

Misinformation about fake news. A systematic, and critical, review of empirical studies on the phenomenon and its status as a “threat”

Abstract: After the 2016 US presidential elections, the term fake news became synonymous with disinformation and a catch-all term for the problems that social networks were bringing to communication. Four years later, there are dozens of empirical studies that have attempted to describe and analyse an issue that, despite still being in the process of definition, has been defined as one of the key Covid-19 cyberthreats by Interpol, is considered a threat to democracy by many states and supra-national institutions and, as a consequence, is subject to regulation or even criminalization. These legislative and criminal policy interventions form part of the first stage in the construction of a moral panic that may lead to the restriction of freedom of expression and information. By analysing empirical research that attempts to measure the extent of the issue and its impact, the present article aims to provide critical reflection on the process of constructing fake news as a threat. Via a systematic review of the literature we observe, firstly, that the concept of fake news used in empirical research is limited and should be refocused as it has not been constructed according to scientific criteria and can fail to include relevant elements and actors, such as governments and traditional media. Secondly, the paper analyses what is scientifically known about the extent, consumption and impact of fake news and argues that it is problematic to establish causal relationships between the issue and the effects it has been said to produce. This conclusion forces us to conduct further research and to reconsider its position as a threat as well as the resulting regulation and criminalization.

Keywords: fake news, misinformation, social networks, threat, criminalization

Introduction

It is not easy to find references to the concept of fake news in academic literature before 2016 (Tandoc, Linm, & Ling, 2017). Those that exist mainly relate it to “news parody” (Marchi, 2012; Brewer, Goldthwaite, & Morreale, 2013), humorous content that imitates the format of the mainstream news - parody news - (Berkowitz & Schwartz, 2016), or the use of real news for comedy - satirical news (Baym, 2005). The term subsequently began to acquire new nuances and be linked to concepts such as political propaganda (Khaldarova & Pantti, 2016) but it was not until the 2016 US presidential election campaign that the concept of fake news gained the popularity and meaning it has today (McGonagle, 2017). First, a series of newspaper articles drew attention to the widespread dissemination of false information on the social network Facebook regarding issues connected to the political contest (Silverman & Alexander, 2016; Silverman, 2016). The subsequent victory of the Republican candidate, contrary to what most surveys had predicted¹, led some authors to turn fake news into a modern version of disinformation that is related to cyberspace and social networks (Tandoc, Linm, & Ling, 2017; Lazer et al., 2018), and to link it to political events such as the aforementioned Donald Trump victory (Bennett & Livingston, 2018) or Brexit (Bastos & Mercea, 2017). This prompted statements of concern from states and institutions about the risks of fake news (European Comission, 2018; Parliament of Singapore, 2018; Gobierno de España, 2019), as well as the proliferation of journalistic initiatives that aim to verify information - fact-checking. In turn, social concern, as is often the case, has led to the emergence of various proposals for regulation and even criminalisation (Lee & Lee, 2019; Tandoc, 2019).

It seems that a certain degree of moral panic (Young, 1971; Cohen, 1972) is being generated with regard to fake news and disinformation, or at a minimum the process of identifying a new phenomenon as a threat to social values and interests is underway,

which constitutes the initial communicative step toward a response based on social control and criminal regulation. Faced with this type of situation, it is always necessary to conduct empirical studies that help substantiate the threat or call it into question (Garland, 2008). In fact, from 2016 there was indeed a significant increase in academic research on the issue and its potential negative impact (Farkas & Schou, 2018). However, the proliferation of studies in disciplines ranging from computer science to information science that quantify and measure the various impacts of fake news has not always been combined with reflection on and conceptualization of the phenomenon. While it is clear that the potential to manipulate society and to alter democratic processes by means of false, distorted or decontextualized information needs to be studied scientifically, this means that it is necessary to adequately review how the concept of fake news is being empirically constructed.

And this is the main objective of the present paper: to identify which conceptualization of fake news is being constructed as an empirically “proved” object, to use Bourdieu’s terminology (2002). This constitutes the first analytical step in order to reflect upon the consequences of the object being analysed and whether it is the object itself that should be of particular concern or whether we should broaden our outlook. To this end, an initial systematic review of the existing empirical literature will be conducted, followed by a critical reflection on the conceptual bases of the works analysed in the review. Because if we do not proceed in this manner, if we do not review what we are measuring, we run the risk of using science to misinform about an issue that generates social concern and that is already being regulated and even criminalized.

Background

The "threat" (and moral panic) of fake news

The dissemination of false information with political or economic objectives is not a new phenomenon and neither can it be said that the dissemination of false information arose with the Internet (Levi et al, 2019). State use of propaganda, as a way of broadcasting the 'correct' worldview, is inherent to their existence. At the end of the 19th century, the magnate William Randolph Hearst, considered the creator of the yellow press (Spencer, 2007), used his media empire to spread distorted news that benefited his interests, even precipitating the Spanish-American War of 1898 (Lowry, 2013). However, it must be acknowledged that today's communicative ecosystem is not identical to that of the 19th century, or even the 20th century. The generalization of the Internet and the increasing prominence of social networks have substantially modified the field of communication, making access cheaper - the investment and knowledge needed to create a website and monetize visits are minimal - and offering new channels for the dissemination of information (Lazer et al., 2018).

In view of this new communicative context, many authors have suggested that, due to the production and dissemination of fake news, there is increased risk of the population being manipulated and becoming incapable of distinguishing true information from false (Lazer et al., 2018). Moreover, this is aided not only by the content, but also by the way in which it reaches the recipient. For example, the diffusion of fake news can be amplified by the use of automated fake-account networks on social networks. These are capable of increasing the diffusion of certain content, which can make it appear more reliable or credible and, therefore, contribute to disinformation (Bastos & Mercea, 2017; Shao, Ciampaglia, Varol, Flammini, & Menczer, 2018). Furthermore, the consequences of this are not limited to the electoral sphere but can also affect other areas of vital importance, such as public health (Wanga, McKeeb, Torbicaa, & Stucklerc, 2019) or the economy (Zhang & Ghorbani, 2020; Fedeli, 2020). In fact, the Covid-19 crisis seems to have

increased the perception that fake news and disinformation are a threat. The WHO warned (2020) that the COVID-19 pandemic was accompanied by an infodemic of disinformation, which seemed to be subsequently corroborated by both Europol (2020) and Interpol (2020). Via its Global Cybercrime Survey, the latter of these organisations confirmed the circulation of false information related to COVID-19 in many countries and, in addition, expressed concern that incorrect information could spread panic in the community and social disorder that have already been exacerbated by the pandemic. Their report also linked fake news to some forms of cyber fraud and to illegal trade in fraudulent medical products.

The status of “threat” that has been assigned to disinformation and fake news in a short period of time has led to many forms of institutional responses. These range from official declarations by state (Government of Spain, 2019) or supranational (European Commission, 2018) bodies about the need for action, to the emergence of regulatory codes (European Commission, EU Code of Practice on Disinformation (2018)) or legislative initiatives all over the world to avoid fake news and disinformation. Regulating this phenomenon is obviously difficult (Pielemeier, 2020) because of the intrinsic risk of limiting freedom of expression (Kaye 2019), which could explain why some regulatory initiatives, such as in the United Kingdom or Italy, have not succeeded. Nevertheless, several laws have already been passed, such as in Germany, where the NetzDG Law of January 1, 2018 requires censorship of content in order not to be punished with heavy fines (Schmitz and Berndt, 2018). In 2018, Malaysia passed an Anti-Fake News Act thereby allowing prison sentences for those who spread fake news, but repealed it the following year (Kaye, 2020). Or, Singapore, whose PAFMA provides significant criminal sanctions for anyone, including service providers, that communicates “false statements”

in Singapore when they know or have reason to believe that it is false and "is likely to be" harmful in a variety of possible ways. (Pielemeier, 2019).

Obviously, this cycle in the emergence of a social concern is not new – it is identified as a potential threat and there is an immediate legislative reaction in the form of criminalisation, with the consequent jeopardization of fundamental rights. It is quite unique that the right that may be affected by these regulations is freedom of expression (Kaye, 2019), and the intimate relationship between the threat and what it entails and the media is certainly unprecedented. The emergence of a concept in the media that quickly becomes a source of social anxieties and object of “disproportionate” demands for criminalization (Hall, Critcher, Jefferson, Clarke, & Roberts, 1978) inevitably leads us to the concept of moral panics coined by Young (1971) and popularized by Cohen (1972). This concept has been used extensively in both criminology and public debate (Garland, 2008) and has been extremely successful in academia to study the response to multiple criminal phenomena. But what is striking is that now it is the media themselves that are at the centre of the equation, not only as the disseminators of that moral panic, but also as the recipients of it. It is in their field of activity (the diffusion of news) that the threat propagates, and some of them (those found in the new informational contexts (Walsh, 2020)) could constitute both the threat as well as the guarantors of security against it. In these circumstances, it has been suggested that social networks have become the object of social anxieties and the source of what has come to be called "technopanics". (Marwick, 2008). In fact, we can already find authors that catalogue the media treatment of fake news as an informational moral panic. Thus, Carlson (2020) maintains that via their discourse on fake news, agents in the field of communication have created a deviant other that encompasses all the concerns derived from the irruption of social networks, while they present themselves as a truthful and trustworthy opposition to that other.

Regardless of the current revision of the concept of moral panic (Garland, 2008; Thompson & Williams, 2013; Jewkes, 2015; Horsley, 2017) and beyond the controversy regarding the concept and the theoretical framework built around it, Cohen's approach has the virtue of drawing attention to the role of the media in the process of creating threats that subsequently legitimize punitive responses. In our view, this approach adequately expresses a process that is underway in relation to fake news, at least in regard to the early stages of identifying something or someone as a threat, and the distorted and simplified dissemination of this threat by the media (Simons, 2019). While concern about fake news and its alleged threat to democracy arises from the media (Silverman & Alexander, 2016; Silverman, 2016), it seems reasonable to consider the possibility that we are witnessing a process, or at least an incipient process, whereby an artificial threat is constructed that can potentially be used by the authorities to implement disproportionate and restrictive regulation of rights. However, what concerns us in this paper is not so much to analyse a "moral panic" to determine whether the aforementioned theoretical framework explains the institutional response to the phenomenon. Rather, the aim is to show that criminalization is being proposed for an issue that is still in its definitional phase, and that it is being ascribed "threat" status when the impact of fake news is only beginning to be measured and when, as we will see below, it is still not clear what is being measured, what is being included, what is being omitted and how, either by action or omission, the authors are producing the process of conceptual attribution of what is later valued as a threat.

Agreements and disagreements about the concept of fake news

Given the political and social interest that the phenomenon of fake news has acquired in recent times and, in particular, given its consideration as a "threat" to the essential

interests of democracies, it is understandable that there have been attempts to empirically "prove" hypotheses of a descriptive nature, such as those relating to the existence of an increase in fake news in certain electoral processes (Silverman, 2016), and hypotheses of the inferential variety, such as those relating the consumption of fake news to certain consequences (Allcott & Gentzkow, 2017). Yet, in reality, a phenomenon is beginning to be measured that, beyond its relationship to the concept of disinformation and the synonymy between them (Wardle & Derekhshan, 2018; Lazer et al., 2018), is still in the process of being defined. As other reviews of the literature have already observed (Tandoc, Linm, & Ling, 2017; Zhang & Ghorbani, 2020), the term fake news is closely linked to concepts such as: satirical news, hyperpartisan news and conspiracy theories. While some research restricts the concept of fake news to manufactured information that pretends to be news (Grinberg, Joseph, Friedland, Swire-Thompson, & Lazer, 2019), others give it a broader meaning (Giglietto, Iannelli, Valeriani, & Rossi, 2019; Bovet & Makse, 2019) and consider fake news to be a category that includes both manufactured news and different combinations of hyperpartisan news, satirical news and conspiracy theories. But these differences in the theoretical contours of fake news, despite having repercussions on an empirical level, are secondary to the articulation of the concept of fake news as a scientific object. The central question, prior to the theoretical delimitation of the concept of fake news, lies in the form of attributing the characteristic "fake" to that which is studied as fake news. This, as we will see below, derives from a conceptual debate that cannot be considered resolved, despite the fact that empirical research is underway.

Perhaps the most established definition of fake news in the academic literature is that of Lazer et al., for whom fake news consists of "fabricated information that mimics news media content in form but not in organizational process or intent" (2018, p. 1094). It

should be noted that this conceptualisation of fake news avoids relating the concept directly to the idea of truth or lies, and emphasises, on the one hand, that this information formally imitates media content and, on the other hand, that it has not been produced following the same processes or with the same intention as news produced by the media. From this definition it is derived that fake news is information that acquires the characteristic of fake because the producer uses neither the same processes nor has the same intentions as the media. In other words, based on this approach the "attribution of fakeness is thus not at the level of the story but at that of the publisher" (Grinberg, Joseph, Friedland, Swire-Thompson, & Lazer, 2019, p. 1).

Leaving process issues aside but maintaining the importance of appearance and of the objectives behind its production, the definition provided by Alcott and Gentzkow (2017) illustrates the other common way to define fake news. For these authors, fake news would be "news articles that are intentionally and verifiably false, and could mislead readers" (2017, p. 213). These authors do introduce the concept of falsehood into the definition and place the attribution of falsehood at the level of the story itself. However, their definition clarifies that not all false information will be fake news, only that which is intentionally false, which can be verified as false and which, moreover, due to its characteristics - of whatever kind - can deceive readers.

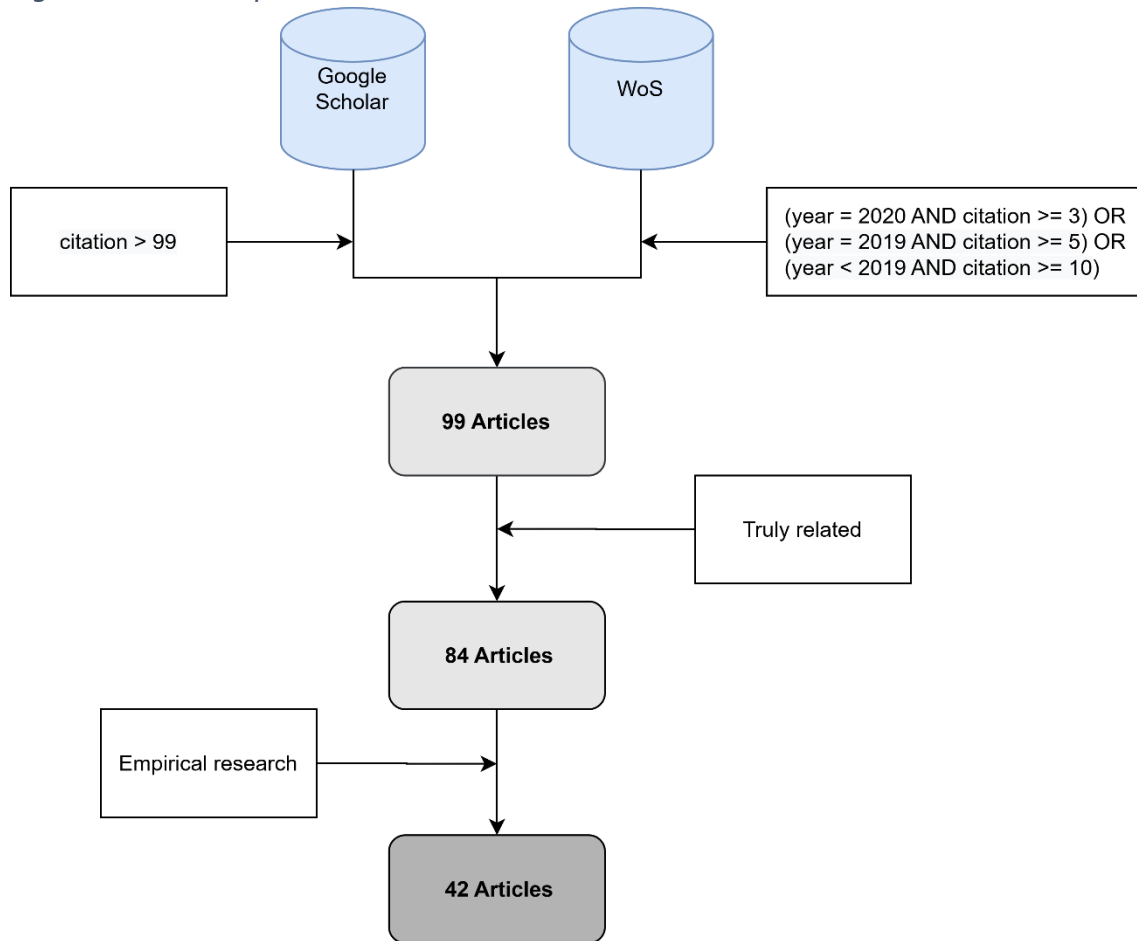
Although the above-mentioned definitions disagree on the origin of fakeness, to consider information fake news both require a certain degree of intentionality on the part of the author and that the information has the potential to not to be recognized as false. Hence, in the systematic revision by Tandoc et al. (2017), on the one hand, the ability of fake news to mimic real news and, on the other hand, the fact that there is no intention to inform, is what has led some authors to blame fake news for generating misinformation (Lazer et al., 2018; Grinberg, Joseph, Friedland, Swire-Thompson, & Lazer, 2019) and

undermining the traditional media system (Tandoc, Linm, & Ling, 2017). Therefore, these definitions leave out cases of systematic and systemic, though unintentional, malpractice, which means omitting from the "problem of fake news" a part of the phenomenon of disinformation that is intimately related to traditional media (Levi, 2019). It can therefore be seen that there is significant agreement on some of the aspects that characterise fake news, as well as significant disagreement and a fundamental unresolved discussion centred on the epistemological question of how the falsehood or "fake" nature of the information should be determined: whether it should be determined on the basis of the content of the news, the characteristics of the issuer or otherwise. Many other discussions can probably be derived from this that are closely related to the difficult question of how to attribute something the condition of truth. In any case, the question that interests us is: Are these consensuses, and also these discrepancies, reflected in the empirical research that is already claiming to study fake news? Do all studies that measure fake news or similar phenomena start from the same concept and use the same method to determine what falls within the object of study? And, above all, does this different conceptual configuration have an impact on the perception of the threat that disinformation and fake news can pose?

In the present paper we inquire about the object that is being studied, about what the empirical research that claims to describe the phenomenon and the impact of fake news is actually measuring. After all, this research can form part of the public debate on the risks that fake news can entail and, therefore, it can influence its regulation, or even criminalization. And if we do not first review the object that empirical research is studying, in an area in which the definitive question clearly seems neither simple nor irrelevant, we run the risk of collaborating in the false configuration of a threat and in the justification of inadequate regulations.

Method

Figure 1: Selection process

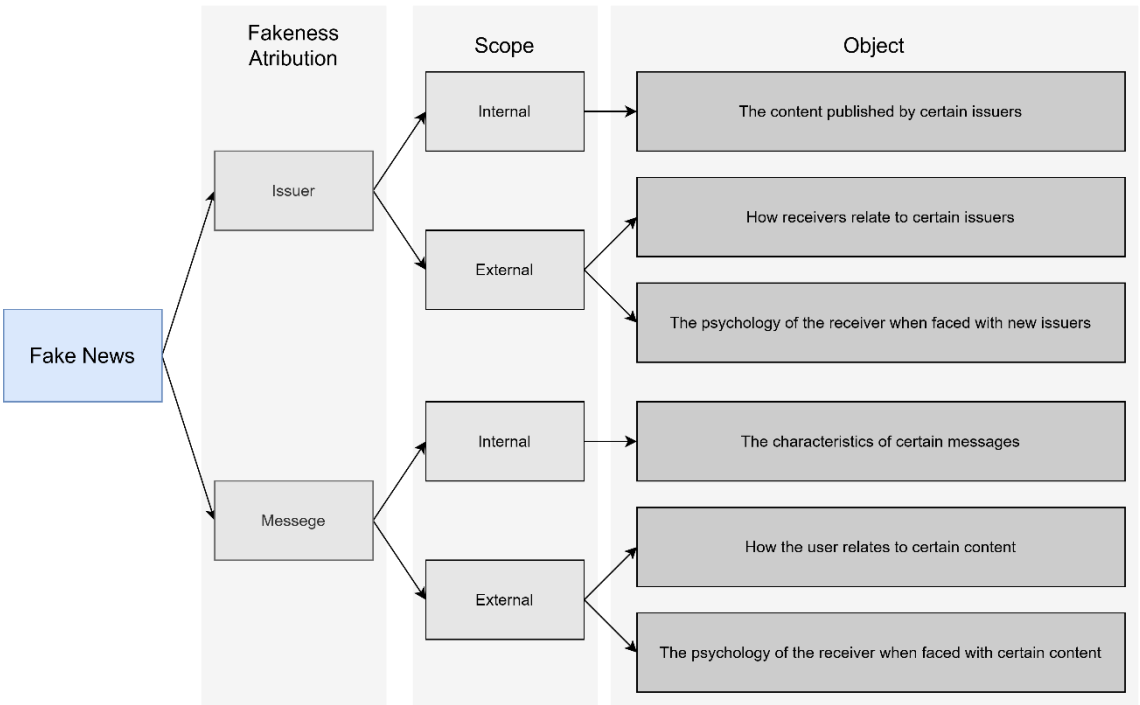


With the aim of addressing the objects that guide empirical research on disinformation, a systematic review of the literature was conducted. To this end, the most relevant articles in which the term fake news was mentioned have been extracted from the Web of Science database. The number of citations was used as the reference criteria to determine the relevance of the articles; thus, we extracted all articles published before 2019 with 10 or more citations, articles published in 2019 with 5 or more citations and articles published in 2020 with three or more citations. This initial search was supplemented by a second Google Scholar search, which yielded 9 articles with more than 100 citations that were not found on Web of Science.

After combining the results from both databases, a list of 99 articles was obtained. These were manually reviewed to distinguish those of a theoretical nature (42 articles) from those that provided their own empirical results, as well as to certify that the articles on the list dealt with the subject matter of interest and did not simply mention the term fake news (15 articles were removed from the final review for this reason). This process produced a list of 42 empirical articles that constitute the object of the present bibliographical review.

Based on the analytical perspective previously outlined and as a result of the bibliographic review, it has been established that research on fake news can be classified into two large groups according to where they seek to identify the origin of its fake nature (*figure 2*).

Figure 2: Classification diagram



Thus, on the one hand, there are studies that attribute the fakeness to the issuer of the message, so that it will be some quality attributable to the issuer that will lead to the information being assumed to be true or false. On the other hand, there are studies in which the characteristics of the information itself will determine, in the eyes of the researchers, whether it is considered fake. In turn, and independently of the method for

attributing fakeness, the different studies can also be divided between those that carry out internal analysis and those that conduct external analysis. The research that conducts external analysis takes fake news as a variable and observes how it relates to other variables, while the research that carries out internal analysis will observe the characteristics of those contents that have been called fake news. When conducting internal analysis, research that attributes the condition of fake news to the issuer, by selecting what they are going to observe based on its source and not on its intrinsic characteristics, are in fact studying the characteristics of the content published by certain issuers; while, when carrying out an external analysis, researchers will be studying how the receivers relate to certain issuers, or, if they take a psychological perspective centred on the subject, they will be studying the psychology of the receiver – personality traits or cognitive skills which could influence how the subject relates to fake news – when faced with these new issuers that seem to have appeared in the field of communication. On the other hand, the studies that attribute some form of falsehood to the messages will be studying the characteristics of these – specifically those to which they attribute the status of fake – when they carry out internal analysis. In the case of external analysis, we can distinguish between: studies on the psychology of the receiver when faced with certain content, if they study the relationship between the human psyche and fake news, and studies on the relationship of the subject with certain content, if we observe how the subject relates to fake news.

Results

As can be seen in table 1, regardless of the way in which news is attributed the characteristic of fake, very few studies carry out internal analyses, with 90% of the reviewed research focusing on external analyses. Within this category, all investigations that attribute fakeness based on the issuer study how the receivers relate to certain issuers.

This perspective has been used by studies that have measured the dissemination and consumption of fake news in certain contexts. It should be noted that this group of research only accounts for 26% of all empirical studies analysed in the present article.

On the other hand, the research that, via external analysis, attributes falsehood as a result of the characteristics of the message are divided between those that study the psychology of the receiver when faced with certain content and those that study how the users relate to that content. This last group, which is mainly composed of proposals for algorithms capable of detecting fake news on social networks, is the most numerous of all, accounting for 40% of all the research reviewed.

Table 1. Number of items per category

Fakeness Attribution (FA)	Scope	Object	N	% in FA category	% total
Issuer	Internal	The content published by certain issuers	2	15.4%	4.8%
	External	How receivers relate to certain issuers	11	84.6%	26.2%
		The psychology of the receiver when faced with new issuer	0	0.0%	0.0%
Message	Internal	The characteristics of certain messages	2	6.9%	4.8%
	External	How the user relates to certain content	17	58.6%	40.5%
		The psychology of the receiver when faced with certain content	10	34.5%	23.8%
Total			42	200.0%	100.0%

About the issuer

Many of the studies analysed have chosen to determine the fakeness of the news not by its content, but by its source, so that the "fake" status of the information is determined by its issuer. Thus, a fake news item is conceptualized, in a broad sense, as a piece of information that comes from a website that pretends to be a real media outlet (Lazer et al., 2018). This way of understanding fake news has been prominent in the empirical studies that have tried to measure the diffusion or consumption of fake news among the population, as it allows the news consumed or shared by the subject to be configured as independent variables, without the need to individually evaluate the content of each piece.

Furthermore, this approach has also allowed research on the patterns of dissemination of certain content through social networks, which has led to botnets being detected as significant collaborators in this task (Bastos & Mercea, 2017; Shao, Ciampaglia, Varol, Flammini, & Menczer, 2018; Shao, et al.). However, this way of approaching fake news via the issuer is not exclusive to external analyses. We also find internal analyses that study the characteristics of the content created and disseminated by these issuers (Bakir & McStay, 2018; Mourão & Robertson, 2019).

Using the issuer to determine the "fake" nature of information makes it necessary to determine which sources are actually considered media, or legitimate media, and which are not. In order to solve this problem, the observed studies resort to categorizations of media elaborated by academics, media or fact-checking organizations (such as pensources.co², Buzzfeed³, FactCheck.org⁴ and politifac.com⁵).

About the message

The other conceptual model for the study of fake news places the condition of "fake" in the study of the content, and not the issuer. From this perspective, research has also been carried out to measure the importance and extent of the phenomenon of fake news in certain political contexts, such as the 2016 US presidential elections (Allcott & Gentzkow, 2017) or the 2014 Ukraine-Russia conflict (Khaldarova & Pantti, 2016). This way of attributing falsehood to fake news has also been used by the different studies on the psychology of the receiver when faced with content considered to be fake (De Keersmaecker & Roets, 2017; Bronstein, Pennycook, Bear, Rand, & Cannon, 2019; Pennycook & Randa, 2019). In both cases, for the construction of the variable "fake news" we no longer find media lists, but we do find fake news lists.

It should be noted that, within the studies that attribute the fakeness of fake news to its content, the most numerous group is made up of proposals for models trained by machine and deep learning for the detection of this type of content. Those models that carry out internal analysis use features based on questions related to the use of language – grammar, syntax, etc. (Potthast, Kiesel, Reinartz, Bevendorff, & Stein, 2018; Pérez-Rosas, Kleinberg, & Lefevre, 2018). On the other hand, research that carries out external analysis uses as features the patterns of dissemination of false information in social networks, as well as characteristics associated with the user posts through which they are disseminated – characteristics of the account, of the message or the presence and characteristics of associated images (Ruchansky, Seo, & Liu, 2017; Jin, Cao, Zhang, Zhou, & Tian, Novel Visual and Statistical Image Features for Microblogs News Verification, 2017; Wang, et al, 2018; Shin, Jian, Driscoll, & Bar, 2018; Reis, Correia, Murai, Veloso, & Benevenuto, 2019; Alrubaian, Al-Qurishi, Mehedi-Hassan, & Alamri, 2018; Jin, Cao, Zhang, & Luo, 2016; Jang, et al.).

As with the previous model, to access this false information, researchers use fake news indexes that are also developed by certain official news agencies or media outlets that are dedicated to fact-checking. Thus, with the exception of the proposals from Pérez-Rodas et al. (2018), which builds its database by inventing fake news, and Alrubaian et al. (2018), in which the participating researchers verify the information, the different algorithms proposed for the detection of fake news have been trained on databases that have been constructed using information labelled as fake news by the aforementioned fact-checking organizations (news agencies such as Xinhua News Agency, or fact-checkers such as snoopers.com). However, in this kind of approach the object is no longer the fake news itself, understood as the "original" false information. The focus of the research, and consequently of the classification model, becomes the characteristics of the

vehicles, accounts and messages through which fake news is disseminated and the patterns of dissemination. In other words, the research focuses on how users relate to certain content but also endows them an active role in the production of misinformation.

The impact of fake news on democracy

Finally, it should be noted that few studies were found that address the impact of fake news. Of the articles reviewed, only 8 based their research on a specific context in such a way that it was possible to use relevant empirical material to attempt to gauge the impact of fake news within that context. With the exception of one of them - related to Brexit (Bastos & Mercea, 2017) - the remaining 7 studies examine fake news in the context of the 2016 American election (Allcott & Gentzkow, 2017; Nelson & Taneja, 2018; Shao et al, 2018; Grinberg, Joseph, Friedland, Swire-Thompson, & Lazer, 2019; Guess, Nagler, & Tucker, 2019; Bovet & Makse, 2019; Guess, Nyhan, & Reifler, 2020). These studies have taken different perspectives (table 2) but their results are consistent, as all found substantially small and highly concentrated diffusion and consumption of fake news amongst a specific profile of subjects, which significantly weakened the initial hypotheses about the relationship between fake news and Donald Trump's victory (Silverman, 2015; Mihailidis & Viotty, 2017). Studies such as Guess, Nyhan and Reifler (2020) have estimated that fake news accounted for 5.9% of the news consumed by each user in the month prior to the elections. With regard only to Twitter, Grinberg and co-authors (2019) observed that during the month prior to the elections, each user was exposed to fake news related to the political campaign 10 times on average, only 1.18% of the user's total exposure to political news. This same research also found that 1% of their sample consumed 80% of the detected fake news. These big consumers of fake news were mainly conservative and were characterized by high consumption of all kinds of news. This

conclusion regarding the profile of consumers of fake news is shared with other studies reviewed herein (Allcott & Gentzkow, 2017; Nelson & Taneja, 2018; Guess, Nyhan, & Reifler, 2020; Guess, Nagler, & Tucker, 2019). These findings are summarized in table 2, and as will be considered in greater detail in the discussion section, none of the existing studies allow a causal relationship to be established between the results of the elections and fake news.

Table 2: Articles that analyse the impact of fake news on the 2016 US presidential elections

Fakeness Attribution (FA)	Main data source	N	% total
Issuer	Poll	0	0%
	Web traffic (navigation history)	2	28.6%
	Social network data	4	57.1%
Message	Poll	1	14.3%
	Web traffic (navigation history)	0	0%
	Social network data	0	0%
Total		7	100%

Discussion and critical reflection

The possible impact of disinformation on the 2016 US presidential elections brought into public debate the potential need to regulate fake news on social networks, either through criminalization or by forcing social media to prevent its dissemination. In turn, this gave rise to an important debate on the role that Internet service providers play in regulating political debate and shaping freedom of expression (Kaye, 2019). The premise on which all the regulatory proposals are founded is that misinformation, as a genus, and fake news as a species, constitute a serious threat to essential values such as democracy, public health, or free public opinion. Empirical studies have begun to emerge, which attempt to confirm some of these premises or to detect "fake news", despite the enormous difficulties posed by the measurement of phenomena that are conceptually not very precise, or to determine causal inferences in political processes or similar areas of social decision making. The results of our study show, however, a tendency to simplify the

characterization of the conditions that constitute "Fake News" for the purposes of its empirical measurement.

We have seen that a large number of studies on fake news introduce a concept that is related to the incursion of new actors into the field of communication who do not follow the traditional verification processes and to whom the characteristic "fake" is implicitly attributed. Without denying the possible relationship between source and truth, this omits certain news from the phenomenon of fake news and disinformation that, because of its content, may be clearly untrue but that does not come from sources categorized as fake or is even disseminated by legitimate media. Furthermore, this ignores the possibility, hypothesized by certain authors (Silverman, 2015; Mihailidis & Viotty, 2017) and rejected by others (Guo & Vargo, 2020), that the content disseminated by these new actors can influence the content of legitimate media sources by filtering into them or conditioning their agenda. Moreover, the vast majority of the empirical research on fake news depends on pre-constituted lists of information or media that are prepared by fact-checkers or government news agencies. In both cases, the criteria for the attribution of the status of "fake" and for the elaboration of the lists are not transparent, and neither are they scientific since none of these organizations principally carry out scientific activities. It is true that the models proposed for the detection of fake news in social networks, which, as we have seen, constitute 36% of all the empirical research reviewed, deserve separate attention. This is because although these studies also depend on the sources from which they extract what they consider to be fake news, the very nature of machine and deep learning is the generation of tools that transcend the database and are capable of learning beyond it. Furthermore, this way of approaching the phenomenon is not limited to observing user interaction with certain agents that disseminate fake news, but it also observes how users disseminate this information, how they modify and re-distribute it,

and how they react to it and to other users. This is, therefore, a broader and more ambitious way of approaching disinformation, since it is not limited to the origin of the information but takes into account the waves -the different reactions- generated by the appearance of false information in a social network. Right now, this kind of research says little about the phenomenon of disinformation because it has been limited to developing tools in "laboratories", that is, within databases created specifically for this purpose, but the application of this type of tools "to the real world" could have great potential to generate knowledge.

It is true that the dependence on the sources from which the potential fakes are extracted is understandable, given the philosophical complexity intrinsic to any phenomenon based on the ideas of "truth" or "lie". However, it is undeniable that such a procedure introduces significant bias into research that aims to study fake news and forces us to be careful when establishing consequences from these studies. The question of disinformation, misinformation and fake news is a complex issue that cannot be reduced to the circulation of information on social networks by websites that pretend to be media and are obviously false or partisan. On several occasions, the theoretical and empirical literature on this subject has drawn attention to the need to observe the role of the media considered legitimate or traditional in disinformation (Brulle, Carmichael, & Jenkins, 2012; Lewandowsky, Ecker, & Cook, 2017). From this perspective, issues such as click-bait (Rochlin, 2017), the primacy of opinion over expert analysis (Case & Given, 2016), or the speed of publication imposed by their own corporate structure (Cooke, 2017) cannot be ignored if we are to understand the phenomenon in its entirety and its potential effects. If one wishes to study the dissemination of fake news and the impact that it can have as a "potential threat to democratic systems", one must at least explore the possibility that governments and large media institutions have some role in the propagation of this type

of content. Likewise, if the objective is more ambitious and one wishes to approach disinformation and misinformation as a social phenomenon, the logics and practices of the main actors in the field of communication -large media institutions, news agencies, press offices, etc.- cannot be completely ignored. Ignoring the questions we have just mentioned and "borrowing" the object of study means that scientific research into fake news runs the risk of being caught up in the game of exchanging meanings and legitimacy that, as in any other field (Bourdieu & Wacquant, 2005), is disputed in the field of communication.

This is not to deny the usefulness, even the necessity, of conducting empirical research on this, let us say, restricted concept of fake news. But in addition to the fact that we must demand an explanation of the concept that is being used as a starting point and of the manner in which the condition of fake is attributed, we believe that it is essential to make it compatible with other empirical research. The present paper establishes the need to adopt a clear, as well as broad, concept of the phenomenon of disinformation that, in turn, can be empirically articulated in different typologies. These should, at a minimum, pay attention to the way in which something is considered fake, whether it is through the medium or through the message, and make the process of attribution explicit so that it is possible to debate and refute the procedure. What we have seen in the literature analysed in the present paper is that, despite the fact that what guides and justifies the research is disinformation and its potential harmful effects, what is ultimately studied, either because of lack of reflection on the concept of fakeness, or because the attribution of fakeness is not dealt with scientifically, are questions related to certain contents or certain issuers whose relationship with fake news is not sufficiently demonstrated -beyond the fact that someone has said at some point that this content was fake news or that this website disseminated fake news. We do not believe it is appropriate to study a social phenomenon

merely as a given natural object that is not previously configured by any type of conceptualization, especially if the intention is to analyse the consequences of this object and to derive from this research responses that are related to control, normative regulation or criminalization.

In any case, and broaching the second question that we intend to discuss in this paper, we must emphasize that the empirical data available right now, despite their limitations, do not allow us to establish a causal relationship between this concept of fake news and the harmful consequences that would theoretically make it a threat and that would justify criminalization. Thus, given what we know right now about the phenomenon, the proposals for criminalization, as well as cataloguing it as a threat, may be remarkably disproportionate. As the United Nations Special Rapporteur on Freedom of Expression has recently emphasized, when a state is preparing to restrict freedom of expression in some way, which includes freedom of information, it is not enough that the objective pursued be legitimate, but that "It must establish a direct and immediate connection between the expression and the threat said to exist", furthermore, this restriction "must be the least intrusive instrument among those which might achieve the desired result" (Kaye, 2020, p. 6). It should also be borne in mind that restricting freedom of expression can have the well-known chilling effect (Miró-Llinares & Gómez-Bellvís, 2020), which is particularly harmful since freedom of expression is a fundamental right that is indispensable for other fundamental rights, such as the right to information or to free public opinion, without which there can be no democratic society (Alcácer-Guirao, 2012). Furthermore, research on the concept of fake news is being conducted empirically using lists of stories and media that have been identified as fake news by actors in the field of communication, ignoring any reflection on the fakeness of the news. Thus, it seems necessary, firstly, to clearly state that we are still in an initial phase of empirical research

that has not been able to certify the threatening nature of fake news, and, secondly, to demand new areas of empirical research on the phenomenon to avoid turning research into a vehicle for the legitimization of the criminalizing discourses that circulate in the field of communication and the political sphere.

¹ It should be noted that some predictions did point to Donald Trump as a possible winner, such as those made by survey aggregator FiveThirtyEight: <https://fivethirtyeight.com/features/the-real-story-of-2016/>

² Opensources.co is a project by Merrimack College that offers a list of news sites categorized in: *fake*, *satire*, *hate* and *clickbait*. This classification is based on the content of the publications, their *about us* section, the fonts used and the writing style.

³ BuzzFeed is a media outlet that, during the 2016 US presidential elections, carried out a series of reports on the dissemination of fake news on Facebook. During the year following the elections, the journalist responsible for the investigations, Craig Silverman, continued to update the list of websites in different reports until it reached 233 websites.

⁴ Politifact is a portal composed of journalists and is primarily dedicated to the verification of news in North America. In December 2016, this organization began to collaborate with Facebook on the verification of news that circulated through the social network. From this experience, they elaborated a list of 324 entries with the websites from which fake news on the social network came most frequently.

factcheck.org is a project that acts as a "consumer advocate" for American voters⁵. Also, as a result of its collaboration with Facebook, in July 2017 it published a *Misinformation Directory* with websites dedicated to the dissemination of *fake news*.

References

- Alcácer-Guirao R (2012) Hate speech and political discourse: In defense of the freedom of the intolerant. *Electronic Journal of Criminal Science and Criminology* 14(2): 1-32.
- Allcott H and Gentzkow M (2017) Social Media and Fake News in the 2016. *Journal of Economic Perspectives* 31(2): 211-236. DOI:10.1257/jep.31.2.211
- Alrubaian M, Al-Qurishi M, Mehedi-Hassan M and Alamri A (2018) A Credibility Analysis System for Assessing Information on Twitter. *IEEE Transactions on Dependable and Secure Computing* 15: 661-674.

- Bakir V and McStay A (2018) Fake News and The Economy of Emotions. *Digital Journalism* 6(2): 154-175. DOI:10.1080/21670811.2017.1345645
- Bastos MT and Mercea D (2017) The Brexit Botnet and User-Generated Hyperpartisan News. *Social Sciences Computer Review* 37(1): 38-54. DOI:10.1177/0894439317734157
- Baym G (2005) The Daily Show: Discursive Integration and the Reinvention of Political Journalism. *Political Communication* 22: 259-276. DOI:10.1080/10584600591006492
- Bennett WL and Livingston S (2018) The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication* 33(2): 122-139. DOI:10.1177/0267323118760317
- Berkowitz D and Schwartz D (2016) Miley, CNN and The Onion: When fake news becomes realer than real. *Journalism Practice* 1(10): 1-17. DOI:10.1080/17512786.2015.1006933
- Bourdieu P and Wacquant L (2005) *Una invitación a la sociología reflexiva*. Buenos Aires: Siglo XXI.
- Bourdieu P, Chamboredon JC and Passeron JC (2002) *El oficio del Sociólogo*. Buenos Aires: Siglo XXI.
- Bovet A and Makse HA (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications* 10(7): 1-14. DOI:10.1038/s41467-018-07761-2
- Brewer PR, Goldthwaite D and Morreale M (2013) The Impact of Real News about “Fake News”: Intertextual Processes and Political Satire. *International Journal of Public Opinion Research* 25(3): 323–343. DOI:10.1093/ijpor/edt015
- Bronstein MV, Pennycook G, Bear A et al. (2019). Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic

Thinking. *Journal of Applied Research in Memory and Cognition* 8(1): 108-117.

DOI:10.1016/j.jarmac.2018.09.005

Brulle R J, Carmichael J and Jenkins JC (2012) Shifting public opinion on climate change: An empirical assessment of factors influencing concern over climate change in the U.S., 2002–2010. *Climatic Change* 114: 169-188. DOI:10.1007/s10584-012-0403-yBu,

Carlson M (2020) Fake news as an informational moral panic: the symbolic deviancy of social media during the 2016 US presidential. *Information, Communication & Society* 23(3): 374–388. DOI:10.1080/1369118X.2018.1505934

Cas DO and Given LM (2016) *Looking for information: A survey of research on information seeking, needs, and behavior*. Bingley: Emerald.

Cohen S (1972). *Folk devils and moral panics: The creation of the mods and rockers*. London: Paladin.

Cooke NA (2017) Posttruth, Truthiness, and Alternative Facts: Information Behavior and Critical Information Consumption for a New Age. *Library Quarterly: Information, Community, Policy* 87(3): 211-221. DOI:10.1086/692298

De Keersmaecker J and Roets A (2017) ‘Fake news’: Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence* 65: 107-110.

European Commission (2018) *Action Plan Against Disinformation*. Available at: https://ec.europa.eu/commission/sites/beta-political/files/eu-communication-disinformation-euco-05122018_en.pdf

Europol (2020) *Catching the virus cybercrime, disinformation and the COVID-19 pandemic*. Available at: <https://www.europol.europa.eu/publications-documents/catching-virus-cybercrime-disinformation-and-covid-19-pandemic>

- Farkas J and Schou J (2018) Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood. *Javnost - The Public* 25(3): 298-314.
DOI:10.1080/13183222.2018.1463047
- Fedeli G (2020) 'Fake news' meets tourism: a proposed research agenda. *Annals of Tourism Research* 80: 1-5.
- Garland D (2008) On the concept of moral panic. *Crime, Media, Culture* 4(1): 9-30.
DOI:10.1177/1741659007087270
- Giglietto F, Iannelli L, Valeriani A and Rossi L (2019) 'Fake news' is the invention of a liar: How false information circulates within the hybrid news system. *Current Sociology Monograph* 67(4): 625-642. DOI:10.1177/0011392119837536
- Government of Spain. (2019). *Annual National Security Report 2019*. Madrid: Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática. Available at: <https://www.dsn.gob.es/es/documento/informe-anual-seguridad-nacional-2019> (accessed 28 June 2020).
- Grinberg N, Joseph K, Friedland L, et al. (2019) Fake news on Twitter during the 2016 U.S. presidential election. *Science* (363): 374-378. DOI:10.1126/science.aau2706
- Guess A, Nagler J and Tucker J (2019) Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advance* 5: 1-8.
DOI:10.1126/sciadv.aau4586
- Guess A, Nyhan B and Reifler J (2020) Exposure to untrustworthy websites in the 2016. *Nature human behaviour* 4: 472–480. DOI:10.1038/s41562-020-0833-x
- Guo L and Vargo C (2020) "Fake News" and Emerging Online Media Ecosystem: An Integrated Intermedia Agenda-Setting Analysis of the 2016 US Presidential Election. *Communication research* 47(2): 178-200. DOI:10.1177/0093650218777177

- Haciyakupoglu G, Yang Hui J, Suguna VS, et al. (2018) *Countering fake news: A survey of recent global initiatives*. Nanyang Technological University. Available at: <https://www.rsis.edu.sg/rsis-publication/cens/countering-fake-news-a-survey-of-recent-global-initiatives/#.XxVZGW5uLHo>
- Hall S, Critcher C, Jefferson T, et al. (1978) *Policing the crisis: Mugging, the state, and the law and order*. London: Macmillan.
- Horsley M (2017) Forget 'Moral Panics'. *Journal of Theoretical and Philosophical Criminology* 9(2): 84-98.
- Interpol (2020) *Cybercrime: COVID-19 impact*. Lyon: Interpol. Available at: <https://www.interpol.int/News-and-Events/News/2020/INTERPOL-report-shows-alarming-rate-of-cyberattacks-during-COVID-19>
- Jang SM, Geng T, Queenie JY, et al. (2018) A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis. *Computers in Human Behaviour* 84: 103-113. DOI:10.1016/j.chb.2018.02.032
- Jewkes Y (2015) *Media and Crime* (Third ed.). London: Sage.
- Jin Z, Cao J, Zhang Y and Luo J (2016) News verification by exploiting conflicting social viewpoints in microblogs. In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. Phoenix: Association for the Advancement of Artificial Intelligence, pp.2972–2978.
- Jin Z, Cao J, Zhang Y, et al. (2017) Novel Visual and Statistical Image Features for Microblogs News Verification. *IEEE TRANSACTIONS ON MULTIMEDIA* 19(3): 598-608. DOI:10.1109/TMM.2016.2617078
- Kaye D (2019) *Special Report to the United Nations Human Right Counil: Freedom of Expression and Elections in the Digital Age*. Genova: United Nations Human Rights

Council. Available at:

<https://www.ohchr.org/Documents/Issues/Opinion/ElectionsReportDigitalAge.pdf>

Kaye D (2020) *Special Report to the United Nations Human Right Council: Disease pandemics and the freedom of opinion and expressio*. Genova: United Nations Human Rights Council. Available at: <https://undocs.org/A/HRC/44/49>

Khaldarova I and Pantti M (2016) Fake news: The narrative battle over the Ukrainian conflict. *Journalism Practice* 10(7): 1-11. DOI:10.1080/17512786.2016.1163237

Lazer DM, Baum MA, Benkler Y, et al. (2018) The science of fake news. *Science* 359(6380): 1094-1096. DOI:10.1126/science.aao2998

Lee H and Lee T (2019) From contempt of court to fake news: public legitimisation and governance in mediated Singapore. *Media International Australia* 173(1): 81-92. DOI:10.1177/1329878X19853074

Lewandowsky S, Ecker UK and Cook J (2017) Beyond Misinformation: Understanding and Coping with the“Post-Truth” Era. *Journal of Applied Research in Memory and Cognition* 6: 353–369. DOI:10.1016/j.jarmac.2017.07.008

Lowry E (2013) The Flower of Cuba: Rhetoric, Representation, and Circulation at the Outbreak of the Spanish-American War. *Rhetoric Review* 32(2): 174-190. DOI:10.1080/07350198.2013.766852

Marchi R (2012) With Facebook, Blogs, and Fake News, Teens Reject Journalistic “Objectivity”. *Journal of Communication Inquiry* 36(3): 246-262. DOI:10.1177/0196859912458700

Marwick AE (2008) To catch a predator? The MySpace moral panic. *First monday* 13(6). DOI:10.5210/fm.v13i6.2152

McGonagle T (2017) “Fake news”: False fears or real concerns? *Netherlands Quarterly of Human Rights* 35(4): 203-209. DOI:10.1177/0924051917738685

- Mihailidis P and Viotty S (2017) Discursive Deflection: Accusation of "Fake News" and the Spread of Mis- and Disinformation in the Tweets of President Trump. *American Behavioral Scientist* 61(4): 441-454. DOI:10.1177/0002764217701217
- Miró-Llinares F and Gómez-Bellví AB (2020). Freedom of expression in social media and criminalization of hate speech in Spain: Evolution, impact and empirical analysis of normative complicity and self-censorship. *Spanish Journal of Legislative Studies* 1: 1-42. DOI:10.21134/sjls.v0i1.1837
- Mourão RR and Robertson CT (2019) Fake News as Discursive Integration: An Analysis of Sites That Publish False, Misleading, Hyperpartisan and Sensational Information. *Journalism Studies* 20(14): 1-19. DOI:10.1080/1461670X.2019.1566871
- Nelson JL and 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. DOI:10.1177/1461444818758715
- Parliament of Singapore (2018). *Report of the select committee on deliberate online falsehoods: Causes, consequences and countermeasures*. Available at: <https://www.parliament.gov.sg/sconlinefalsehoods>
- Pennycook G and Rand DG (2019) Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188: 39-50. DOI:10.1016/j.cognition.2018.06.011
- Pérez-Rosas B, Kleinberg B and Lefevre A (2018) Automatic Detection of Fake News. In: *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe: Association for Computational Linguistics, pp.3391-3401.
- Pielemeier, JS (2020) Disentangling Disinformation: What Makes Regulating Disinformation So Difficult? *Utah L. Review* 917, Available at: <https://ssrn.com/abstract=3629541>

- Potthast M, Kiesel J, Reinartz K, et al. (2018) A Stylometric Inquiry into Hyperpartisan and Fake News. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. Melbourne: Association for Computational Linguistics, pp. 231-240.
- Reis JC, Correia A, Murai F, et al. (2019) Supervised Learning for Fake News Detection. *IEEE Intelligent Systems* 34(2): 76-81. DOI:10.1109/MIS.2019.2899143.
- Rochlin N (2017) Fake news: belief in post-truth. *Library Hi Tech* 35(3): 386-392. DOI:10.1108/LHT-03-2017-0062
- Ruchansky N, Seo S and Liu Y (2017) CSI: A Hybrid Deep Model for Fake News Detection. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. DOI:10.1145/3132847.3132877
- Schimitz, S and Berndt, C (2018) The German Act on Improving Law Enforcement on Social Networks (NetzDG): A Blunt Sword? *SSRN*, 14 december. Available at <https://ssrn.com/abstract=3306964>
- Shao C, Ciampaglia GL, Varol O, et al. (2018) The spread of low credibility by social bots. *Nature Communications* 9: 1-9. DOI:10.1038/s41467-018-06930-7
- Shao C, Hui PM, Wang L, et al. (2018) Anatomy of an online misinformation network. *PLoS ONE* 13(4): 1-23. DOI:10.1371/journal.
- Shin J, Jian L, Driscoll K and Bar F (2018). The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior* 83: 278-287. DOI:10.1016/j.chb.2018.02.008
- Silverman C (2015) *Lies, Damn Lies and Viral Content*. Tow Center for Digital Journalism. Available at: <https://academiccommons.columbia.edu/doi/10.7916/D8Q81RHH>

- Silverman C (2016) This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook. *BuzzfeedNews*, 16 November. Available at: <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>
- Silverman C and Alexander L (2016). How Teens In The Balkans Are Duping Trump Supporters With Fake News. *BuzzfeedNews*, 3 November. Available at: <https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo>
- Simons G (2019) The Anatomy of a Moral Panic: Western Mainstream Media's Russia Scapegoat. *Changing Societies & Personalities* 3(3): 189–206. DOI:10.15826/csp.2019.3.3.071
- Spencer DR (2007) *The Yellow Journalism: The Press and America's Emergence as a World Power*. Evanston: Northwestern University Press,.
- Tandoc EC (2019) The facts of fake news: A research review. *Sociology Compass* 13(9): 1-9. DOI:10.1111/soc4.12724
- Tandoc EC, Linm ZW and Ling R (2017) Defining “Fake News”: A typology of scholarly definitions. *Digital Journalism* 6(2): 137-153. DOI:10.1080/21670811.2017.1360143
- Thompson B and Williams A (2013) *The Myth of Moral Panics: Sex, Snuff, and Satan*. New York: Routledge.
- Walsh JP (2020) Social media and moral panics: Assessing the effects of technological change on societal reaction. *International Journal of Cultural Studies*: 1-20. DOI:10.1177/1367877920912257
- Wang Y, Ma F, Jin Z, et al. (2018) EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. In: *Proceedings of the 24th ACM SIGKDD*

International Conference on Knowledge Discovery & Data Mining. New York: Association for Computing Machinery, pp. 849–857.
DOI:10.1145/3219819.3219903

Wanga Y, McKeen M, Torbica A and Stuckler D (2019). Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. *Social Media and Medicine* 240: 1-12. DOI:10.1016/j.socscimed.2019.11255

Wardle, C and Derakhshan H (2017) Information disorder: Toward an interdisciplinary framework for research and policy making. Strasbourg: Council of Europe report.

WHO (2020) *Coronavirus Disease 2019: Situation Report-45*. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200305-sitrep-45-covid-19.pdf?sfvrsn=ed2ba78b_4

Young J (1971) *The Drugtakers: The social meaning of drugs use*. London: Paladin.

Zhang X and Ghorbani AA (2020) An overview of online fake news: Characterization, detection, and discussion. *Information Processing and Management* 57(2): 1-26.
DOI:10.1016/j.ipm.2019.03.004