

Verification in the Covid-19 infodemic. Newtral's case study

Verificación en la infodemia de la Covid-19. El caso Newtral

Guadalupe Aguado-Guadalupe. University Carlos III of Madrid. Spain.

maguado@hum.uc3m.es

[CV]  

Itziar Bernaola-Serrano. University Carlos III of Madrid. Spain.

ibernaol@hum.uc3m.es

[CV]  

How to cite this article / Standardized reference

Aguado-Guadalupe, G. & Bernaola-Serrano, I. (2020). Verification in the Covid-19 infodemic. Newtral's case study. *Revista Latina de Comunicación Social*, 78, 289-308. <https://www.doi.org/10.4185/RLCS-2020-1478>

ABSTRACT

Introduction: This work describes Newtral's verification procedures in the infodemic that has occurred during the Covid-19 crisis and the characteristics of the verified content. **Methodology:** 205 verified pieces of content published between January 25 and May 25, 2020, are analyzed; verification processes are followed up and an unstructured personal interview with Newtral's head of contents is conducted. **Results:** The number of verified content increased considerably as of March 10, four days before the state of alarm was decreed. Regarding verification methods, sources were consulted in 90.5% of cases, while technological tools were only used in 9.5%. The most widespread were texts on social media, detecting 66% of cases in more than one social network. It was impossible to detect the origin in 91.7% of the verified pieces of contents. **Discussion and conclusions:** Consulting various sources is the most widely used verification method, far more than tools such as search engines or geolocation apps. Consultations of official and impersonated sources prevailed. The most common were fabricated and imposter content disseminated through social media and WhatsApp.

KEYWORDS: Covid-19; Newtral; verification; infodemic; disinformation; fake content; fact-checking.

RESUMEN

Introducción: En este trabajo se describen los procedimientos de verificación de Newtral en la infodemia acontecida durante la crisis de la Covid-19 y las características de los contenidos verificados. **Metodología:** Se analizan 205 contenidos verificados entre el 25 de enero y el 25 de mayo de 2020, se hace un seguimiento de los procesos de verificación y se realiza entrevista personal no estructurada con el director de contenidos de Newtral. **Resultados:** El número de contenidos verificados aumentó considerablemente a partir del 10 de marzo, cuatro días antes de decretarse el estado de alarma. En cuanto a los sistemas de verificación, en el 90,5% de los casos se recurrió a consulta de fuentes, empleándose herramientas solamente en el 9,5%. Lo más difundido fueron textos en redes sociales, detectándose el 66% de los casos en más de una red social. Fue imposible detectar el origen en el 91,7% de los contenidos contrastados. **Discusión y conclusiones:** La consulta

a diversas fuentes fue el método de verificación más utilizado, muy por delante de herramientas como motores de búsqueda o de geolocalización. Prevalcieron las consultas a fuentes oficiales y suplantadas. Lo más habitual fueron contenidos fabricados e impostores difundidos a través de redes sociales y WhatsApp.

PALABRAS CLAVE: Covid-19; Newtral; verificación; infodemia; desinformación; contenidos falsos; *fact-checking*.

CONTENT: 1. Introduction. 1.1. Fake content and verification. 2. Methodology and research questions. 3. Results. 3.1. Newtral's verification methods. 3.2. Evolution of verifications performed by Newtral over time. 3.3. Fake content formats. 3.4. Dissemination channels. 3.5. Origin of verified content. 3.6. Content typology. 3.7. Risks that verified fake content entails. 4. Discussion. 5. Conclusions. 6. References.

Translation by **Carlos Javier Rivas Quintero** (University of the Andes, Mérida, Venezuela)

1. Introduction

The first pieces of news in Spain about an atypical type of pneumonia, subsequently known as COVID-19, were published at the beginning of January 2020 (Aldama, 2020; De la Cal, 2020; Santirso, 2020). Since then, the importance of verification systems has increased as a result of the rise in fake content and disinformation originated during the COVID-19 crisis, generating what Tardáguila has described as “the worst wave of disinformation in history” (Valera, 2020). A clear example of this situation is that, during the pandemic, the number of consultations Newtral received through its on-demand verification service via WhatsApp registered a sixteen-fold increase. According to Joaquín Ortega, head of contents of that company, the entire fact-checking team had to be reorganized and focused on addressing the amount of requests from people who needed certainty in one of the moments of greatest uncertainty experienced in recent decades.

It is true there are previous experiences in this regard, such as the one lived due to the Ebola virus health crisis in 2014, during which a hoaxes dissemination trend was observed on the Internet, especially after the announcement of the first case detected outside Africa; a situation that generated what Jurado-Salván and Jurado-Izquierdo (2014) called “the Gadiana effect” of health information, since there was “first, crisis; next, health alert; then, disinformation” (Jurado-Salván and Jurado-Izquierdo, 2014, p.90).

Possibly, dissemination of fake content during this new health crisis has been boosted by a context in which the 8Ps or motivations for disinformation identified by First Draft (Wardle, 2017) exist: poor journalism, to parody, to provoke or to “punk”, passion, partisanship, profit, political influence, and propaganda. Added to this is the fact that false news diffuses “farther, faster, deeper, and more broadly” (Vosoughi, Roy and Aral, 2018).

This situation of overabundant information, its rapid spread and the disinformation originated during the pandemic, have caused the World Health Organization (WHO) to warn about the threat of rumors, hoaxes and fake data being disseminated, giving rise to what is now known as infodemic. This situation led WHO to activate a risk communication and infodemic management team to track information actively, whilst enabling a section on its website to refute the myths circulating on social media. For its part, the Spanish Government carried out the campaign “*Coronavirus, siempre fuentes oficiales*” [EN: Coronavirus, always official sources], to caution the population (Martín-Barato *et al.*,

2020). Additionally, WhatsApp, Twitter, and Facebook have created strategies to fight such false news. In this line, the International Fact-Checking Network (IFCN) launched a project to combat disinformation on WhatsApp during the pandemic, connecting users with more than 80 independent fact-checkers from 74 countries.

According to IFCN's associated director, Cristina Tardáguila, this novel coronavirus has been the biggest challenge fact-checkers have ever faced (Suárez, 2020). It can be asserted that the pandemic of hoaxes and disinformation around COVID-19 is global. More than 100 fact-checker teams from different countries, including Newtral's team, are working to combat this infodemic. Likewise, according to this organization's head of contents, Joaquín Ortega, Newtral collaborates in alliance with IFCN's international fact-checkers against disinformation on the coronavirus. Furthermore, thanks to efforts in collaboration with Facebook, they have managed to develop a chatbot so that WhatsApp users can find there the denials posted by all members of this alliance, both in English and Spanish.

This work analyzes the verification methods Newtral has been implementing during the COVID-19 health crisis and the characteristics of the verified content.

This startup, founded in 2018, has been selected since it is one of the Spanish fact-checker teams that obtained the certification of the International Fact-Checking Network -together with *Maldita.es*, *Efe Verifica* and *AFP España*-, and which has collaborated in a partnership with other IFCN members to combat disinformation globally during the pandemic. It is important to point out that all these fact-checkers have to undergo annual evaluations by the IFCN to verify they abide by a series of commitments embodied in its Code of Principles (<https://ifcncodeofprinciples.poynter.org/>), regarding their funding, nonpartisan and transparent methods and sources, so that they all base on common standards in their way of working, hence this study's results, addressing the verifications conducted by Neutral during the pandemic, can be largely extrapolated to other fact-checker teams.

This study is focused on the 205 false pieces of news that were verified and published by Newtral from January 25, 2020, the date when Newtral detected the first fake content related to the coronavirus, to May 25, 2020, the date when all the Spanish provinces came out of phase zero.

This study's objectives are:

1. To detail the verification parameters, methods and tools used by Newtral in fake content related to COVID-19.
2. To analyze the characteristics of the fake content verified by Newtral during the infodemic, in terms of number, format, dissemination channel, origin, type and possible risks to society.
3. To confirm whether the data resulting from the verification processes allow observing significant changes in the amount and typology of the content throughout the pandemic.

1.1. Fake content and verification

Although fake and unverified content and hoaxes have always existed, it can be asserted that they are a problem that has proliferated in the digital environment, among other issues because, as stated by Elías (2018, p.2), "in the second decade of the 21st century the truth is not so relevant anymore, since followers or entries matter more than the prestige of the source or the professional who signs". There is also the emergence of infoxication websites based on fake news and making this content go viral rapidly, increasing disinformation and loss of credibility. This has led to the question of what the reasons promoting this fake content and its subsequent consequences are, bearing in mind how

factors such as social networks influence on its dissemination, and what its proliferation entails for both the media and citizens (McNair, 2018).

Precisely, diffusion of fake content during the COVID-19 crisis has promoted research interest in this issue, addressing its impact on the media system, regarding detection capabilities and credibility (Casero-Ripollés, 2020); its dissemination on platforms such as Twitter, Instagram, YouTube, Reddit and Gab (Cinelli *et al.*, 2020), and its types (Brennen *et al.*, 2020). All this without losing sight of the analysis of what has motivated the spread of fake news during the pandemic, noting the role the lack of attention social media users' give to the context of disinformation has played, which has led to considering the need of improving the accuracy of information about the pandemic (Rand *et al.*, 2020).

Although it is true this infodemic occasioned by the health crisis has increased the interest in the analysis of fake pieces of content, they had already been subject of study in terms of their global nature (Pérez-Tornero and Varis, 2010), terminological scope (Wardle and Derakhshan, 2017) and conceptual delimitation, leading, for example, to considering the differentiating nuances between fake news and post-truth (Blanco-Alonso, 2020).

Regarding terminological aspects, it should be clarified that in spite of several authors (Wardle, 2017; Zuckerman, 2017; Boyd, 2017 and Jack, 2017) deeming the term “fake news” inadequate to describe a phenomenon as complex as misinformation and disinformation (Wardle and Derakhshan, 2017), the term “fake news” is present in a majority of the most cited articles on the Web of Science and Scopus (Blanco-Alfonso *et al.*, 2019).

Other aspects that have attracted attention in investigations focused on false pieces of information are their approach and their content, to make fake news more likely to be shared online by people than real news (Vosoughi, Roy and Aral, 2018). Precisely, among the aspects that influence their dissemination, research has detected: trust in the information shared, trust in the media, and the credibility of fake news (Montero-Liberona and Halpern, 2019).

It is exactly this ease of spreading and making fake content go viral what has led the different investigations that have been conducted in this regard to highlight the necessity for media literacy, thus “if new generations obtain information from social networks and other online resources, they must learn how to decode what they read” (Fernández-García, 2017, p.75).

The importance of fake news detection has also been underlined in the light of its potential for extremely negative impacts on individuals, although we must be aware it is a challenging task (Shu, *et al.*, 2017). It is important to bear in mind that fake news is a difficult problem to address, even with the detection tools available, since “the linguistic, terminological and semantic analysis is based on the study of frequencies and expressions that can be measured and countered” (Blázquez-Ochando, 2018, p.15), adding to it the real-time detection dilemma. Therefore, some authors reflect on the convenience of a regulatory framework, even proposing two perspectives in this regard: “1) the necessity of regulating the content regardless of the form in which it is presented; 2) the necessity of legislating on the form how content is presented as a transparency and defense mechanism of our rights as consumers” (Magallón-Rosa, 2018a, p. 1).

Other aspects worthy of notice have been the risk of content filtering and labeling (Pauner-Chulvi, 2018), its impact on electoral results (Allcott, 2017) and political affairs (Aparici, García and Rincón-Manzano, 2019), as well as its intentionality (Blanco-Herrero and Arcilla-Calderón, 2019).

All this without losing sight of the changes that have been incorporated into work routines (Rodríguez-Fernández, 2019) and the role of journalism as an antidote, taking into consideration that fake news pushes journalism towards a new context, in which “the media must understand that fighting disinformation and fake news requires greater training for their journalists and greater transparency regarding editorial policies” (Amorós-García, 2019, 40). Which also entails the necessity for new professional profiles (Ufarte-Ruiz, Peralta-García and Murcia-Verdú, 2018), where priorities are editors’ training in verification, transparent rectification policies, and being helpful to citizens (Jiménez-Cruz, 2019). In any case, we must keep in mind, as stated by Cebrián-Enrique (2012, p. 238), that “this social media context is an opportunity to stand for and assert journalism, by virtue of verification, among other values”.

Precisely, the increase in fake content and its rapid spread has led to the need of contemplating how to combat it; bearing in mind, as pointed out by Redondo (2018), that verifying consists in checking and contrasting. Although *ante hoc* fact-checking becomes complex in today’s context, since there is rapid access to information, as well as to disinformation, as a result of content going viral through social networks (Ortiz de Guinea and Martín-Sáez, 2019). Therefore, it is necessary to take into consideration, as stated by Silverman (2014), that technology can lead us astray as much as it can help us when there is lots of information circulating at a very hectic pace.

The importance of verification became evident in the interest that fake news identification platforms have acquired, giving rise to studies focused on their functioning, such as the case of the analysis conducted on *Maldita.es*, which demonstrated the relevance of mobile devices and social networks to make these types of projects work, as well as the implementation of a rigorous multi verification process for denials (Bernal-Triviño and Clares-Gavilán, 2019).

Likewise, fact-checking communication strategies have been worthy of notice, both with regard to refuting rumors and the type of circulating disinformation (Magallón-Rosa, 2018b). Precisely, disinformation has made verification one of the big information challenges on the online media that seek to become benchmarks for content consumption on social networks (Marcos-Recio, 2017), as well as it represents an opportunity for journalism in terms of credibility, leading to contemplating how to reverse this situation from within the sector itself, and what risks and opportunities are there for journalism in this context (Pérez-Rey and Calderón, 2019). It should be noted that in the last decade alone, before the COVID-19 pandemic, fact-checking platforms have emerged in more than 50 countries (Fernández-García, 2017).

In this regard, messaging services and social networks have also been part of researchers’ interest, such as Twitter (Magallón-Rosa, 2018b), WhatsApp (Palomo and Sedano, 2018) and Facebook (Guess, Nagler and Tucker, 2019), which have been worthy of attention in terms of processes of verification, dissemination, and audiences’ engagement in this task.

2. Methodology and research questions

This research focuses on Newtral since knowing of its verification methods bears value in itself, and the amount of cases analyzed allows comprehending the outreach and some characteristics of the infodemic.

The research questions to achieve the objectives proposed in this article are hereunder:

- Q.1. what are the verification methods Newtral has used during the COVID-19 crisis?
- Q.2. what types of sources have they used?

- Q.3. what verification tools have been used?
- Q.4. how have verifications evolved over time?
- Q.5. what has been the most common format of the fake content detected?
- Q.6. via which channels has the content been disseminated?
- Q.7. has the origin of the verified fake content been identified?
- Q.8. what types of content have been identified?
- Q.9. what possible risks has the dissemination of said content entailed?

To carry out the analysis, a triangulation methodology was used, combining qualitative and quantitative analyzes. A study of the verification procedures, an analysis of the verified content, and an unstructured personal interview with Newtral's head of contents were conducted.

The method followed these phases:

1. We selected 205 pieces of content verified and published by Newtral between January 25, 2020, the date when this entity verified the first piece of false news related to the coronavirus, until May 25, 2020, date when every Spanish region had officially come out of the so-called phase 0, hence the whole country had initiated the de-escalation process, with some provinces in phase 1 and others in phase 2. All the verifications subject of study were published in the *Zona de Verificación* [EN: Verification zone] (Fakes) on Newtral.es. The study includes all those pieces of content verified by Newtral during the COVID-19 crisis, somehow related to the pandemic, not only regarding health matters, but also other issues linked to the situation generated by the coronavirus and its consequences, such as lockdown or the political measures established in those days. Fake content disseminated during the study period not related to COVID-19 was not included in the sample. The 205 pieces of news analyzed were either detected by Newtral's own fact-checkers team or received through its WhatsApp "*verification on demand*" service, a channel the company offers to the public so that any person can contact Newtral via this private messaging platform to check if certain information is real or fake.
2. The verification procedures were studied first in the analysis of the sample. To this end, the specifications Newtral provides in each of its verifications were taken into account. According to the aforementioned IFCN Code of Principles, these must be transparent regarding the methodology used and the sources consulted. In this sense, we based on a typology that includes: consultation of various types of sources, other media, and fact-checkers, or the use of different verification software tools such as InVID, TinEye, Google Images, or Yandex, among others. In the event that sources were used for fact-checking, these are specified in the verification itself; therefore, we proceeded to their categorization differentiating between: expert sources, official sources, documentary sources (in most cases consultations of the Official State Gazette (*BOE*) of Royal Decrees that have been approved and published during the crisis), original sources of the image that was manipulated or taken out of context, impersonated sources (in the case of imposter content), or sources quoted or alluded to.
3. The following information was collected from each of the 205 verified pieces of content analyzed:
 - News title of the verification published on Newtral.es.
 - Link to the verification on the web.
 - Verification date. The publication date of the content analyzed is specified, or in most cases, it is the verification publication date instead, generally a few days after its detection.

- Format: text, video, voice message, image, alleged document, alleged piece of news, and email message.
 - Network on which the content was detected: stating if it was collected on Facebook, Twitter, Instagram, or WhatsApp. Although WhatsApp is a private messaging application, not a social network, it was included under the “networks” descriptor since it has played an equally important role in fake content dissemination.
 - Origin of fake content.
 - Type of fake content. Wardle’s typology, published on the First Draft website (2017), was used to classify fake news in an ascending order of harmfulness, based on their intent to deceive. Thus, it is categorized as: satire or parody, false connection (headlines, visuals or captions that do not support the content), misleading content, false context, imposter content (impersonated source), manipulated content, and fabricated content. Another category not contemplated by Wardle was also considered, since it has been identified in this COVID-19 disinformation crisis, consisting in alleged information mixing fake and genuine content.
 - Possible risks the disseminated content entails, identifying four: to misinform, to cause alarm, to cause rejection or harm (to a person, group, or country), and fraud or scam.
4. In order to contrast the data considered in the study and to know Newtral’s fact-checkers team perception, the information obtained through observation and analysis has been compared with an unstructured personal interview with Joaquín Ortega, head of contents of this organization, with the purpose of controlling researchers’ personal bias and correcting possible deficiencies, seeking to increase this study’s validity as well as it has allowed obtaining the necessary data to know the subject of study.

3. Results

3.1. Newtral’s verification methods

After monitoring the specifications detailed by Newtral in each of its verifications, according to the transparency policy in the methodology embodied in the IFCN Code of Principles, we observed that the verification method used by its fact-checkers team based on the utilization of different types of sources and various technological tools. In sources, consultations or collaboration with other IFCN fact-checker teams and with other media were included.

In most cases, more than one verification method is used, and when just one single method is implemented, it is common to utilize either more than one source and of different types, or different verification tools. We counted up to eight different sources and/or tools being used in the same verification.

In the sample under analysis, consultation of sources is much more frequent (90.5% of cases) than the use of technological tools (9.5%). Newtral’s head of contents, Joaquín Ortega, confirms this, “perhaps the verification method we have used the most during this pandemic is the least technological of all, which is consulting with experts in virology, epidemiology, medicine, etc., by phone or via email. Much of the content that was verified contained hoaxes about false remedies or scientific discoveries that were not real. In our methodology, we force ourselves to consult with several recognized experts from credible institutes or entities to be absolutely certain that what we are verifying is either fake or not. When it comes to verifying health matters, we have to have everything very connected, be clear and straightforward when explaining it”.

This study demonstrates that the most consulted sources are the official ones, from which data is frequently requested. Among the most contacted sources there are the Ministry of Health, WHO, the National Police, the Civil Guard and Regional Police Departments. Secondly, impersonated sources appear, followed by the sources quoted or alluded to in the fake content. Finally, other IFCN fact-checker members such as Alt News (India), Chequeado (Argentina), Ecuador Chequea, Full Fact (The United Kingdom), Snopes, AfricaCheck, AFP Factual, or Les Décodeurs (Le Monde) are consulted.

The next most consulted sources are the original ones, which can either be the original video being reported, the original statements of a public figure, or the website from which the information originates and turns out to be a satirical website, for example. Finally, and in this order, documentary sources (legal texts in almost all cases), expert sources (generally linked to different universities) and the media, mostly foreign (mainly from China) or local ones are consulted.

Regarding the tools used in the verifications analyzed, related to the coronavirus crisis case, they did not need to be very sophisticated or technological. All of them are free applications available on the Internet. “One of the tools we use most is the search for content already published and stored on platforms such as Google, Yandex, Bing or TinEye. Why? Because people who dedicate to creating fake content usually utilize elements available on the Web. Photos of people, situations, buildings, videos on social networks, etc., that already exist, and that are manipulated with audiovisual editing tools or events that are decontextualized by placing them in a different situation or implicating people who were never there”, said Ortega.

The most used tools are search engines such as Google Images or Yandex, which allow tracking images on the Internet by performing a reverse search to find similar images to the one being verified, hence checking if it is prior to what is being claimed in the false information, since using old images as if they were recent is a very common practice.

Secondly, they turn to InVid, a specific tool for video verification. This tool, of the French news agency AFP, provides basic data on the publication/posting, production, management rights, and the broadcaster of a video. It also allows executing reverse searches based on frames of a video.

Thirdly, geolocation tools are used -such as Google Maps, Google Earth or Baidu Maps- to check if an event did really occur in a certain place. Next, the most used method is what has been labeled as content analysis, which includes some simple verification practices, such as checking the account from which a tweet originated or analyzing a video to make sure it is not Spain, for example.

The next tool used is TinEye, a website that allows knowing when an image was uploaded to the Internet for the very first time, a key datum when it comes to verifying fake content that sets a photograph on a false date.

Table 1: *Newtral's verification methods*

SOURCES	286 (90.5%)	TOOLS	30 (9.5%)
Official Source	102	Search engines	13
Impersonated source	42	InVid	6
Quoted source	40	Geolocation	5
Other <i>fact-checkers</i>	27	Content analysis	3
Original Source	21	TinEye	2
Documentary source	20	Other	1
Expert source	19		

The media	13		
-----------	----	--	--

Source: authors' own creation based on data retrieved from newtral.es

3.2. Evolution of verifications performed by Newtral over time

The first fake news detected by Newtral's team related to the coronavirus was spotted on Facebook, on January 25, 2020, and verified on January 28. It was a video, shared over 2,000 times, claiming to show images of the Wuhan seafood market (focal point of the coronavirus outbreak in China), but it was actually recorded in Indonesia. It was verified using the InVid tool, proving that these same images had been uploaded to YouTube in June 2019.

From this first fake news related to COVID-19, until the beginning of the de-escalation process across Spain on May 25, when all the regions that were still in phase 0 changed to phase 1 and some of them to phase 2, there were 205 pieces of content verified by Newtral, which means an average of 1.7 verified pieces of content per day.

This chronological analysis permits observing the evolution of the infodemic. Most of the dates registered correspond to those when the verifications were published, which often occur a day or two after their detection on the Internet. Only in one exceptional case has a hoax been verified months after its detection, such as the one published under the title "*El 'bicho' extraído de un labio en un video viral no vive en «las latas de refrescos» ni es efecto del coronavirus*" [The "bug" pulled out of a lip in a viral video does not live in soda cans nor is it an effect of the coronavirus], which was verified in February 2020, but was first detected in October 2019.

The least amount of fake content was detected and verified between January 25 and March 9 (five days before the state of alarm was decreed). During this period, the verified pieces of content ranged from 0 to 4 per day, with a mean of 0.7. March 10 meant a turning point since the number of verified pieces of content soared over 10. As of the next day (March 11), the verified pieces of content ranged from 0 to 6 per day, being the average of that second period significantly higher than the first one, with 2.3 verified pieces of content per day.

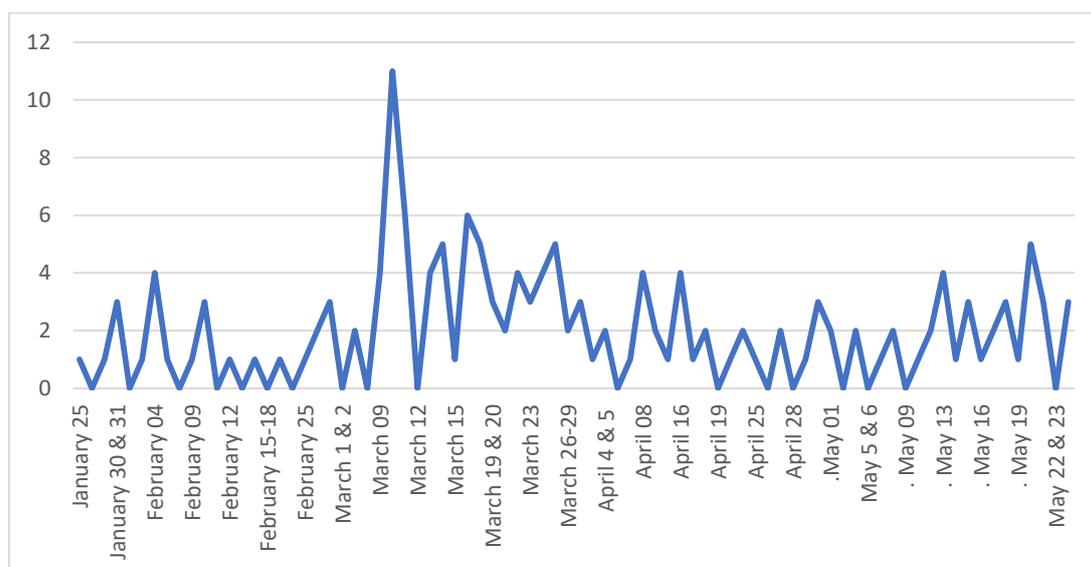


Chart 1: Number of fake pieces of content verified by Newtral per day

Source: authors' own creation based on data retrieved from newtral.es

3.3. Fake content formats

As for the formats of the verified pieces of content, the most disseminated are text messages shared on social networks (Facebook posts, Instagram stories, and tweets) or through WhatsApp chain letters. No sources are quoted and they can include all type of content: false remedies for coronavirus, false information on political measures taken, etc.

The second most used format is video, followed by photograph or screenshots. Thirdly, we have fake documents, such as an alleged letter or statement from an official entity. In this sense, for example, false announcements from the City Council of Barcelona, the University of Seville, the Xunta de Galicia, the Ministry of Education, and UNICEF, among others, have been found. Fourthly, alleged information coming from recognized news agencies such as *La Vanguardia*, *la Cadena SER* or the *Agencia EFE*, for example. Fake tweets from the media have been included in this category of alleged news since they aim to cause the same effect.

The fifth most used format is voice messages. Newtral's head of contents highlights the increase in this format: "voice messages were already being used to disseminate false warnings or fabricated stories, but they have multiplied during the pandemic." Finally, and a lot less frequently, we find email messages.

Table 2: *Formats of the pieces of contents verified by Newtral*

FORMAT	N° pieces of content	%
Text	85	41,4
Video	42	20,4
Image	28	13,6
Alleged document	18	8,7
Alleged piece of news	15	7,4
Audio	13	6,4
Email	4	2,0

Source: authors' own creation based on data retrieved from newtral.es

Note: One piece of content was identified in two different formats. Both were included.

3.4. Dissemination channels

As for how the fake pieces of content were disseminated, the majority (66%) were detected on more than one social network, including WhatsApp messaging application. Regarding the pieces of content disseminated through a single channel, most of them went viral via WhatsApp, followed by Twitter and Facebook, in this order. Four pieces of content that were disseminated via email or SMS were also detected in the sample; all of them were scams. Finally, two fake or manipulated videos were also broadcasted on TV.

Table 3: *Channel through which fake content was disseminated*

CHANNEL	N° pieces of content
Several social networks	136
WhatsApp	36
Twitter	18
Facebook	8

Email and SMS	4
Other / TV	2
Digital media	1

Source: authors' own creation based on data retrieved from newtral.es

3.5. Origin of verified content

It was impossible to detect the origin in 91.7% of the verified pieces of content, that is, to spot where the content originated or who created it for the first time. In those cases the origin is classified as “unknown”. The account was identified only when the content came from a specific user of a social network such as Twitter or Facebook.

It has indeed been identified when the origin is an online medium, which other digital portals and several messages on social networks have subsequently echoed, making the content viral. Likewise, in one case, it was possible to conclude that the origin was in a different country from Spain (France), but with no further specifications. Only on one occasion it was possible to identify the person who disseminated the content, who mixed accurate and wrong information without any will to misinform. Finally, in some cases, it was not possible to identify the origin, but it was possible to detect some of the individuals who contributed to making the fake news go viral, such as the case in which the deputy of *Partido Popular*, Rafael Hernando, echoed an alleged image of a group of Muslims walking on the street during the confinement but it was actually from 2018, hence he later rectified his mistake on the same social network.

Table 4: *Origin of verified content*

ORIGIN	N° pieces of content
Unknown	188
Twitter account	10
Facebook Account	1
TikTok Account	1
Online media	2
Satirical website	1
Identified Person	1
Email account	1

Source: authors' own creation based on data retrieved from newtral.es

3.6. Content typology

It has been noted that the most frequent pieces of content are precisely the most harmful ones, according to Wardle's classification (2017), the so-called fabricated content. An example is the one that was disseminated on Facebook and Twitter and was later verified on January 29 with the title “*Es falso que haya informes sobre más de 10.000 muertos en Wuhan a causa del coronavirus*” [EN: It is false there are reports about more than 10,000 deaths in Wuhan due to the coronavirus]. Secondly, there is imposter content, which impersonates a source, as in the case of an alleged document from the Ministry of Health released on February 9 that recommended “not to supply products from China”, which was actually a fabrication. Thirdly, there is the content with false contextual information (false context), such as a video of an alligator apparently recorded in Getxo (Biscay) linked to changes in nature caused by the coronavirus. The video was actually filmed in 2015, in Alabama State, (The United States).

On May 24, a curious case was identified, since true information is claimed to have false contextual information. In this case, the verification consists precisely in demonstrating that the context of the video was actually true (an ambulance trying to go through a Vox demonstration in Santander on May 23, 2020).

Finally, the analysis conducted detected very little satire and false connection content. In addition, there were five cases observed that do not fall within any of the other seven classifications, since they mix true and false information, and it is not clear whether they bear malice.

Table 5: *Verified content typology scaled from least to most harmful*

TYPE OF CONTENT	Nº of pieces of content	%
Mixing true and false information	5	2,4
Satire or parody	4	2,0
False connection	3	1,5
Misleading content	17	8,3
False context	47	22,9
Imposter content	48	23,4
Manipulated content	12	5,9
Fabricated content	69	33,6

Source: Authors' own creation.

Within the imposter content group, an analysis of the impersonated sources was carried out, noting the ones of the media as the most frequent. There were impersonations of *El País*, *El Mundo*, *La Vanguardia*, *La Razón*, RTVE, Agencia EFE, Cadena SER, *La Voz de Galicia*, *Público*, *eldiario.es* and its director, Ignacio Escolar.

Secondly, the most commonly impersonated sources are those grouped under the descriptor "Companies and Organizations" -including companies such as *Netflix*, *Caixabank*, *Leroy Merlin* and *Carrefour*, as well as organizations such as *Caritas*, *CORREOS*, Social Security, and the Spanish Tax Agency-. There are risks of fraud or scam in these cases.

Thirdly, we found sources of the Presidency of the Government and of three ministries (Health, Interior, and Education), followed by various regional sources of Madrid, Galicia, Catalonia, and the Valencian Community. Four universities (University of Seville, University of Jaén, Stanford University, and the University of Barcelona), official entities such as WHO, UNICEF and UN, city councils (Barcelona, Malaga and Tarragona), foreign health authorities, Police and Civil Guard, and Spanish health sources, specifically, the Head of Cardiology at the Hospital Gregorio Marañón were also impersonated.

Table 6: *Impersonated sources in the imposter content group*

IMPERSONATED SOURCE	Nº of pieces of content
Media	13
Companies and entities	8
Government	7
Autonomous Communities	6
Universities	4
Official International Organizations	3
City Councils	3
Foreign Health Authorities	2

Police and Civil Guard	2
Spanish Health Authorities	1

Source: authors' own creation.

Several cases of the same fake content being spread through different mediums have been detected, such as images of body bags lying on the ground that were recorded in Ecuador and were disseminated in several countries as if they had been filmed in local hospitals in the United States and Spain (in Madrid and Barcelona).

In other cases what changes is the content format and the regional community to which it refers, such as the alleged piece of news disseminated in Catalonia and Madrid (as a voice message and an alleged document, respectively) about a group of thieves disguised as doctors who had entered homes on the pretext of conducting coronavirus tests but with intent to rob.

On another note, we observed that some fake pieces of content have been disseminated in different moments throughout the pandemic crisis. For example, some alleged rules that had to be obeyed during the state of alarm were disseminated in mid-March and went viral again in mid-April.

3.7. Risks that verified fake content entails

Regarding the possible risks verified content entails, mere disinformation appeared 113 times (55.1% of the sample). In the case of COVID-19, many hoaxes of this type also pose health risks since they recommend supposed treatments for the virus that can be harmful. Such is the case of a message shared via WhatsApp and verified on March 10, which claimed that by holding your breath you would know if you were infected by coronavirus. These types of messages were detected 14 times.

Generating social unrest was the second most frequent risk, being observed in 45 cases (21.9% of the total). An example of this was the message that circulated on social networks and that was verified on March 16, which claimed that during the state of alarm decreed due to the coronavirus, insurance companies were not covering traffic accidents. The third, which was detected 37 times (18.0%), aimed to cause rejection or harm towards a person, institution or group, such as an alleged piece of news warning that a Romany resident group from Haro had refused to follow the health protocols.

Finally, in fourth place (10 cases detected, 4.8%) there were frauds or scams, such as the message disseminated on the Internet at the beginning of April indicating that *Caritas* offered daily coupons of up to 1,000 euros to anyone who requested it by filling out an application form with personal data and bank account numbers. No other similar content had been registered before March 23 and most of them are cases of phishing, a fraudulent attempt that seeks to obtain someone's sensitive information.

In those cases where there were more than one risk associated, the one deemed predominant was counted.

Within the section "rejection or harm", we analyzed who might be harmed by false content. In this sense, it is observed that content inclined to harm a politician were the most numerous. We identified potentially harmful pieces of content aimed at Spanish politicians such as Manuela Carmena, Irene Montero, Pablo Iglesias, Teresa Rodríguez, Pablo Casado, Santiago Abascal, and José María Aznar; foreigners such as Christine Lagarde and Vladimir Putin, and at political parties (*Podemos*) and the political class in general. All of them have been detected as of March 10.

Secondly, there are those pieces of content that entail risks of harming the whole Government or some of its members, mainly the President. And thirdly, those that could harm the image of a certain group (content against Romany, Muslims, transgender people, and protesters against the Government have been detected) or a country (all those found in the sample refer to China). Finally, content hostile to the king and other referring to a journalist were detected.

Table 7: *Pieces of content that entail risks of generating rejection or harm towards certain groups*

GROUP	Nº of pieces of content
Politicians	14
Government	12
Groups	5
Country	4
The King	1
Journalist	1

Source: authors' own creation.

4. Discussion

Based on this study's results, after analyzing the 205 fake pieces of content verified by Newtral between January 25 and May 25, 2020, and taking into consideration the initial research questions, it is noted how Newtral has used a wide-ranging variety of verification methods. Although consultation sources are used the most to contrast the content compared to the scarce use of technological tools. This result leads one to think of the importance of traditional journalistic methods in verification processes, beyond the additional support new tools such as search engines or more sophisticated ones may provide to this end.

It is remarkable that when observing the sources consulted in the verifications, impersonated sources were used in second place after official sources. This is even more relevant if we take into account that imposter content (which impersonates sources) are the second most detected ones in the sample analyzed. Among the most impersonated sources identified in the results section, various media, companies, and official and government organizations stand out. This situation raises the necessity for these actors to have expert teams to detect and verify fake content disseminated on the Internet.

Regarding the evolution of the verified content over time, an increase is observed as of March 9, although this evolution is not constant. It is significant how the greatest increase in verifications throughout the analyzed period (four months) occurred between March 9 and 12. Bearing in mind that the dates registered in Chart 1 correspond mainly to those when the verifications were published and they tend to be one or two days after their detection on social networks. Hence it is understood that the spike in fake content dissemination corresponds to the days after March 8, coinciding with the controversies generated around the demonstrations held on that date and with WHO declaring the pandemic. It is striking that the largest amount of fake content detected corresponded to these dates before the state of alarm declaration in Spain.

As for the channels used to disseminate fake content, it is interesting to note that in most cases it happened through more than one social network, although the predominant channel was WhatsApp private messaging application, a communication system especially used in Spain.

By analyzing the data, certain limitations were noticed when it came to detecting the origin of the content analyzed. This is a consequence, on some occasions, of the information coming from anonymous accounts, which might be a field of study worthy of notice from different perspectives; informative, technological and legal.

Regarding the risks the disseminated fake content entails, in this case, since it is a large-scale health crisis, mere disinformation entails, in many cases, risks to the health of those who receive the content; hence, the fact of it going viral is particularly dangerous.

5. Conclusions

Just as stated by Cebrián-Enrique (2012), the diversity and complexity of the journalistic reality prevents the establishment of a universal verification method. This becomes evident when analyzing Newtral during this period, in which the use of different verification systems and tools was noted, although there is predominance of some of them over others. Therefore, it is proven that consultation of various sources is the most used method during the COVID-19 crisis, especially official sources and those that have been impersonated. It is striking how the use of sources prevails over verification tools, which are characterized by being very basic; with search engines such as Google Images or Yandex being the most used ones.

Regarding the fake pieces of content detected, it is noted how the number of verifications reached its highest on March 10, four days before the Government decreed the state of alarm, and it remained higher throughout the second period analyzed, during which Spain was in the midst of the lockdown and the health crisis. At the same time, in accordance with previous studies, such as the one of Cinelli *et al.* (2020) and McNair (2018), the role of social networks in the dissemination of the content verified by Newtral has been demonstrated, noting the predominance of text messages disseminated on social networks or via private messaging platforms such as WhatsApp. Similarly, we have observed that most of the pieces of content were disseminated through more than one social network, although their origin could not be identified in most cases.

As for the verified content typology, the most widely disseminated was fabricated content, deemed by authors such as Wardle (2017) to be the most harmful one, followed by imposter content that impersonates a source (mainly media, companies, and organizations -in the case of scams-, and central and regional government sources) and those providing false context. Regarding the latter type and in the case of videos, these frequently went viral globally. Precisely, this type of content, just as authors like Ortiz de Guinea and Martín-Sáez (2019) or Silverman (2014) already warned in previous studies, together with the rapid dissemination on social networks, have once again promoted disinformation to be the most frequent risk, followed, in this case, by the possibility of generating social unrest and causing rejection or harm towards a person, group or institution.

Likewise, we have observed that the content typology differs throughout the entire period analyzed. In this regard, it was found that some pieces of content, such as those with intent to harm politicians or involving fraud or scam, were not disseminated during the first weeks and only appeared as of mid-March, when the magnitude of the crisis was already noticeable.

The disinformation risks this type of crisis like the COVID-19 entails have cause not only for fact-checker teams to conduct verification tasks, but also for various sources to issue denials which are typically disseminated on the same social networks through which false content is spread. In this sense, it seems reasonable to conclude that communication departments of official organizations and large private companies will gradually incorporate professional experts on verification to their teams

to face this type of infodemics, hence the emergence of a new field of work worthy of interest for future communication investigations.

6. References

- Aldama, Z. (2020, enero 5). Un brote de neumonía atípica en China revive el fantasma de la epidemia mortal". *Heraldo de Aragón*. <https://www.heraldo.es/noticias/internacional/2020/01/05/un-brote-de-neumonia-atipica-en-china-revive-el-fantasma-de-la-epidemia-mortal-1351994.html>
- Allcott, H. y Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31 (2), 2011-2036. <https://doi.org/10.1257/jep.31.2.211>
- Amorós-García, M. (2019). Los medios de comunicación contra las noticias falsas: todo es falso, ¿salvo el periodismo? *Cuadernos de Periodistas*, (38), 21-40.
- Aparici, R., García-Marín, D. y Rincón-Manzano, L. (2019). Noticias falsas, bulos y trending topics. Anatomía y estrategias de la desinformación en el conflicto catalán. *El profesional de la Información*, 28 (3). <https://doi.org/10.3145/epi.2019.may.13>
- Bernal-Triviño, A. y Clares-Gavilán, J. (2019). Uso del móvil y las redes sociales como canales de verificación de Fake News. El caso de Maldita.es. *El profesional de la información*, 28 (3), 1-8. <https://doi.org/10.3145/epi.2019.may.12>
- Blanco-Alfonso, I., García Galera, C. y Tejedor Calvo, S. (2019). El impacto de las fake news en la investigación en Ciencias Sociales. Revisión bibliográfica sistematizada. *Historia y comunicación social*, 24 (2), 449-469. <https://doi.org/10.5209/hics.66290>
- Blanco-Alfonso, I. (2020). Posverdad, percepción de la realidad y opinión pública. Una aproximación desde la fenomenología. *Revista de Estudios Políticos*, (187), 167-186. <https://doi.org/10.18042/cep/rep.187.06>
- Blanco-Herrero, D. y Arcilla-Calderón, C. (2019). Deontología y noticias falsas: estudio de las percepciones de periodistas españoles". *El profesional de la información*, 28 (3), 1-13. <https://doi.org/10.3145/epi.2019.may.08>
- Blázquez-Ochando, M. (2018). El problema de las noticias falsas: detección y contramedidas. En: XV Seminario Hispano-Mexicano de Investigación en Biblioteconomía y Documentación, Ciudad de México.
- Boyd, D. (2017, marzo 27). Google and Facebook can't just make Fake News Disappear. En *Wired*. <https://www.wired.com/2017/03/google-and-facebook-cant-just-make-fake-news-disappear/>
- Brennen, S., Simon, F., Howard, Ph. y Nielsen, R. (2020, abril 7). Types, sources and claims of COVID-19 misinformation. <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>
- Cinelli, M., Quattrociocchi, W., Galeazzi, C., Brugnoli, E., Schimdt, A; Zola, P., Zollo, F. y Scala, A. (2020). The COVID-19 Social Media Infodemic. arXiv:2003.05004v1 [cs.SI]. Cornell University, 1-18. <https://arxiv.org/pdf/2003.05004v1.pdf>

- Casero-Ripollés, A. (2020). Impacto del Covid-19 en el sistema de medios. Consecuencias comunicativas y democráticas del consumo de noticias durante el brote. *El Profesional de la información*, 29 (2), 1-12. <https://doi.org/10.3145/epi.2020.mar.23>
- Cebrián-Enrique, B. (2012). Al rescate de la verificación periodística. *Zer*, 17 (33), 227-241. <https://ojs.ehu.es/index.php/Zer/article/view/10633>
- Coromina, Ó., Prado, E. y Padilla, A. (2018). The grammatization of emotions on Facebook in the elections to the Parliament of Catalonia 2017. *El profesional de la información*, 27 (5), 1004-1011. <https://doi.org/10.3145/epi.2018.sep.05>
- De la Cal, L. (2020, enero 8). La misteriosa neumonía que está poniendo nerviosa a China. *El Mundo*. <https://www.elmundo.es/ciencia-salud/salud/2020/01/08/5e146e4121efa044108b4631.html>
- Elías, C. (2018). Fakenews, poder y periodismo en la era de la posverdad y 'hechos alternativos'. *Ámbitos. Revista Internacional de comunicación* (40), 1-6. <https://revistascientificas.us.es/index.php/Ambitos/article/view/8913>
- Fernández-García, N. (2017). Fake News: una oportunidad para la alfabetización mediática. *Nueva Sociedad*, (269), 66-77.
- Guess, A., Nagler, J. y Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5 (1). <https://doi.org/10.1126/sciadv.aau4586>
- Jack, C. (2017). Lexicon of Lies. *Data & Society*. https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf
- Jiménez-Cruz, C. (2019). En la era de la desinformación, periodismo para que no te la cuelen. *Cuadernos de Periodistas*, (38), 13-20. <http://www.cuadernosdeperiodistas.com/en-la-era-de-la-desinformacion-periodismo-para-que-no-te-la-cuelen/>
- Jurado-Salván, E. y Jurado-Izquierdo, M. (2014). Los errores de comunicación en la crisis del Ébola. *Cuadernos de Periodistas*, (29), 90-99. <http://www.cuadernosdeperiodistas.com/los-errores-de-comunicacion-en-la-crisis-del-ebola/>
- Magallón-Rosa, R. (2018a). Leyes fake news. El problema para la libertad de información de no legislar. *Telos*, <https://telos.fundaciontelefonica.com/las-leyes-las-fake-news-problema-la-libertad-informacion-no-legislar/>
- Magallón-Rosa, R. (2018b). Nuevos formatos de verificación El caso de Maldito Bulo en Twitter. *Sphera Pública*, 1 (18), 41-65. <http://sphera.ucam.edu/index.php/sphera-01/article/view/341>
- Marcos-Recio, J.C. (2017). Verificar para mejorar la información en los medios de comunicación con fuentes documentales. *Hipertext.net: Revista académica sobre documentación digital y comunicación interactiva*, (15), 36-45. <https://doi.org/10.2436/20.8050.01.44>
- Martín-Barato, A.; López-Doblas, M.; Luque-Martín, N. y March-Cerdá, J. C. (2020, abril 15). Fake news y bulos contra la seguridad y la salud durante la crisis del coronavirus. Escuela Andaluza de

Salud Pública. <https://www.easp.es/web/coronavirusysaludpublica/fake-news-y-bulos-contra-la-seguridad-y-la-salud-durante-la-crisis-del-coronavirus/>

- McNair, B. (2018). *Fake News. Falsehood, Fabrication and Fantasy in Journalism*. Routledge.
- Montero-Liberona, C. y Halpern, D. (2019). Factores que influyen en compartir noticias falsas de salud en línea. *El profesional de la información*, 28 (3). <https://doi.org/10.3145/epi.2019.may.17>
- Ortiz de Guinea, Y. y Martín-Sáez, J. L. (2019). De los bulos a las fake news. Periodismo, contenidos generados por el usuario y redes sociales. *Creatividad y Sociedad. Revista de la Asociación para la Creatividad*, (30), 104-124.
- Palomo-Torres, B. y Sedano-Amundarain, J. (2018). WhatsApp como herramienta de verificación de fake news. El caso de B de Bulo. *Revista Latina de Comunicación Social*, 73 (13), 1384-1397. <https://doi.org/10.4185/RLCS-2018-1312>
- Pauner-Chulvi, C. (2018). Noticias falsas y libertad de expresión e información. El control de los contenidos informativos en la red. *Teoría y Realidad Constitucional*, (41), 297-318. <https://doi.org/10.5944/trc.41.2018.22123>
- Pérez-Tornero, J.M y Varis, T. (2010). *Media Literacy and New Humanism*. UNESCO Institute for Information Technologies in Education.
- Pérez-Rey, J. y Calderón, M.E. (2019). Combatir el ruido de los bulos desde la esencia del periodismo. *Interactiva: Revista de la comunicación y el marketing digital*, (187), 50-54.
- Rand, D., Pennucook, G., McPhetres, J. y Zhang, Y. (2020). *Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy nudge intervention*. <http://ide.mit.edu/publications/fighting-covid-19-misinformation-social-media-experimental-evidence-scalable-accuracy>
- Redondo, M. (2018). *Verificación digital para periodistas. Manual de bulos y desinformación internacional*, Barcelona: Editorial UOC.
- Rodríguez-Fernández, L. (2019). Desinformación: retos profesionales para el sector de la comunicación. *El profesional de la información*, 28 (3), 1-11. <https://doi.org/10.3145/epi.2019.may.06>
- Santirso, J. (2020, enero 11). Un virus similar al SARS, responsable de la misteriosa neumonía china. *El País*. https://elpais.com/sociedad/2020/01/09/actualidad/1578556344_366873.html
- Silverman, C. (2014). *Verificación Handbook*. Reino Unido: European Journalism Centre. <https://verificationhandbook.com/downloads/verification.handbook.pdf>
- Shu, K., Sliva, A., Wang, S., Jiliang, T. y Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *AMC SIGKDD Explorations Newsletter*, 19 (1), 22-36. <https://doi.org/10.1145/3137597.3137600>

- Suárez, E. (2020, marzo 31). How fact-checkers are fighting coronavirus misinformation worldwide. Reuters Institute. <https://reutersinstitute.politics.ox.ac.uk/risj-review/how-fact-checkers-are-fighting-coronavirus-misinformation-worldwide>
- Ufarte-Ruiz, M.J., Peralta-García, L. y Murcia-Verdú, J. (2018). Fact checking: un nuevo desafío del periodismo. *El profesional de la información*, 27(4), 733-741. <https://doi.org/10.3145/epi.2018.jul.02>
- Valera, S. (2020, abril 22). Cristina Tardáguila: Estamos ante la peor ola de desinformación de la historia". Asociación de la Prensa de Madrid. <https://www.apmadrid.es/cristina-tardaguila-estamos-ante-la-peor-ola-de-desinformacion-de-la-historia/>
- Vosoughi, S., Roy, D., Aral, S. (2018). The Spread of true and false news online. *Science*, 359 (6380),146-1151. <https://doi.org/10.1126/science.aap9559>
- Wardle, C. (2017, marzo 14). Noticias falsas. *First Draft*. <https://es.firstdraftnews.org/2017/03/14/noticias-falsas-es-complicado>
- Wardle, C. y Derakhshan, H. (2017). *Information disorder. Toward an interdisciplinary framework for research and policymaking*. Estrasburgo: Council of Europe report DGI. <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c>
- Zuckerman, E. (2017, enero 30). Stop saying Fake News, It's not helping. *My Heart is in Accra*. <http://www.ethanzuckerman.com/blog/2017/01/30/stop-saying-fake-news-its-not-helping/>

AUTHORS:

Guadalupe Aguado-Guadalupe

She is a Full Professor at the Department of Communication of the University Carlos III of Madrid. She has a Ph.D. in Information Sciences from the Complutense University of Madrid. Member of the Research Group Journalism and Social Analysis: Evolution, Effect and Trends (PASEET). Her line of research is focused on trends and business models of communication in the digital environment.

H-Index: 15

Orcid ID: <https://orcid.org/0000-0001-7314-2403>

Google Scholar: https://scholar.google.es/citations?user=_BxwRkAAAAJ&hl=es

Itziar Bernaola-Serrano

She is a professor at the Department of Communication of the University Carlos III of Madrid. She is a Doctoral Candidate at the Media Research Program of the University Carlos III of Madrid. Her line of research is focused on verification processes, agenda-setting and use of journalistic sources.

H-Index: 1

Orcid ID: <https://orcid.org/0000-0002-1607-2661>

Google Scholar: <https://scholar.google.es/citations?user=K6PQsfMAAAAJ&hl=es&oi=ao>