the development of communication and journalism, initiating discussions about its practice and its role in society (McNair, 2017). Although there are evidence of misinformation since the Roman Empire (Burkhardt, 2017), its major development took place with the invention of the print in the 15th century. The possibility of disseminating written information in a faster and easier way made possible the circulation of falsehood too.

Hence, the advent of new means of communication, increased the presence of deliberated false content, not always with harmful purposes. In this regard, one of the greatest examples of misinformation of our times was the radio broadcast of *The War of the Worlds* directed by Orson Welles on October 30th, 1938. That radio show was followed by thousands of listeners, and some of them believed that the Earth was under an alien attack, thanks to the narration of Welles' cast (Gorbach, 2018).

This radio show wanted to entertain the audience using an alteration of reality. However, manipulation of the truth has been used as a weapon in military conflicts over the centuries in order to ascribe malicious acts or characteristics to the enemy (Bloch, 1999, p. 182). A good example of this use of misinformation was the sequence of news published after the explosion in the boilers of the United States Navy ship USS Maine on February 15th, 1898. In the middle of the fight for being the most read against Joseph Pulitzer's *The New York World*, William Randolph Hearst, editor of the *New York Journal*, sent a journalist to Cuba with the objective of telling the readers the details of a Spanish attack to this ship. Thus, when the correspondent arrived at the island reported that alleged attack did not exist. Nonetheless, the newspaper published a series of stories detailing the attack—even when they knew they were not accurate—causing a climate of hate against Spain and acceptance of the coming war. Finally, The United States declared the war against Spain (Amorós, 2018, p. 34). After this conflict, misinformation continued to be used against the enemy in war times. Thus, it is possible to identify strategies of its use in recent conflicts like the World War I and World War II, the Vietnam War or the Gulf War (Peters, 2018).

2.2. Fake News as a Threat to Journalism

Falsity has cast a shadow over the discipline of communication throughout history. One of the newest forms of misinformation is fake news, pieces that imitate the appearance of journalistic information, but deliberately altered (Rochlin, 2017). This form of deception has coexisted with true news. However, the current communicative scenario, marked by the utilisation of high speed and low contrast means of communication—and among all social media—provides a fertile soil for the dissemination of any form of misinformation (Lazer et al., 2018).

Platforms like Facebook or Twitter are now among the primary news sources for Internet users (Bergström & Jervelycke-Belfrage, 2018). Fake news producers are aware of this fact. As a result, they have made the web the main channel for false content distribution, taking advantage of the possibility of communicating anonymously provided by certain spaces (Vosoughi, Roy, & Aral, 2018). Furthermore, fake news producers have the chance of reaching as large audiences as consolidated journalistic brands (Fletcher, Cornia, Graves, & Nielsen, 2018), which makes the verification of this falsities more difficult.

During the last few years, there have been different proposals for classifying fake news. Among them, the one developed by Tandoc, Lim, and Ling (2018) is perhaps the most exhaustive: news satire, a very common form of fake news with a large presence in magazines, websites and radio or TV shows; news parody, which shares some of the characteristics of news satire, but it is not based on topical issues. These pieces are fictional elements specifically produced for certain purposes; news fabrication, unfounded stories that try to imitate the structure of news published by legacy media. The promoters of these pieces try to deceive by blending them among the truthful ones; photo manipulation—alteration of images—and more recently videos—for building a different reality; advertising and public relations—dissemination of advertising by masking it to look as journalistic reporting; and propaganda—stories from political organizations with the objective of influencing citizens' opinion on them. Like some of the previous ones, they imitate the formal structure of news pieces.

Regarding its formal structure, fake news try to imitate news items' formal appearance. Thus, visual codes and elements like headlines, images, videos hypertext and texts conceived like journalistic pieces are common features of this misinformation strategies (Amorós, 2018, p. 65). Nonetheless, its major particularity is that fake news tries to attack the readers' previous opinion, especially on controversial issues related to racism, xenophobia, homophobia and other forms of hate (Bennett & Livingston, 2018; Waisbord, 2018). This connection makes possible the rapid replication of such content thanks to the ease of sharing through spaces like social media platforms. Thus, episodes like electoral processes (Lowrey, 2017), or more recently the Covid-19 pandemic (Salaverría et al., 2020; Shimizu, 2020), resulted in a deep growth of fake news circulation, at times using simple methods but at times taking advance of the most advanced technology.

2.3. Deepfake: A Novel Form of Fake News

Deepfakes, a combination of 'deep learning' and 'fake' (Westerlund, 2019), are "highly realistic and difficult-to-detect digital manipulations of audio or video" (Chesney & Citron, 2019). It can be defined as "a technique used to manipulate videos using computer code" (Fernandes et al., 2019, p. 1721), generally replacing the voice or the face of a person with the face of the voice of another person. Although the photo and video manipulation have existed for a long time, the use of artificial intelligence methods for these purposes has augmented the number of fakes and its quality. Some of these videos are humorous, but the majority of them are damaging (Maras & Alexandrou, 2019). Hence, this is a recent movement whose beginnings date back to 2017, starting then a rapid popularisation until now (Deeptrace Labs, 2018, pp. 2–4).

This technique is the result of using Generative Adversarial Networks, algorithms designed to replace human faces or voices in thousands of images and videos in order to make them as realistic as possible (Li, Chang, & Lyu, 2018). The main advantage of these algorithms is that these systems are learning how to improve themselves by creating deepfakes. Therefore, future creations will be improved thanks to past experiences. This feature makes this misinformation procedure more dangerous, especially due to the emergence of mobile apps and computer programmes that allow users without computer programming training to produce deepfakes (Nirkin, Keller, & Hassner, 2019; Schwartz, 2018).

Farid et al. (2019, pp. 4–6) tried to label the different forms adopted by deepfakes in four categories: 1) face replacement or face swapping—this method involves changing one person's face, the source, for another one, the target; 2) face re-enactment—manipulation of the features of the features of one person's face like the movement of the mouth or the eyes, among others; 3) face generation—creation of a completely new face using all the potential provided by Generative Adversarial Networks; and 4) speech synthesis—alteration of someone's discourse in terms of cadence and intonation, or generation of a completely new one.

As with other technologies, the same algorithms used for creating deepfakes could have a beneficial application in the field of psychology, building digital synthetic identities for voiceless users; or in robot sketches through advanced facial recognition for law enforcement, for example (Akhtar & Dasgupta, 2019; Zhu, Fang, Sui, & Li, 2020). Notwithstanding, its use seems to be more harmful than beneficial nowadays with examples of the use of these technologies in acts of fraud and crime (Stupp, 2019).

Hence, one of the biggest challenges of deepfakes is to find out how to counteract them knowing that the debunking methods' development is always late regarding the production of misinformation (Galston, 2020). However, a great deal of effort has been made—and is still made—to develop technology-based tools for detecting and correcting it, both from public and private organizations (Deeptrace Labs, 2018, p. 2). These tools will be helpful in almost all areas of communication, especially for journalism.

2.4. Fact-Checking: Journalism's Response to the Misinformation Wave

In light of the above, verified information seems to be a necessity in our communicative context (Ekström, Lewis, & Westlund, 2020), especially because disruptive episodes like the coronavirus outbreak resulted in a clear increase of citizens' informative consumption (Masip et al., 2020). Furthermore, political communication has shifted to a model in which political leaders share their messages online instead of doing it through traditional media (López-García & Pavía, 2019).

At this juncture, the media has increased the importance of verification processes for correcting both internal and external errors (Geham, 2017). Consequently, a new professional profile—the fact-checker—has emerged with the mission of debunking misinformation and prevent audiences of its consumption. These professionals try to go to the origin of an information or a claim for gathering all the available data and contrasting it (Graves, 2016, p. 110). Fortunately, journalists have also benefited from the development of new technological tools designed for verifying images, videos or websites in an efficient manner (Brandtzaeg, Lüders, Spangenberg, Rath-Wiggins, & Følstad, 2016).

Although verification has always been part of any journalistic process, the rapid growth of the fake news phenomenon over the past few years made this activity more important than ever. Thus, the census created by the University of Duke Reporters' Lab counts now almost 300 fact-checkers in more than 60 countries by the by middle of 2020, a hundred more than on the same date in 2019 (Stencel & Luther, 2020). Regarding this, it is possible to talk about fact-checking as a transnational movement (Graves, 2018) were both legacy and independent media organizations try to restore the trust lost by the media (Bennett & Livingston, 2018).

3. Method

The starting point of this research will be the application of the Systematic Literature Review method (Kitchenham, 2004) as a method to set an approach on how deepfakes are being addressed and studied. Due to the novelty of this reality, this method will let us understand in an exhaustive way (Codina, 2017) what are researchers doing to assess this phenomenon and what efforts are being done to stop its spread.

Hence, our method consisted in the following phases: 1) topic identification—'deepfake' and 'deep fake'—and the period of analysis—all the available literature; 2) source selection—Web of Science's SCI-Expanded, SSCI, CPCI-S, CPI-SS, CPCI-SSH, and Scopus; 3) search in databases—the selection of Web of Science and Scopus is justified by the importance of these two databases, which contain the most relevant contributions for the Social Sciences field in general and deepfakes specifically; and 4) identification of the studied variables for each item—descriptive data (article title, date, journal or conference, number of authors, and keywords), type of study, research techniques (observation, survey, interview, content analysis, case study, experimental or non-specified), principal contribution, DOI or URL, and institution and country.

This search resulted in 54 different research items: 28 presented at international conferences and 26 published in academic journals—all of that after deleting duplicities and texts that did not fit the criteria, such as editorial articles, call for papers, or interviews, among others. These 54 examples comprise our sample that will be addressed in the next section in order to understand the path followed by researchers on this subject.

Concerning the second stage of our study, it will analyse the approach taken by three renowned media outlets and news agencies—*The Washington Post*, *The Wall Street Journal*, and *Reuters*—and three of the most important Internet platforms—Google, Facebook and Twitter—in neutralizing the spread of deepfakes. Thus, case study of these six organizations will be applied in order to understand how they are managing to identify, label and notify deepfakes through different approaches—protocols, use of technology, collaboration with institutions, and funding of innovative projects. This will be done through the analysis of the available reports and statements of these six organizations.

Consequently, the main goal of our study will be to identify the coincidences and disparities in the strategy of three major media outlets and three of the most important online platforms when trying to stop the diffusion of deepfakes. This will be relevant in order to understand if six of the main representatives from these two communicative fields are joining efforts and strategies in limiting or not its spread, and how these procedures could be improved.

4. Findings

4.1. Results of the Systematic Literature Review

We will start by depicting the state of the research on deepfakes, especially the contributions indexed in the two main databases—Web of Science and Scopus. As shown in Table 1, research about this issue started in 2018 with four conference papers. However, it was quintupled in 2019, and during the first half of 2020 almost a half more of works on deepfakes than the previous year were published. Furthermore, the most salient element of this table is that this form of fake news used to have presence at conferences, but in 2020 they become a topic addressed in academic journals too. Nonetheless, it is necessary to note that the situation resulting from the Covid-19 pandemic

Table 1. Evolution of the studies on deepfake indexed in WoS and Scopus.

Year	Conference paper	Journal article	Total
2018	4	0	4
2019	16	5	21
2020 (1st half)	8	21	29

has provoked the cancellation or postponement of many conferences.

Regarding the authorship of this research, the most common approach is the participation of three authors. Thus, the arithmetic mean—3,13 authors—and the mode—14 articles have three authors—serve to confirm this. Also concerning the authorship, researchers from 24 countries were identified, most of them located in the United States and Asia—China, Japan, South Korea, India, or Taiwan.

Finally, this review shows a large degree of uniformity concerning the type of studies published on deepfakes. Almost all the reviewed articles and conference papers take a descriptive approach. This is because 32 of the items are the result of experimenting with new tools and algorithms to counteract it. Another important group of research is review articles on deepfake detection and prevention or even about legal framework and legal concerns of this form of misinformation, something that was found 21 times.

In sum, the novelty of deepfake implies a certain degree of youth for its research. At present, it is possible to see two trends: Studies that present new forms to stop its spread, or studies that try to create context on its emergence and development.

4.2. Counteracting Deepfakes at The Wall Street Journal*,* The Washington Post *and* Reuters

Recent advances in artificial intelligence and their democratisation have allowed average users to create deepfakes. This represents a major challenge for our society due to the potential harmful impact of these creations, especially before electoral processes. Looking to the United States 2020 general election, *The Wall Street Journal* has created a division of 21 journalists whose unique objective is detecting, labelling and debunking misinformation, particularly deepfakes (Southern, 2019). This team is a joint effort of Standards & Ethics and R&D departments, and this work is very linked to the use of technology with presence of journalist with video, photo, visuals, research and news experience that have been trained for deepfake detection (Marconi & Daldrup, 2018). Furthermore, *The Wall Street Journal* provides specialized training in fake news and deepfake identification in partnership with different researchers. This has led to the development of a protocol to find examples of this kind of misinformation with three stages: source examination (contact with the source, authorship identification, and metadata check, among others), search for older versions of the footage available online, and footage examination with video and photo editing programs.

Meanwhile, *The Washington Post* has applied to deepfake detection very similar criteria to other fake news detection. Thus, the *The Washington Post* has added video experts to the tasks developed by the team led by Glenn Kessler—also known as 'The Fact Checker' (Kessler, 2019). The most important contribution of this publication regarding this problem is the elaboration of a taxonomy to classify and label deepfakes. *The Washington Post* was also pioneering in the use of scales to highlight the degree of truth and lie of any content. Regarding doctored videos, the newspaper sets out three categories of manipulation (Ajaka, Samuels, & Kessler, 2019): missing context (presentation of the video without context or with a context intentionally altered), deceptive editing (rearrangement and edition of the video in certain parts or details), and malicious transformation (complete manipulation and transformation of the footage resulting in a completely new fabrication).

A third approach to this reality could be the one adopted by the news agency *Reuters*. The news services provider reports its awareness and concern on the deepfake spread (Crosse, n.d.). Hence, it has started a collaboration with Facebook for detecting as much doctored user-generated content as possible among all the videos and photos that run on the platform (Patadia, 2020).

In this regard, *Reuters* has started a blog whose objective is verifying doctored materials in English and Spanish. All of that with the objective of debunking as much information as possible ahead of the 2020 United States election. This is a clear example of the emerging collaboration among technological platforms and the media, a joint effort in trying to stop the rapid growth of fake news and deepfakes in such significant moments like a presidential run-up.

4.3. Internet Giants' Strategies Against Deepfakes

The spread of falsehood through social media platforms and other Internet spaces is now a challenge for providers like Google, Facebook or Twitter. As a result, over the last few months they have started different initiatives whose unique objective is finding efficient ways to detect and stop the misinformation and, more recently, deepfakes.

Regarding this, these three companies show different approaches against this matter. Google, for instance, has made available to the research community a large set of manipulated and non-manipulated videos (Dufour & Gully, 2019). With this initiative, they want to help in the development of identification techniques by taking advantage of the great amount of information saved in their files. In addition, they collaborate with the Defense Advanced Research Projects Agency to fund different researchers that are developing media forensic tools.

On the other hand, Facebook is financing different research projects within its 'Deepfake Detection Challenge.' This initiative, boosted by companies like Facebook, Microsoft and Amazon Web Services and research units from various universities across the United States, tries to assist researchers that are working on the development of artificial intelligence-based deepfake detection tools. Thus, a corpus of more than 100,000 videos was available to these researchers that

fight for presenting useful mechanism in order to win different awards.

Furthermore, Mark Zuckerberg's social network tries to counteract this form of misinformation by deleting doctored videos or photos, or labelling it as fake news with the help of fact-checking media outlets (Bickert, 2020). This is particularly important for those related to the 2020 United States run-up due to the influence that fake news could have in this process.

Finally, Twitter shows a simpler approach towards this problem. They summarize their strategy in the following four rules (Harvey, 2019): Identification through a notice of Tweets with manipulated content, warning of its manipulated condition before sharing it, inclusion of a link to news articles or other verified sources in which users can find out why and how the content has been doctored, and elimination of all that manipulated content potentially harmful or threatening to anyone's safety.

These diverging strategies on behalf of the major online platforms are in part the product of self-regulated methods for fighting deepfakes, as there is still incipient intervention on behalf of the states in regulating content on social media and other outlets. The question to be asked here is whether it is the online platforms' sole responsibility to tackle misinformation or if there are any social interests in this situation for which other public entities should allocate resources to.

The European Commission already pointed out in 2018 the need for governments to invest in research and detection of misinformation, while also prompting these to hold social media companies accountable (European Commission, 2018). So far, in the last two years the EU has launched a series of initiatives to tackle the issue: a code of practice against disinformation, the creation of the Social Observatory for Disinformation and Social Media Analysis and the set-up of the Rapid Alert System, among other R&D projects such as PROVENANCE, SocialTruth, EUNOMIA or WeVerify (European Commission, 2020). Despite a lot of efforts being made to avoid the spread of misinformation in the EU, deepfakes are still not as much on the agenda as other academics are asking for, while also describing their worry for seemingly understaffed programs (Bressnan, 2019). Measures taken by countries to prompt social media companies in acting against fake news contain different levels of intervention and are mainly dedicated to counteracting disinformation in political advertisement. France and Germany, for example, require online platforms to establish an easily accessible and visible way for users to flag false information, while Australia requires all paid electoral advertising, including advertisements on social media, to be authorized and to contain an authorization statement (Levush, 2019). In the United States, some states have already taken specific measures to counter deepfakes, although these are still merely reactive and not preventive, such as Texas passing a law that criminalizes publishing and distributing deepfake videos with the intention to harm a candidate during the electoral process; or California, where a law was passed last October making it illegal for anyone to intentionally distribute deepfakes for deceiving voters or perjure a candidate (Castro, 2020).

The implications of these incipient interactions between governments and social media companies might have relevant governance questions in the forthcoming years, all the while these companies are also starting to take new approaches to their governance structures, such is the case of the Facebook, who set up the Independent Oversight Board, which "aims to make Facebook more accountable and improve the decision-making process," in the words of Nick Clegg, currently Facebook's VP of Global Affairs and Communications and former Deputy Prime Minister of the United Kingdom (Moltzau, 2020).

5. Discussion

As shown in the previous section, the media and Internet platforms have initiated different strategies to fight misinformation and, more particularly, the spread of deepfakes. In this regard, there are some similarities and differences among the strategies of these two communicative sectors.

First of all, it seems clear that the collaboration among platforms and media outlets increases over time. Example of this could be the agreements among *Reuters* and Facebook whose objective is to detect fake news and share its correction. Furthermore, other fact-checking organizations collaborate with this social network in labelling false content and warn users about this.

Another coincidence is the use of technology as a weapon in the battle against deepfakes. Both news media and digital platforms have understood that high technology and the use of algorithms as powerful as those used for creating fakes is the only chance to counteract them. Thus, media outlets are increasingly training journalists and interdisciplinary teams in the use of these mechanisms that allow them to identify this form of misinformation.

The third match could be the growing synergies between the academic and communicative sides. Thus, media outlets and platforms try to collaborate with researchers and institutions specialized in fake news detection, both in training and to apply their methods.

Regarding the divergences when dealing deepfakes, online platforms are able to fund research projects whose objective is developing artificial intelligence-tools for identifying this form of fake news. The media, however, does not have such possibilities due to the expenditure of these activities.

Another difference in dealing with this issue could be that the media use to correct misinformation instead deleting it. As shown before, some of the social media platforms have the elimination of doctored content among their strategies. This presents a clear challenge. Although deleting manipulated videos or photos ends

with the problem for future or potential users, does not for those users that have seen them. In contrast, labelling these materials as false or manipulated—the approach followed by verification media outlets—could be helpful for future users.

6. Conclusions

Deepfakes have become a reality in our communicative system. Media outlets and Internet services providers try to counteract it with different outlooks. However, the development of the techniques for producing misinformation seems to advance faster than those for debunking it. Regarding this, the available research is mainly focused on two aspects: On one hand detection tools, and on the other hand, the implications of this form of fake news for democracy and national security. The fact that so far only big technological giants are capable of introducing hi-tech expensive solutions for fighting deepfakes motivates that the available mass of research on the subject is fundamentally dedicated to address the questions raised by these corporations which are mainly technological. On the other hand, journalism focused media, which are not able to invest large amounts of money on deepfake detection are therefore unable to push their concerns into the research agenda. For this reason, producing research on the implications of deepfakes for journalism and under journalistic premises presents itself as elemental, as well as further investigation on how the media trains its professionals for detecting advanced misinformation.

The novelty of this deceiving technique provokes its understudied situation, but the constant growth of works on this matter show that it will be an important field for researchers on misinformation and media forensics in the following years. However, the study is able to show to some extent that the media and digital platforms' have notable similarities and differences when it comes to their strategies. This could be due to the different nature of their business models, but nevertheless sometimes it seems to be a matter of investment. Digital platforms have joined efforts with technological, academic or entrepreneurial partners, spending large amounts of money in this field, which is something that many media outlets cannot accomplish.

However, these collaborations go in some cases beyond the private sector. The growing pressure on behalf of states and governments, pushing for different levels of regulation for these social media companies is being noted around the world. Countries have started to put mechanisms in place that allow for a more scrutinized assessment of these corporations, since there is an increasingly wider knowledge on how their inner social interaction tools can affect basic democratic elements (elections, political advertisement, etc.). A representative example of this is Facebook's recent signing of Nick Clegg, former Deputy Prime Minister of the United Kingdom, in a move that seems to state its willingness to reform its governance structures towards more state-sensible policies. The tendency seems to describe a tension between the states and governments' purpose to regulate, and the effort of these companies to maintain their self-regulation.

Nonetheless, it is necessary to highlight the limitations of our study and our sample. As the reader may have noticed, our sample shows a western-oriented cultural bias, both for the media and digital platforms analysis. This could make it difficult to generalize our findings to similar organizations from other territories. However, this research presents a descriptive approach whose objective is to analyse how major communication companies are trying to counteract the spread of deepfakes, the most novel and hi-tech-based form of misinformation. Furthermore, the novelty and high level of sophistication of this new way of producing fake news requires a similar level of human and technological resources for the debunking process, something that only large companies—mainly United States-based—can afford at this particular time. In the future, it will be interesting to follow the progress of both deepfakes production and identification strategies as well as the paths adopted by researchers in this subject. As mentioned, this is a recent phenomenon, but its rapid growth will make necessary the setting up of protocols for containing its spread in both media outlets and online platforms.

Future lines of research regarding this topic might include a deep assessment of how journalists perceive deepfakes in their daily routines, and the challenge of verification for journalists, which includes questions of technological literacy and guides for good practices.

Acknowledgments

This article has been developed within the research projects "Digital Native Media in Spain: Storytelling Formats and Mobile Strategy (RTI2018–093346-B-C33)" and "New Values, Governance, Funding and Public Media Services for the Internet Society: European and Spanish Contrasts (RTI2018–096065-B-100)" funded by the Ministry of Science, Innovation and Universities (Government of Spain) and co-funded by the European Regional Development Fund (ERDF). Furthermore, authors Ángel Vizoso and Martín Vaz-Álvarez are also beneficiary of the Training University Lecturers' Program (FPU), funded by the Spanish Ministry of Science, Innovation and Universities (Government of Spain).

Conflict of Interests

The authors declare no conflict of interests.

References

Ajaka, N., Samuels, E., & Kessler, G. (2019). The Washington Post's guide to manipulated video. *The Washington Post.* Retrieved from https://www.

washingtonpost.com/graphics/2019/politics/fact-checker/manipulated-video-guide

Akhtar, Z., & Dasgupta, D. (2019). A comparative evaluation of local feature descriptors for deepfakes detection. In *2019 IEEE international symposium on technologies for homeland security, HST 2019*. Woburn, MA: IEEE. https://doi.org/10.1109/hst47167.2019.9033005

Amorós, M. (2018). *Fake news: La verdad de las noticias falsas* [Fake news: The truth of false news]. Barcelona: Plataforma Editorial.

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. https://doi.org/10.1177/0267323118760317

Bergström, A., & Jervelycke-Belfrage, M. (2018). News in social media. *Digital Journalism*, *6*(5), 583–598. https://doi.org/10.1080/21670811.2018.1423625

Bickert, M. (2020, January 6). Enforcing against manipulated media. *Facebook*. Retrieved from https://about.fb.com/news/2020/01/enforcing-against-manipulated-media

Bloch, M. (1999). *Historia e historiadores* [History and historians]. Madrid: Akal.

Brandtzaeg, P. B., & Følstad, A. (2017). Trust and distrust in online fact-checking services. *Communications of the ACM*, *60*(9), 65–71. https://doi.org/https://doi.org/10.1145/3122803

Brandtzaeg, P. B., Lüders, M., Spangenberg, J., Rath-Wiggins, L., & Følstad, A. (2016). Emerging journalistic verification practices concerning social media. *Journalism Practice*, *10*(3), 323–342. https://doi.org/10.1080/17512786.2015.1020331

Bressnan, S. (2019, September 19). Can the EU prevent deepfakes from threatening peace? *Carnegie Europe*. Retrieved from https://carnegieeurope.eu/strategiceurope/79877

Burkhardt, J. M. (2017). Chapter 1: History of fake news. *Library Technology Reports*, *53*(8), 5–9.

Carlson, M. (2017). *Journalistic authority: Legitimating news in the digital era*. New York, NY: Columbia University Press.

Castro, D. (2020). Deepfakes are on the rise: How should government respond? *Government Technology*. Retrieved from https://www.govtech.com/policy/Deepfakes-Are-on-the-Rise-How-Should-Government-Respond.html

Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war. *Foreign Affairs.* Retrieved from https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war

Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., & Flammini, A. (2015). Computational fact checking from knowledge networks. *PLOS ONE*, *10*(6). https://doi.org/10.1371/journal.pone.0128193

Codina, L. (2017, April 20). Revisiones sistematizadas y cómo llevarlas a cabo con garantías: Systematic reviews y SALSA Framework [Systematized reviews and how to conduct it with guarantees: Systematic reviews and SALSA Framework]. *Lluiscodina*. Retrieved from https://www.lluiscodina.com/revision-sistematica-salsa-framework

Crosse, G. (n.d.). Truth and authentication: The frightening world of video fakes. *Reuters.* Retrieved from https://agency.reuters.com/en/insights/articles/articles-archive/truth-and-authentication-the-frightening-world-of-video-fakes.html

Deeptrace Labs. (2018). *The state of deepfakes: Reality under attack*. Amsterdam: Deeptrace Labs. Retrieved from https://s3.eu-west-2.amazonaws.com/rep2018/2018-the-state-of-deepfakes.pdf

Dufour, N., & Gully, A. (2019, September 24). Contributing data to deepfake detection research. *Google Blog.* Retrieved from https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html

Ekström, M., Lewis, S. C., & Westlund, O. (2020). Epistemologies of digital journalism and the study of misinformation. *New Media & Society*, *22*(2), 205–212. https://doi.org/10.1177/1461444819856914

European Commission. (2018). *Action plan against disinformation*. Brussels: European Commission. Retrieved from https://eeas.europa.eu/sites/eeas/files/action_plan_against_disinformation.pdf

European Commission. (2020). *Tackling online disinformation*. Brussels: European Commission. Retrieved from https://ec.europa.eu/digital-single-market/en/tackling-online-disinformation

Farid, H., Davies, A., Webb, L., Wolf, C., Hwang, T., Zucconi, A., & Lyu, S. (2019). *Deepfakes and audiovisual disinformation*. London: Centre for Data Ethics and Innovation. Retrieved from https://www.gov.uk/government/publications/cdei-publishes-its-first-series-of-three-snapshot-papers-ethical-issues-in-ai/snapshot-paper-deepfakes-and-audiovisual-disinformation

Fernandes, S., Raj, S., Ortiz, E., Vintila, I., Salter, M., Urosevic, G., & Jha, S. (2019). Predicting heart rate variations of deepfake videos using neural ODE. In *Proceedings: 2019 international conference on computer vision workshop, ICCVW 2019* (pp. 1721–1729). Seoul: IEEE. https://doi.org/https://doi.org/10.1109/iccvw.2019.00213

Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2018). *Measuring the reach of fake news and online disinformation in Europe*. Oxford: Reuters Institute for the Study of Journalism. Retrieved from https://reutersinstitute.politics.ox.ac.uk/our-research/measuring-reach-fake-news-and-online-disinformation-europe

Galston, W. A. (2020). *Is seeing still believing? The deepfake challenge to truth in politics*. Washington, DC: Brookings. Retrieved from https://www.brookings.edu/research/is-seeing-still-believing-the-deepfake-

challenge-to-truth-in-politics

Geham, F. (2017, July 12). *Le fact-checking: Une réponse à la crise de l'information et de la démocratie* [Fact-checking: The answer to the information and democracy crisis]. Paris: Fondapol. Retrieved from http://www.fondapol.org/etude/farid-gueham-le-fact-checking-une-reponse-a-la-crise-de-linformation-et-de-la-democratie

Gorbach, J. (2018). Not your grandpa's hoax: A comparative history of fake news. *American Journalism*, *35*(2), 236–249. https://doi.org/10.1080/08821127.2018.1457915

Graves, L. (2016). *Deciding what's true*. New York, NY: Columbia University Press.

Graves, L. (2018). Boundaries not drawn. *Journalism Studies*, *19*(5), 613–631. https://doi.org/10.1080/1461670X.2016.1196602

Harvey, D. (2019, November 11). Help us shape our approach to synthetic and manipulated media. *Twitter.* Retrieved from https://blog.twitter.com/en_us/topics/company/2019/synthetic_manipulated_media_policy_feedback.html

Kessler, G. (2019, June 25). Introducing The Fact Checker's guide to manipulated video. *The Washington Post.* Retrieved from https://www.washingtonpost.com/politics/2019/06/25/introducing-fact-checkers-guide-manipulated-video

Kitchenham, B. (2004). *Procedures for performing systematic reviews*. Keele: Keele University. Retrieved from http://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf

Kovach, B., & Rosenstiel, T. (2014). *The elements of journalism, revised and updated: What newspeople should know and the public should expect* (3rd ed.). New York, NY: Three Rivers Press.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., . . . Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Levush, R. (2019). Government responses to disinformation on social media platforms. *Library of Congress*. Retrieved from https://www.loc.gov/law/help/social-media-disinformation/compsum.php

Li, Y., Chang, M.-C., & Lyu, S. (2018). In ictu oculi: Exposing AI created fake videos by detecting eye blinking. In *2018 10TH IEEE international workshop on information forensics and security (WIFS)* (pp. 1–5). Hong Kong: IEEE. https://doi.org/10.1109/wifs.2018.8630787

López-García, G., & Pavía, J. M. (2019). Political communication in election processes: An overview. *Contemporary Social Science*, *14*(1), 1–13. https://doi.org/10.1080/21582041.2018.1479040

Lowrey, W. (2017). The emergence and development of news fact-checking sites. *Journalism Studies*, *18*(3), 376–394. https://doi.org/10.1080/1461670X.2015.1052537

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

Marconi, F., & Daldrup, T. (2018, November 15). How The Wall Street Journal is preparing its journalists to detect deepfakes. *NiemanLab.* Retrieved from https://www.niemanlab.org/2018/11/how-the-wall-street-journal-is-preparing-its-journalists-to-detect-deepfakes

Masip, P., Aran-Ramspott, S., Ruiz-Caballero, C., Suau, J., Almenar, E., & Puertas-Graell, D. (2020). News consumption and media coverage during the confinement by Covid-19: Information overload, ideological bias and sensationalism. *El Profesional de la Información*, *29*(3). https://doi.org/10.3145/epi.2020.may.12

McIntyre, K., & Gyldensted, C. (2017). Constructive journalism: An introduction and practical guide for applying positive psychology techniques to news production. *The Journal of Media Innovations*, *4*(2), 20–34. https://doi.org/https://doi.org/10.5617/jomi.v4i2.2403

McNair, B. (2017). *Fake news: Falsehood, Fabrication and Fantasy in Journalism*. London: Routledge.

Moltzau, A. (2020, January 9). What strategy does Europe have to tackle deepfakes? *Medium*. Retrieved from https://medium.com/dataseries/what-strategy-does-europe-have-to-tackle-deepfakes-fb159040f0c

Nelson, J. L., & Taneja, H. (2018). The small, disloyal fake news audience: The role of audience availability in fake news consumption. *New Media & Society*, *20*(10), 3720–3737. https://doi.org/10.1177/1461444818758715

Nirkin, Y., Keller, Y., & Hassner, T. (2019). FSGAN: Subject agnostic face swapping and reenactment. *Nirkin*. Retrieved from https://nirkin.com/fsgan

Palomo, B., & Masip, P. (2020). Journalistic reinvention for an automated and polarized scenario. In C. Toural Bran, Á. Vizoso, S. Pérez-Seijo, M. Rodríguez-Castro, & M.-C. Negreira-Rey (Eds.), *Information visualization in the era of innovative journalism* (pp. 161–170). New York, NY: Routledge.

Patadia, D. (2020, February 12). Reuters launches fact-checking initiative to identify misinformation, in partnership with Facebook. *Reuters*. Retrieved from https://www.reuters.com/article/rpb-fbfactchecking/reuters-launches-fact-checking-initiative-to-identify-misinformation-in-partnership-with-facebook-idUSKBN2061TG

Peters, M. A. (2018). The information wars, fake news and the end of globalisation. *Educational Philosophy and Theory*, *50*(13), 1161–1164. https://doi.org/10.1080/00131857.2017.1417200

Rochlin, N. (2017). Fake news: Belief in post-truth. *Library*

Hi Tech, *35*(3), 386–392. https://doi.org/10.1108/LHT-03-2017-0062

Salaverría, R., Buslón, N., López-Pan, F., León, B., López-Goñi, I., & Erviti, M.-C. (2020). Disinformation in times of pandemic: Typology of hoaxes on Covid-19. *El Profesional de la Información*, *29*(3). https://doi.org/https://doi.org/10.3145/epi.2020.may.15

Schwartz, O. (2018, November 12). You thought fake news was bad? Deep fakes are where truth goes to die. *The Guardian.* Retrieved from https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth

Shapiro, I., Brin, C., Bédard-Brûlé, I., & Mychajlowycz, K. (2013). Verification as a strategic ritual how journalists retrospectively describe processes for ensuring accuracy. *Journalism Practice*, *7*(6), 657–673. https://doi.org/10.1080/17512786.2013.765638

Shimizu, K. (2020, February 29). 2019-nCoV, fake news, and racism. *The Lancet*, *395*(10225), 685–686. https://doi.org/10.1016/S0140-*6736*(20)30357-3

Southern, L. (2019, July 1). 'A perfect storm': The Wall Street Journal has 21 people detecting 'deepfakes.' *Digiday*. Retrieved from https://digiday.com/media/the-wall-street-journal-has-21-people-detecting-deepfakes

Stencel, M., & Luther, J. (2020, June 22). Annual census finds nearly 300 fact-checking projects around the world. *Duke Reporters' Lab*. Retrieved from https://reporterslab.org/annual-census-finds-nearly-300-fact-checking-projects-around-the-world

Stupp, C. (2019, August 30). Fraudsters used AI to mimic CEO's voice in unusual cybercrime case. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news." *Digital Journalism*, *6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, *6*(1). https://doi.org/10.1177/2056305120903408

Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, *20*(5), 2028–2049. https://doi.org/10.1177/1461444817712086

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Waisbord, S. (2018). Truth is what happens to news. *Journalism Studies*, *19*(13), 1866–1878. https://doi.org/10.1080/1461670X.2018.1492881

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, *9*, 40–53. https://doi.org/http://doi.org/10.22215/timreview/1282

Zhu, B., Fang, H., Sui, Y., & Li, L. (2020). Deepfakes for medical video de-identification: Privacy protection and diagnostic information preservation. In *AIES 2020: Proceedings of the AAAI/ACM conference on AI, ethics, and society* (pp. 414–420). New York, NY: ACM. https://doi.org/https://doi.org/10.1145/3375627.3375849

About the Authors

Ángel Vizoso is a Researcher at Novos Medios Research Group (University of Santiago de Compostela) and Beneficiary of the Training University Lecturers´ Program (FPU) funded by the Spanish Ministry of Science, Innovation and Universities. (Government of Spain). His research lines are mainly information visualization, fact-checking and journalistic production for online media and he was Visiting Scholar at Universidade Nova de Lisboa (Portugal).

Martín Vaz-Álvarez is a Researcher at Novos Medios Research Group (University of Santiago de Compostela) and Beneficiary of the Training University Lecturers' Program (FPU) funded by the Spanish Ministry of Science, Innovation and Universities. (Government of Spain). His research lines are focused on co-creation in European public broadcasters, innovation and new Technologies.

Xosé López-García is a Professor of Journalism at University of de Compostela, PhD in History and Journalist. Coordinator of Novos Medios Research Group since 1994, whose research lines are focused on the study of digital and printed media, the analysis of the implications of technology for mediated communications, as well as the performance and funding of cultural industries or the history of communication, among others.