Fake news on the Internet: the strategy to battle misinformation

Diciembre 2017

CELE

UP
Universidad
de Palermo

Fake news on the Internet:
the strategy to battle misinformation

Carlos Cortés y Luisa Isaza**

I. Introducción

On December 4, 2016, a month after the US presidential election, Edgar Maddison Welch, a 28-year-old man from North Carolina, fired three shots at the Comet Ping Pong pizzeria in Washington DC. He was determined to enter, investigate and rescue the children who were being exploited by an alleged child sex trafficking network run by Hillary Clinton and other members of the Democratic Party.[1]

Maddison Welch was convinced of the existence of such network from news he had read on the Internet. After his arrest and despite the fact that the police denied the story - known as *Pizzagate* - the man apologized but never admitted that the information that had motivated his attack was fake.[2] The outrageous story was spread through social media and discussion forums on the Internet, along with hundreds of false stories related to the two candidates or other members of their parties.[3]

Weeks before the election, millions of people saw in their Facebook News Feed a piece of news about an unprecedented statement from Pope Francis proclaiming his support for the candidacy of Donald Trump. This piece of fake news received 960,000 interactions in the social network (comments, reactions and shares), more than any other real news about the election.[4] As a matter of fact, according to a BuzzFeed study released days after the election, the top twenty fake news available in Facebook during the three months prior to the election had more engagement than the twenty main actual stories from the most well-known media outlets (New York Times, Washington Post, Los Angeles Times, Wall Street Journal, FOX News, among others) published on the same social media network.

The Guardian and BuzzFeed revealed that many of these stories were being produced by a group of young Macedonians, who through suggestive headlines that produced clicks - a deceptive technique known as clickbait - made thousands of dollars in advertising thanks to the traffic on their Internet sites.[5] Over a hundred websites were created

---

*   Lawyer, University of Los Andes (Colombia), Master's Degree, Media and Communication Governance, London School of Economics. He is a consultant in freedom of expression and Internet regulation, and external advisor to Twitter's public policy team.

** Lawyer, Universidad Javeriana (Colombia). Legal advisor, Defense and Attention to Journalists Office, Fundación para la Libertad de Prensa (FLIP), Colombia.

1   Siddiqui, Faiz and Svrluga, Susan, "N.C. man told police he went to D.C. pizzeria with gun to investigate conspiracy theory", *Washington Post*, December 5, 2016. Retrieved from: http://wapo.st/2gW0yFo.

2   Goldman, Adam, "The Comet Ping Pong Gunman Answers Our Reporter's Questions", *The New York Times,* December 7, 2016. Retrieved from: http://nyti.ms/2Er1xrM.

3   Other pieces of fake news related to the elections and widely circulated stated, for example, that it had been confirmed that Hillary Clinton had sold weapons to ISIS, that Donald Trump was offering one-way tickets to Africa and Mexico for those who wanted to leave the country and that an ISIS leader was asking Muslims to vote for Clinton. Silverman, Craig, "Here Are 50 Of The Biggest Fake News Hits On Facebook From 2016", *Buzzfeed,* December 30, 2016. Retrieved from: http://bzfd.it/2Ge4ZXo.

4   Silverman, Craig, "This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook", *Buzzfeed,* November 16, 2016. Retrieved from: http://bzfd.it/2ssm265.

5   Tynan, Dan, "How Facebook powers money machines for obscure political 'news' sites", *The Guardian,* August 24, 2016. Retrieved from: http://bit.ly/2bhzDzv; Silverman, Craig y Alexander, Lawrence "How Teens In The Balkans Are Duping Trump Supporters With Fake News", *BuzzFeed,* November 3, 2016. Retrieved from: http://bzfd.it/2EubVDu; Ohlheiser, Abby, "This is how Facebook's fake-news writers make money", *The Washington Post,* November 18, 2016. Retrieved from: http://wapo.st/2Bt9wGk.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

2

for this purpose in the city of Veles, Macedonia, all designed to look like authentic news portals. Other fake news factories operated directly within the United States.[6] According to their own creators, much of the traffic in these sites came from clicks originated in Facebook, where they also had hundreds of thousands of followers.

The false information was spread until the day of the election in the form of hoaxes, fake Twitter accounts, misinformation tweets and even Google search results.[7] After the vote, Mediaite reported that the first result in the Google search engine to "final vote count 2016" was a site called 70News where it was falsely stated that Donald Trump had won both the electoral and popular votes.[8]

After Donald Trump's surprising victory, the discussion about fake news exploded.[9] Some people - including a fake news author- said that misinformation in social networks had directly influenced the outcome of the elections.[10] And although at the moment there is no study to measure this impact in a clear way, it is undeniable that Facebook, Twitter and Google did play an important role as the main source of information for many people.[11]

The services of these companies were not only exploited by astute young people to make profits out of advertising. According to the findings of American intelligence, the Russian government used these platforms to spread fake news and propaganda, seeking to influence public debate during the campaign and benefit Trump's candidacy to the detriment of Clinton's.[12] More recently, the Prime Minister of the United Kingdom, Theresa May, made similar accusations against the Russian government.[13]

6   Silverman, Craig y Singer-Vine, Jeremy, "The True Story Behind The Biggest Fake News Hit Of The Election", *BuzzFeed,* December 16, 2016. Retrieved from: http://bzfd.it/2BZ7kHs.

7   Rogers, Kathy and Bromwich, Jonah, "The Hoaxes, Fake News and Misinformation We Saw on Election Day", *The New York Times,* November 8, 2016. Retrieved from: http://nyti.ms/2o4NhOA.

8   Abrams, Dan, "Now Even Google Search Aiding in Scourge of Fake, Inaccurate News About Election 2016", *Mediate,* November 13, 2016. Retrieved from: http://bit.ly/2ssmbq9.
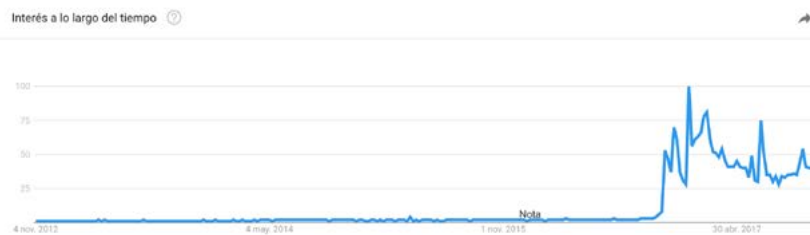
9   A Google Trends search reveals that searches of the term *fake news* shot up in November, 2016, after being next to insignificant for the past few years. The peak of searches happened in the week of January 8 to 14, 2017, which is when the famous incident occurred in which Donald Trump refused to answer a question from a CNN journalist, saying " You are fake news! " Google Trends, http://bit.ly/2Gfn3QZ, last access: December 4, 2017.

10 Read, Max "Donald Trump Won Because of Facebook", *New York Magazine,* November 9, 2016. Retrieved from: http://nymag.com/selectall/2016/11/donald-trump-won-because-of-facebook.html; Dewey, Caitlin, "Facebook fake-news writer: 'I think Donald Trump is in the White House because of me'", *The Washington Post,* November 17, 2016. Retrieved from: http://wapo.st/2f3NllC; Parkinson, Hannah, "Click and elect: how fake news helped Donald Trump win a real election", *The Guardian,* November 14, 2016. Retrieved from: http://bit.ly/2fSyaDH.

11 According to the latest study on digital news from the Reuters Institute for the Study of Journalism at the University of Oxford, a survey conducted in 36 countries revealed that 54% of Internet users use social media as a source of news (Reuters Institute for the Study of Journalism, "Digital News Report ", http://bit.ly/SBryvD, last access: December 14, 2017). In the case of the United States, a study by the Pew Research Center shows that 67% of the population consumes news through these platforms. Living up to its name, Facebook's News Feed is the main source of news among respondents who used social media as a news source (Pew Research Center , "News Use Across Social Media Platforms 2017", http://pewrsr.ch/2vMCQWO, last access: December 14, 2017).

12 Office of the National Director of National Intelligence of the National Intelligence Council of the United States of America, "Assessing Russian Activities and Intentions in Recent US Elections", January 6, 2017, retrieved from: http://bit.ly/2iRbS9b.

13 Mason, Rowena, "Theresa May accuses Russia of interfering in elections and fake news", *The Guardian,* November 14, 2017. Retrieved from: http://bit.ly/2mnQaML.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

3

Note. Google searches of the term "fake news" during the last five years (Google Trends).

Several US authorities initiated investigations into Russian interference in the election, for which they have requested information from Internet companies. Initially, Facebook wanted to downplay the problem: two days after the election, Mark Zuckerberg publicly argued that thinking that fake news articles had an impact on the election was a "pretty crazy" idea.[14] But later the company admitted before the United States Senate that millions of users had seen advertising produced by Russia, published on Facebook and Instagram.[15] Meanwhile, Twitter reported that it had found 2,752 accounts which were being controlled by Russians and that Russian bots had tweeted 1.4 million times during the election. Google found on YouTube more than a thousand videos on the subject.

This scenario became the perfect storm for Internet companies, which for a great variety of issues and on different fronts face pressure from governments and civil society to make their practices transparent.[16] Misinformation is not an isolated problem nor is it new. In the background is the question of harmful content -defamatory, incendiary, which violates privacy, among others- and the companies' response to these. The moderation of content is, therefore, at the center of this discussion and its implications reach the exercise of online freedom of expression.

As it will be seen below, Facebook and Google are adopting measures to deal with the problem. Some of them focus on prevention, seeking to educate the citizen to make informed decisions about the content he or she consumes. Others have a direct effect on the information that is published on the platforms, either through invisible changes in the algorithm or visible warnings about the veracity of the information. Most of these measures are trials or have partial geographic coverage;[17] others are meant only for specific moments, such as elections. In any case, it is not easy to determine their scope and depth because they are essentially a series of announcements whose implementation is not entirely clear.

This document exposes the measures announced by Facebook and Google to combat misinformation. It also includes a brief allusion to YouTube and Twitter. The focus is set on those measures that have a direct effect on the platform and on the information received by its users, and not on preventive and educational measures that have been in development in parallel. Likewise, the document tries to determine the geographic coverage of these measures. Subsequently, as a conclusion, the possible problems behind the proposed solutions are divided in four points: i) scale and time, ii) impact, iii) the role of civil society, and iv) transparency.

---

14 Interview conducted by journalist David Kirkpatrick, founder of Techonomy Media, on November 10, 2016, retrieved from: http://bit.ly/2o6NMrv.

15 Shaban, Hamza, Timberg, Craig and Dwoskin, Elizabeth "Facebook, Google and Twitter testified on Capitol Hill. Here's what they said", *The Washington Post,* October 31, 2017, retrieved from: http://wapo.st/2gZM9fx.

16 In June 2017, the German parliament passed a law that requires Internet companies with over two million users to remove hate speech and other illegal content from their platforms, such as defamatory fake news in less than 24 hours, on pain of fines of up to 50 million euros (Miller, Joe, "Germany votes for 50m euro social media ends", BBC, June 30, 2017, retrieved from: http://bbc.in/2C0rT-na). More recently, a spokesperson for Theresa May reported that the UK is evaluating the role of Facebook and Google in the supply of news and what their responsibilities might be ("Britain looking at Google, Facebook role in news: PM May's Spokesman," Reuters, October 10, 2017, retrieved from: http://reut.rs/2swt75t).

17 The "Explore" Facebook experiment, developed in six countries, separates friend posts and advertising into one tab and public content from other Facebook accounts into another. Facebook has announced that it has no plans to make this measure definitive or extensive to everyone. Facebook, "Clarifying Recent Tests", http://bit.ly/2zwW0NF, last access: December 14, 2017.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

4

II. Misinformation and manipulation: a tentative classification

In November of 2017, Collins Dictionary chose "fake news" as the word of the year.[18] The expression, which was used 365% more this year, will be included in the next edition of that dictionary as "false information, frequently sensational, disseminated under the guise of news reporting". However, fake news does not have one single connotation. The general public does not only use it to refer to false reports, but in general to express a discontent with the misinformation, especially online.[19] Without having to look too far for an example, President Donald Trump uses it to disqualify any information he does not agree with.

The truth is that, beyond the political use of the term, fake news is also related to extremist opinions, propaganda and manipulation. Many might consider a politician's uninformed and alarmist opinion on Twitter to be fake news, to the same degree as a news story which reports falsely about the death of a world leader in deliberate bad faith. In both cases we find differences in content (an opinion versus a piece of news), format (a tweet versus a web page) and possibly motivations (the politician wants to rally his or her base while the website wants clicks).

Below there is a classification to help understand how content is produced and how it reaches the reader. This classification does not set mutually exclusive categories. For example, content can be fake news and have, at the same time, a propagandistic approach. It is not fully comprehensive either, it specially excludes contents of satirical journalism and errors of reporting committed in good faith. The latter could be part of a debate about fake news, but are not related to the purpose of this document.

1.Fake news

This refers to deliberately false content that is published on websites whose appearance tries to be formal and authentic.[20] Sometimes the design of the site and its URL take the place of a well-known news portal. The clear purpose is to deceive the user. Generally these contents proliferate in social networks through the own accounts of those portals, either in an organic way - with likes, retweets and shared by users- or with promoted actions, that is, paying for these contents to be advertised by the platforms.

---

[18] Flood, Alison, "Fake news is 'very real' word of the year for 2017", *The Guardian,* November 2, 2017, retrieved from: http://bit.ly/2iTWYk4.
[19] *"Our findings suggest that, from an audience perspective, fake news is only in part about fabricated news reports narrowly defined, and much more about a wider discontent with the information landscape— including news media and politicians as well as platform companies. Tackling false news narrowly speaking is important, but it will not address the broader issue that people feel much of the information they come across, especially online, consists of poor journalism, political propaganda, and misleading forms of advertising and sponsored content."* Nielsen, Rasmus and Graves, Lucas, "'News you don't believe': Audience perspectives on fake news", Reuters Institute for the Study of Journalism, October 2017, retrieved from: http://bit.ly/2o4Exb6.
20   Media Matters, "Understanding The Fake News Universe", http://bit.ly/2EDq8NH.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

5

Note. False story published in the abcnews.com.co site, created to mimic ABC News, whose web address is abcnews.go.com. According to this false news, some of the protesters against Donald Trump were paid to protest. Despite the falsehood of the news, days after his election, Trump himself suggested that these people were "professional demonstrators".[21]

Note. A picture of voting in Arizona was altered to include the image of an arrest. The Twitter user who posted it said that an undocumented immigrant had been arrested for trying to vote.[22]

Pieces of fake news in the strict sense can have economic or political motivations or a little of both. In the first case, they are commercial operations that seek to generate traffic from false contents and, above all, sensationalist headlines that people click on, but where the information does not make sense nor has any relevance. In the second case, they try to appear authentic not so much to generate traffic and profits but to manipulate the public debate in favor of certain political interests. An example of this category is the aforementioned false information on the support of Pope Francis to Donald Trump's candidacy in 2016.

[21] This news was shared on Twitter by Eric Trump, Donald Trump's son, by Corey Lewandowski and Kellyanne Conway, two of Trump's campaign leaders. Jacobson, Louis, "No, someone wasn't paid $3,500 to protest Donald Trump; it's fake news", *Politifact,* November 17, 2016, retrieved from: http://bit.ly/2flpuUR; Stahl, Lesley, "President-elect Trump speaks to a divided country", *CBS News,* November 13, 2016, retrieved from: http://cbsn.ws/2swFxKz.
[22] Wardle, Claire, "6 types of misinformation circulated this election season", *Columbia Journalism Review,* November 18, 2016, retrieved from: http://bit.ly/2CkRdjh.

Pope Francis Shocks World,
Endorses Donald Trump for
President, Releases Statement

TOPICS: Pope Francis Endorses Donald Trump

Note. Fake news about the Pope's support for Donald Trump.

Political interest in the piece of fake news does not necessarily exclude the economic interest. While some pieces of fake news are created with either one of these motivations, in many cases both orbits can converge. In the case of misinformation around the presidential campaign in the United States, young Macedonians may have been indifferent to who won the election; but not the Russian operatives who also influenced it. In the latter case, the economic benefits derived from web traffic and interaction added to the underlying political agenda.[23]

2. Propaganda

Jacques Ellul considers propaganda an elusive concept that develops around psychological and war action, re-education, brainwashing and public relations between humans. In that sense, the French sociologist thinks that propaganda is, above all, a technique to influence the actions of groups or individuals.[24]

Among other manifestations, propaganda may include false information or certain information presented with a deceptive approach.[25] For example, some facts are reported but others are omitted; information is out of context; content is manipulated; theories or opinions are presented as facts; highly disputable information is given as credible; information is denied in order to create confusion, or one statement is proclaimed as the only truth in opposition to the 'other' - the strategy of the nationalist movements–.[26]

Propaganda has been a part of politics and communications at least since the beginning of last century. Therefore, it is not a digital phenomenon. However, the reach of these contents online is especially significant: through the advertising tool -particularly in Facebook- propaganda is tailored to specific communities and groups based on tastes, political bias and friendship circles. In the case of the elections in the United States, this level of detail reached an alarming level:

Dark ads - as they are known due to the impossibility of knowing who sees them - allowed reaffirming political convictions, stirring up differences and, in general, polarizing the electorate. A white man from a republican state saw an adver-

---

[23] "These Macedonians on Facebook didn't care if Trump won or lost the White House. They only wanted pocket money to pay for things--a car, watches, better cell phones, more drinks at the bar". Subramian, Samanth, "*Welcome To Veles, Macedonia, Fake: News Factory To The World", Wired,* March, 2017, retrieved from: http://bit.ly/2o7BOxQ.

[24] Ellul, Jacques, *Propaganda: The Formation of Men's Attitudes,* New York, Vintage Books, 1965.

[25] "*Propaganda is false or misleading information or ideas addressed to a mass audience by parties who thereby gain advantage. Propaganda is created and disseminated systematically and does not invite critical analysis or response".* Huckin, Thomas, "Propaganda Defined", in: Henderson, Gae Lyn and   Braun, M.J (Eds), *Propaganda and Rhetoric in Democracy: History, Theory, Analysis*, Carbondale, Southern Illinois University Press, first ed, 2016, pp. 118-136.

[26] Powell, John A. "Us vs them: the sinister techniques of 'Othering' – and how to avoid them", *The Guardian,* November 8, 2017, retrieved from: http://bit.ly/2iGLAUX.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

7

tisement against immigration or in defense of the use of guns; an African-American saw an advertisement that recalled the racial persecution of the Ku Klux Klan; a Catholic saw Hillary as the incarnation of the Devil in a fight against Jesus.[27]

Note. A television ad for Donald Trump's campaign stated that the then-candidate would stop illegal immigration on the "southern border" by building a wall that would be paid by Mexico, while showing a video of dozens of people crossing the border between Morocco and the Spanish city of Melilla, not between Mexico and the United States.[28]

3. Conspiracy theories

Conspiracy theories seek to explain a particular event as the result of a plan carefully coordinated by an individual or a group. Motivations are generally secret and malicious, and actions are carried out to the detriment of the general interest.[29] These theories swarm in video channels and Internet pages, and are often presented as news despite their scant factual foundation.

In Colombia, there has been a theory for some years according to which President Juan Manuel Santos was recruited to secretly work for the Cuban government.[30] In Argentina, after the disappearance of Santiago Maldonado, multiple conspiracy theories erupted on the Internet, including one that stated that the website santiagomaldonado.com, created to demand the appearance of the activist alive, had been set up before his disappearance.[31]

The Pizzagate conspiracy theory, mentioned in the introduction of this paper, was examined by several media outlets in the United States with the purpose of identifying its origin.[32] In October 2016, a Twitter user posted a Facebook message in which a woman claimed that an anonymous source from the New York City Police Department had told her there was evidence that Hillary and Bill Clinton were involved in a child sex trafficking network. The tweet reached thousands of retweets and hours later an user of a discussion forum on the Internet posted a message saying that "internal sources" had confirmed the existence of the pedophilia ring, which would be exposed in a matter of hours.[33] The next day, the fake news site YourNewsWire.com posted a story based on the comments of a 4chan user - a tro-

---

[27] Cortés, Carlos, "El algoritmo imposible, redes sociales y noticias falsas", *Revista Arcadia,* December, 2017, retrieved from: http://bit.ly/2CmcdpT.

[28] Edds, Carolyn, "Donald Trump's first TV ad shows migrants 'at the southern border,' but they're actually in Morocco", *Politifact,* June 4, 2016, retrieved from: http://bit.ly/1mvNQi9.

[29] Media Matters, "Understanding The Fake News Universe", http://bit.ly/2EsVDdZ, last access: December 14, 2017. According to Michael Barkun, conspiracy theories are characterized by following three principles, namely: i) nothing happens by accident, ii) nothing is what it seems and iii) everything is connected. Barkun, Michael, *A Culture of Conspiracy: Apocalyptic Visions in Contemporary America*, Berkeley, University of California Press, 2003, pp. 3-4, retrieved from: http://bit.ly/2Btakek.

[30] "Los Santos y su militancia castrocomunista", *Periodismo Sin Fronteras,* June 1, 2013, retrieved from: http://bit.ly/1hjVlzx.

[31] "El sitio de Santiago Maldonado fue creado antes de su desaparición", *Data 24,* September 18, 2017, retrieved from: http://bit.ly/2w3l0ed.

[32] The following articles can be read on this: (i) Aisch, Gregor, Huang, Jon and Kang, Cecilia, "Dissecting the #PizzaGate Conspiracy Theories", *The New York Times,* December 10, 2016, retrieved from: http://nyti.ms/2jcCzlu. (ii) Silverman, Craig, "How The Bizarre Conspiracy Theory Behind 'Pizzagate' Was Spread", *Buzzfeed,* December 5, 2016, retrieved from: http://bzfd.it/2CjSQOl. (iii) LaCapria, Kim, "Chuck E. Sleaze", *Snopes,* November 21, 2016, retrieved from: http://bit.ly/2xX7xta.

[33] The discussion forum "Breaking: Its worse then classified emails. Political Pedophile Sex Ring exposed" retrieved from: http://bit.ly/2ErmerY, last access: December 14, 2017.
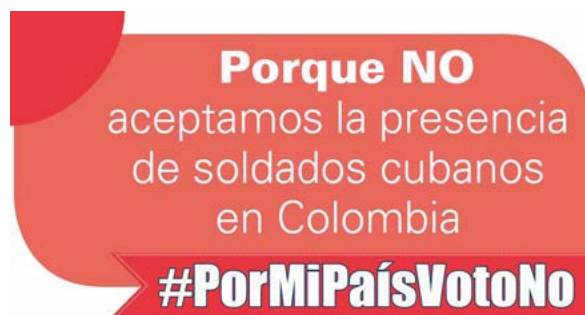
lling-heavy online board - where it was reported that an FBI source had confirmed the accusations.[34] The story was replicated and expanded by other fake news sites and shared on Facebook.

Among all the published versions, the story reached hundreds of thousands of interactions on Facebook and Twitter, and began to go viral under the #PizzaGate label.[35] Fake news articles were created with manipulated photos. Discussion forums and comments sections talked about a network of underground tunnels, torture chambers, Satanism and cannibalism in the basements of several restaurants. When the media refuted the theory they were accused by believers of wanting to hide the truth. Even weeks after the arrest of the man who shot the Comet Ping Pong pizzeria - where, of course, no illegal network was operating - some people still hinted that the story was true.[36]

4. False information, rumors, chains, memes

A broader category of misinformation, which includes different forms of fake news, propaganda and conspiracy theories, consists of the contents mixed and spread through instant messaging services, mainly WhatsApp. On this platform, information moves from hand to hand in the form of images, videos or memes, without an identified or identifiable author. In the end, the contact who shared it gives legitimacy and authority to the content.

In countries like Colombia this kind of content had a seemingly far reach last year. During the campaigns for and against the Havana peace agreements, prior to the October 2016 plebiscite, false, inaccurate and decontextualized content was disseminated through WhatsApp, aimed at capturing the vote against the agreements. Like one of the leaders of the "No" campaign admitted, the goal was to generate anger and outrage.[37]

Note. Messages with fake information in social networks and WhatsApp during the campaign prior to the October 2016 plebiscite.[38]

Whether as eminently false news, propaganda, conspiracy theories or rumors, fake news is not an isolated phenomenon of social and political reality, and in no way is it an externality of technology. "Today, when we talk about people's relationship with the Internet, we tend to adopt the uncritical language of computational science. 'Fake news' is described as a 'virus' among users 'exposed' to online misinformation" explains North American sociologist Katherine

---

[34] "FBI Insider: Clinton Emails Linked To Political Pedophile Sex Ring", *YourNewsWire.com*, October 31, 2016, retrieved from: http://bit.ly/2fesMdl.

[35] Silverman, Craig, "How The Bizarre Conspiracy Theory Behind 'Pizzagate' Was Spread", BuzzFeed, December 5, 2016, retrieved from: http://bzfd.it/2CjSQOl.

[36] Allam, Hannah, "Conspiracy peddlers continue pushing debunked 'pizzagate' tale", *Miami Herald,* December 5, 2016, retrieved from: http://hrld.us/2CFYGKw.

[37] Ramírez, Juliana, "El No ha sido la campaña más barata y más efectiva de la historia", *Asuntos Legales,* October 4, 2016, retrieved from: http://bit.ly/2EHxlwc.

[38] "¿Qué tan ciertos son algunos memes de la campaña del "NO" en el plebiscito?", *Pacifista,* August 30, 2016, retrieved from: http://bit.ly/2bP16MN.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

9

Cross.[39] Misinformation originates and feeds back in and from human actions and it is in this relationship that we can locate the dimension of the problem and the limitations of the proposed solutions.

III. Affordances and spaces

Just as the phenomenon of misinformation responds to different social contexts, it does not manifest itself in the same way in all digital media: a space such as WhatsApp enables the exchange of anonymous content -without author or apparent source-; Facebook gives greater relevance to content that is massively shared and on Twitter the user chooses the voices he or she wants to hear. In this dialog between the platform and the user the production and exchange of information acquires particular connotations. Fake news do not take form in the same way in all these spaces.

The affordances are the properties that arise from the relationship between an object and a person. The concept of affordances, introduced by the psychologist James Gibson, refers to the possibilities of action in a given context, and is used to talk about how users interact with objects, environments or technologies to achieve certain results.[40] Identifying the offerings of an object allows the understanding of the different ways in which the object can be used for various purposes. Some of those offerings are inscribed in the design of the object; others are "discovered" by the individual. A broom is used to sweep, but also to hit a distant object, like an orange in a tree; a table is used to put objects or sit on it; a courier service is used to spread your own messages or those from strangers. French intellectual Bruno Latour defines design itself as a process of inscription of modes of use: "The resulting products carry with them the 'scripts' that inhibit or preclude certain actions while inviting or demanding others."[41]

When those objects, environments or technologies allow social actions, we are faced with a social affordance[42] In the context of social technologies, different studies have focused, for example, on how social networks are used by people to organize their private lives; by governments to have direct contact with their citizens or to monitor them; by companies to stimulate teamwork; by educational institutions to promote pedagogical purposes, or by political organizations to motivate citizen participation.[43] The affordances of the service are also influenced by the use policies and algorithms, which allow these interactions and access to content or users (through the recommendations of the News Feed, for example). This adoption of the product by the user and the communities helps identify how misinformation is designed, dissemi-

---

[39] Cros, Katherine, "The Art of the Real: Disinformation vs. democracy", *The Baffler,* June, 017, retrieved from: http://bit.ly/2Er1KLR.

[40] Gibson, James, *The Ecological Approach to Visual Perception,* Hillsdale, Cornell University, 1986. Evans, Pearce and others define offerings as the multifaceted relational structure between an object or technology and the user, which allows or restricts potential behavioral outcomes in a particular context. To understand the concept they propose the following example: the built-in camera of a smartphone is a function of the phone; an offering is the phone's ability to record (e.g., images or videos of a person) and a possible result is the documentation of a human rights violation with that camera. They explain that functions or tools are static, while offerings are dynamic. Evans, Sandra K. and others , "Explicating Affordances: A Conceptual Framework for Understanding Affordances in Communication Research", in: *Journal of Computer-Mediated Communication,* Vol. 22, No. 1, Blackwell Publishing Ltd, 2017, retrieved from: http://bit.ly/2EEfYfM.

[41] Dow Schüll, Natasha, *Addiction by Design: Machine Gambling in Las Vegas,* Princeton, Princeton University Press, 2012 (Ed. Kindle). Informal translation.

[42] "Along the same lines as Gibson's, our hypothesis is that the richest and most elaborate environmental affordances are provided by other animals and other people". Kaufmann, Laurence and Clément, Fabrice, "How Culture Comes to Mind: From Social Affordances to Cultural Analogies", *Intellectiva,* No. 46-47, 2007, retrieved from: http://bit.ly/2sv5lqD.

[43] See: Kursat Ozenc, Fatih and Farnham, Shelly "Life "modes" in social media": *CHI '11 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Vancouver, May 7 to 12, 2012; Mergel, Inés, "Implementing Social Media in the Public Sector", October, 2013, retrieved from: http://bit.ly/2o6K0hT; Zeiller, Michael and Schauer, Bettina "Adoption, motivation and success factors of social media for team collaboration in SMEs",: *i-KNOW '11 Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies,* Graz, September 7 to 9, 2011; Wang, Qiyun, Lit Woo, Huay y Lang Quek, Choon, "Exploring the Affordances of Facebook for Teaching and Learning",: *International Review of Contemporary Learning Research*, No. 23-31, 2012, retrieved from: http://bit.ly/2Gd7ol3; Blegind Jensen, Tina and Dyrby, Signe "Exploring Affordances Of Facebook As A Social Media Platform In Political Campaigning",: *Proceedings of the 21st European Conference on Information Systems,* ECIS 2013 Completed Research, 2013, retrieved from: http://bit.ly/2ElGgxm.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

10

nated and consumed. In many cases, it is simply the common use of the service: a piece of fake news is an informative unit like any other. In other cases, it is an unintended consequence that the product offers: using a deceptive headline to generate clicks and viralize a lie. One way or another, these are uses that are not alien to the product.

Take the aforementioned Pizzagate example. The genesis of the story was in social networks: a rumor shared on Facebook ("an anonymous source of the NYPD said ...") and replicated on Twitter. This false content arises organically, that is to say, through the normal information-sharing mechanisms of the service. The configuration of both platforms, which rewards interaction in various ways, allows consuming this content massively. Its veracity is irrelevant. At the same time, a website that produces fake news presents that rumor as an article and, in addition to publishing it organically, promotes it. That is to say, it pays a platform like Facebook to display that information as advertisement to get more interaction and visibility. At this point, the product offers a tool to make that advertisement reach a defined audience. It is logical to assume that the inscribed use -the desired use - of the service is to promote truthful products, but its design offers - affords - the possibility of promoting fraudulent information.

Let's look at Twitter. The design of the network allows and promotes the creation of open and decentralized conversations. This offer has allowed planning social protests, rights movements and countless joint actions. But the incorporation of the product allows these joint actions to be concerted, which opens the possibility for a group of apparently authentic accounts to devote themselves to promoting false content.[44] And in this open and participatory environment, many users consider that this volume of exchanges is authentic and that, therefore, the information conveyed is accurate.

Let's finish with WhatsApp. As a messaging service, it is designed to preserve the privacy of communications. This closed configuration makes it impossible, in principle, to monitor the contents, either by the platform itself or by third parties (hence it is known as a "dark social" environment).[45] Furthermore, the only people who can share this content are the contacts that are part of a conversation -individual or collective-.[46] As individuals are the ones sharing videos, images, audios or texts, the information does not necessarily include a source, but it is still shared in a context of intimacy and with a known interlocutor, all of which gives it legitimacy. The object has the designed use of connecting known people in a closed environment, which does not mean that this channel is not used to share any type of information.

Being aware of these differences is important for two reasons. On the one hand, it allows the understanding of the proposed solutions and identifying their inherent limitations. On the other, it places the problem of fake news in the orbit of the social incorporation of a technology. The use of technology - and therefore, the questions that arise from using it - is a process mediated by people and not an isolated equation: "Technologies are not inserted in everyday life, causing a revolution or a radical break as people say; on the contrary, this insertion usually entails a gradual evolution, a negotiation between the inherited practices and the desire for the transformation of societies".[47]

IV. The solution to the problem

Long before a debate arose around misinformation, social networks already faced the general challenge of moderating content online. Arbitrating the flow of information is perhaps the greatest sign of the power of these platforms

---

[44] A study carried out by the University of Edinburgh identified hundreds of false accounts operated from Russia with the aim of influencing the referendum on the permanence of the United Kingdom in the European Union (Brexit). Booth, Robert and others, "Russia used hundreds of fake accounts to tweet about Brexit, data shows", The Guardian, November 14, 2017, retrieved from: http://bit.ly/2iUPF8b.

[45] Communications surveillance schemes are not ruled out, but the central point is that the product is not designed to generate and measure interaction and consumption of content such as Facebook or Twitter.

[46] If an unknown person gets one's phone and shares something on WhatsApp, it will possibly have an alienating effect. The recipient will reject the message or will not give relevance to the information.

[47] Gómez, Rocío and others (comp.), *Facebook como obra mundana. Poetizar la vida y recrear vínculos personales*, Universidad del Valle, Editorial Program, 2016, pp. 66.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

11

in their condition of intermediaries:[48] They do not produce the content, but they make important decisions about that content: what is going to be distributed and to whom, how are users going to be connected and how are their interactions going to be managed, and what is going to be rejected".[49] By moderating content, these actors try to apply their community rules to build user loyalty with the service and maintain an operation free of undesired interference.

Hate speech and terrorism, discrimination against minorities, harassment against women and toxic content, in general, forced these companies to seek a complicated balance between the free circulation of content and timely restriction. In this practice, they increasingly face questions about the transparency and accountability of these processes. The role of algorithms in content decisions, the suspension of accounts and the appeal mechanism, among others, are part of an agenda of demands that both governments and civil society ask from these companies.[50]

This is the context in which companies like Facebook, Google and Twitter try to respond to the problem of fake news. And although it would be a topic to develop in another opportunity, it is relevant to locate those answers within the content moderation policies already in use by these platforms -and not as a separate issue-. For example: while Facebook has a policy of real names, Twitter does not prohibit pseudonyms or parody accounts. This starting point delineates different developments regarding the moderation of misinformation.

1. Facebook

In mid-2017 the English newspaper The Guardian, through the project "The Facebook Files", offered a panorama of the company's moderation of content practices.[51] Instances of fake news in particular are part of what they call "information operations", which Facebook interprets as "the actions taken by actors organized to distort the national or foreign political sentiment."[52] Information operations are divided into fake news, misinformation and false accounts. Note how the company recognizes the joint and organized nature of these actions, understanding that it is an adapted and particular use of the product.

With that understanding, Facebook has announced solutions to face fake news with a focus on two areas. On the one hand, it has announced measures to promote literacy in news (known as news literacy) in its users, to help them make informed decisions about news and sources that can be trusted. To this end, it has developed two projects: the Facebook Journalism Project and New Integrity Initiative. [53] These preventive and educational measures are not addressed in this document. On the other hand, it announced the adoption of technical measures that have an impact on the contents received by users on the platform. As published by the company, these are the characteristics of said measures:

---

[48] *See* Cortés, Carlos "Las llaves del ama de llaves: la estrategia de los intermediarios en Internet y el impacto en el entorno digital", in: Bertoni, Eduardo (comp.), *Internet y Derechos Humanos. Aportes para la discusión en América Latina.* Buenos Aires, CELE, Universidad de Palermo, 2014.

[49] Gillespie, T. Cited in: Myers West, S. 'Raging Against the Machine: Network Gatekeeping and Collective Action on Social Media Platforms', Media and Communication (ISSN: 2183–2439) 2017, Volume 5, Issue 3, p. 28 (informal translation).

[50] See, among other initiatives, www.onlinecensorship.org and http://responsible-tech.org/. The Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression of the United Nations, for example, will make a report in June 2018 on the regulation of content in the digital age.

[51] See, "The Facebook Files", *The Guardian*, retrieved from: http://bit.ly/2r6h5yb.

[52] Facebook, "Information Operations and Facebook", April 27, 2017, retrieved from: http://bit.ly/2oOOS9s.

[53] Through this initiative, Facebook offers training tools for journalists. "Facebook Journalism Project", http://bit.ly/2ioDPAO, last access: December 14, 2017. The News Integrity Initiative is Facebook's big bet on the subject of news literacy. This project led to the creation of a consortium of technology industry leaders, academic institutions and other organizations, which will aim to help people make informed decisions about the news they read and share on the Internet. "News Integrity Initiative", http://bit.ly/2D5udGi, last access: December 14, 2017.

1.1. Report, fact-checking and flagging

In December 2016, Facebook announced that it was testing different mechanisms to facilitate its users the reporting of possible fake news.[54] Community reports and "other signs" -not specified by Facebook- would be considered to select stories that they would send to independent organizations to do fact-checking. Facebook does not detail how many reports are necessary to generate an alert.

If the organizations dispute the content after the verification process, Facebook will display warnings (flagging) indicating that the article has been disputed. A link to get more detailed information will be included next to the flag. If despite these messages a user wants to share the content, Facebook can show a new message in which it warns the user that they will share content that has been disputed. Moreover, according to Facebook, these stories may have less prevalence in the News Feed, and may not be promoted or converted into advertisements.

Mark Zuckerberg has said that Facebook does not want to be an arbitrator of the truth.[55] The task of verifying the facts was entrusted to external organizations that are signatories of Poynter's International Fact-Checking Network's code of principles.[56] Initially, the company is working with ABC News, FactCheck.org, Associated Press, Snopes and Politifact.[57] Between the end of 2016 and 2017, these measures have been announced as trials in countries such as the United States, Germany, France and the Netherlands, and are not available permanently.[58]

When making this announcement, Facebook published some images of the reporting scheme. A test made from a connection with IP in the United States confirmed the mechanism. However, the same exercise carried out in August of 2017 from an IP in Colombia showed that although users could point to the option "It is fake news", it was not possible to send a report to Facebook. The only options available were to block, stop following, delete or send a message to whomever had shared the disputed information. A new test carried out in December 2017 shows that the reporting option is disabled in both countries, at least with our Facebook user.

---

[54] "Enfrentando engaños y noticias falsas", http://bit.ly/2HgPW0k, last access: December 14, 2017.

[55] http://bit.ly/2fOsvha

[56] International Fact-Checking Network, "International Fact-Checking Network fact-checkers' code of principles", retrieved from: http://bit.ly/2BwtakU.

[57] Mullin, Benjamin, "Facebook rolls out plan to fight fake news", *Poynter,* December 15, 2016, retrieved from: http://bit.ly/2h5bPm5.

[58] See: Facebook, "News Feed FYI: Addressing Hoaxes and Fake News", December 15, 2016, retrieved from: http://bit.ly/2gFFvVw; Facebook, "Umgang mit Falschmeldungen", January 15, 2017, retrieved from: http://bit.ly/2gFveyn; Facebook, "Facebook annonce le lancement de son outil de fact-checking en France", February 6, 2017, retrieved from: http://bit.ly/2swl4Ws; Facebook, "Addressing Fake News Globally", January 15, 2017, retrieved from: http://bit.ly/2jyZ1FV.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

13

Note: screenshots of the fake news reporting scheme.

1.2. Changes in the news feed and measures to counter the false amplification.

As stated above, the content that the verifiers mark as false may have less relevance in the News Feed. Additionally, there are other signs that can lead to Facebook giving less relevance to these contents: for example, if reading an article people are less inclined to share it, it can be a sign that the story is misleading.[59] Or when users regularly post many publications per day with external content that can be considered of low quality -deceptive or sensationalist- the publication with an external link can be "punished".[60] On the other hand, additional measures have been announced in the news feed:

- According to an announcement made in April 2017, Facebook is trying to make it so that articles related to a publication appear before the user reads the desired content, so that they have quick access to other perspectives on the same subject, including articles that have passed the fact checking filter.[61]

Note: Related articles in Facebook

---

59 Facebook, "Enfrentando engaños y noticias falsas", December 15, 2016, retrieved from: http://bit.ly/2HgPW0k.
60 Facebook, "News Feed FYI: Showing More Informative Links in News Feed", June 30, 2017, retrieved from: http://bit.ly/2uzG8HM.
61 Facebook, "News Feed FYI: New Test With Related Articles", April 25, 2017, retrieved from: http://bit.ly/2q2zxHe.

- In an August 2017 update, Facebook announced that it will start using automated learning (machine learning) to detect more deceptive content. Facebook will be able to show the fact-checking stories below the original articles.

- Since July 2017, Facebook has eliminated the option to customize the image, title or preview description of the links that are published in Facebook profiles or groups. These previews will depend solely on the metadata of the website that is linked.[62] This prevents modifying this information to attract clicks through headlines or misleading images.

- Since 2013, Facebook has announced the adoption of a new algorithm to detect and give more relevance to "high quality" content. At the time, for example, whether the content was relevant or if the sources were reliable were both taken into consideration.[63] On this matter, in 2016 Facebook established a policy to prevent advertisers with low quality content pages from advertising on the platform.[64]

Facebook has also announced measures to counter "false amplification", which it defines as the coordinated activity that seeks to manipulate the political discussion.[65] The creation of false accounts (often executed on a large scale) is part of this irregular practice; the coordinated distribution of content or repeated messages; the likes or coordinated reactions; the creation of groups with the purpose of distributing sensationalist news or headlines, and the creation of memes, videos or manipulated photos; among others.

To this end, the company announced the development of technological tools that allow it to identify false likes and false comments that come from false accounts, malware or so-called "click farms" (groups of accounts that have an apparently organic activity that are created to generate a false interaction).[66]

2. Google

According to the New York Times, last October fake news ads used Google's advertising system and even appeared in fact-checking portals such as Snopes and Politifact. A fake Vogue website announced, for example, that the first lady of the United States, Melania Trump, had refused to live in the White House.[67] However, this was not Google's only front of misinformation. Fake news portals have appeared in its search engine results, and videos of conspiracy theories are frequently recommended in YouTube.[68]

Some measures taken by the company to deal with misinformation are focused on digital literacy about news consumption, such as the Internet Citizens program on YouTube, which gives workshops to young people between the ages of 13 and 18. On the other hand, Google finances hundreds of projects in Europe to produce original journalism and to guide citizens on reliable content.[69] Beyond this, the response of the service has focused on two fronts: search services (Google Search and Google News) and advertising services (Google AdWords, Google AdSense and YouTube).

---

62 Facebook, "API Change Log: Modifying Link Previews", retrieved from: http://bit.ly/2u1ndq6.
63 Facebook, "News Feed FYI: Showing More High Quality Content", retrieved from: http://bit.ly/2BZYU2U.
64 See: "Contenido molesto o de baja calidad", retrieved from: http://bit.ly/2q9hkFk, last access: December 14, 2017.
65 Facebook, "Information Operations and Facebook", April 27, 2017, retrieved from: http://bit.ly/2oOOS9s.
66 See: Facebook, "Breaking new Ground in the Fight Against Fake Likes", April 17, 2015, retrieved from: http://bit.ly/1yCsKDM; Facebook, "Disrupting a major spam operation", April 14, 2017, retrieved from: http://bit.ly/2oQJZQX.
67 Wakabayashi, Daisuke and Qiu, Linda, "Google Serves Fake News Ads in an Unlikely Place: Fact-Checking Sites", *The New York Times,* October 17, 2017, retrieved from: http://nyti.ms/2hO00Ca.
68 Abrams, Dan, "Now Even Google Search Aiding in Scourge of Fake, Inaccurate News About Election 2016", *Mediaite,* November 13, 2016, retrieved from: http://bit.ly/2C0uaic.
69 In this regard, see the Cross Check project and the Digital News Initiative.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

15

1.1. Changes in search services: Google Search and Google News

Google has announced several changes that could affect the experience of users of the search service: fact checking for search results, changes in the search algorithm and user feedback. Moreover, Google has announced that it will make the way searches work more transparent.

- Fact checking

Since 2009, Google has implemented a labeling system to mark some of the results shown in Google News searches.[70] For example, the content opinion, Blog or Satire labels help the users to identify the type of content they will encounter when opening an article. In October 2016, Google announced that it would implement the Fact checking label, with which it intends to label articles that have gone through a fact-checking process.

For example, for months President Trump has said that his government will make the largest tax cut in history. A search of this topic in Google News (November 2017) shows among the results an article of FactCheck.org marked with the label Fact-Check, where it is explained to which extent is that statement truthful and plausible.

Note. The FactCheck.org story appears under the subtitle "Related Coverage". Taken from Google News.

Originally, this function was available in the United States and the United Kingdom. However, on April 2017, Google announced that the label would be available worldwide and that it would be extended to its general Google Search system in all languages.[71] Thus, when a user does a Google search and the results show contents that have been verified, Google will display this information indicating what was said, who said it, who checked the information and the result of that checking.[72] The result of the checking can be not only true or false, but also mostly true or partially true. To index this information, Google starts with criteria such as the reliability of the source, checked facts, sources and quotes, and the conclusions reached through this review.[73]

Publishers who want their fact checks to show up in search results can point them to Google in two ways: i) using the ClaimReview label[74] in the publication code, based on Google's guidelines for data verifiers; ii) using the Share The Facts widget, which acts as a quality seal for participating organizations and can be embedded - as if it were a tweet - on a web page.[75] We see then how Google limits itself to highlighting the checks made by third parties based on certain

---

[70] Google, "¿Qué significa cada etiqueta (p. ej., "blog")?", http://bit.ly/2F2tiYW, last access: December 14, 2017.

[71] Google, "Fact Check now available in Google Search and News around the world", April 7, 2017, retrieved from: http://bit.ly/2tZyRSf.
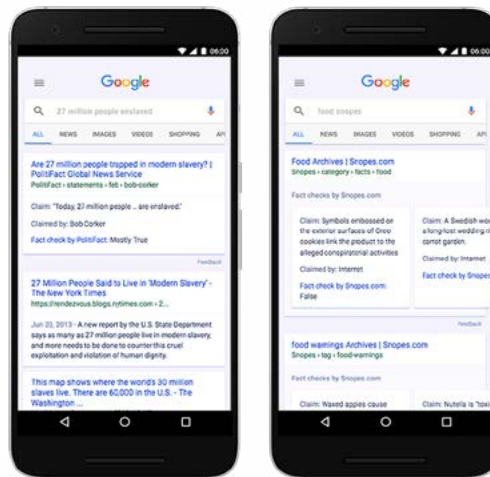
[72] Different verifiers can reach different conclusions. Several conclusions can be presented in the results.

[73] See data verifications in the search results, http://bit.ly/2CjpcJ5, last access: December 14, 2017.

[74] See: Schema.org, "ClaimReview", retrieved from: http://bit.ly/2EpwIDP, last access: December 14 2017; Google, "Verificaciones de datos", retrieved from: http://bit.ly/2CmGPrd, last access: December 14, 2017.

[75] Share the Facts, "Share the Facts", retrieved from: http://bit.ly/2srvQxi, last access: December 14, 2017. To date, it works together with twelve organizations, none of them in Latin America: PolitiFact, The Washington Post, FactCheck.org, Gossip Cop, Pagella Politica, AGI, La Stampa, The Ferret, Climate Feedback, Demagog, Newsweek and VERA Files Fact Check (http://www.sharethefacts.org/about/).

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

16

criteria, even if they have reached different conclusions.[76] In contrast, Facebook incorporates the checking of some organizations in its own conclusions, either to decide the prominence of the story or to make a particular warning.
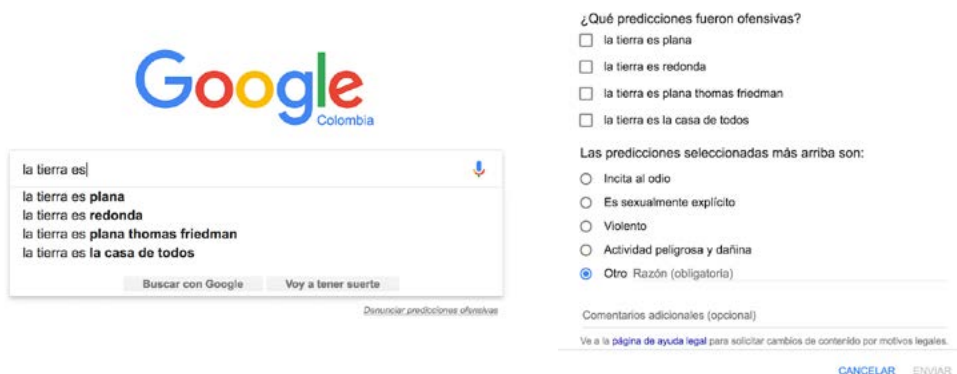

Note. Examples of how checked information appears in Google searches.

- Improvements in the search ranking

Although Google states that only 0.25% of searches show offensive or deceptive content, in April 2017 it announced changes in Google Search so that the results reflect more reliable content.[77] This includes updating the guidelines used by teams that evaluate the quality of search results.[78]

- Feedback from users

In April 2017, Google announced that it would improve the users' reporting mechanisms in relation to the terms suggested in the autocomplete function and the contents highlighted in the search results (featured snippets). According to Google, new mechanisms for obtaining feedback include pre-established categories of offensive content that facilitate reporting. These reporting mechanisms not only work to signal fake news, but also other problematic content (sexually explicit, violent, dangerous, etc.).[79]

Note. Users can send comments on the autocomplete function by clicking on the option 'Report offensive query', which displays several reporting options.

---

[76] Google, "Fact Check now available in Google Search and News around the world", April 7, 2017, retrieved from: http://bit.ly/2tZyRSf.

[77] Google, "Our latest quality improvements for Search", April 25, 2017, retrieved from: http://bit.ly/2sya4ro.

[78] Google, "General Guidelines", July 27, 2017, retrieved from: http://bit.ly/1ZdLFyy.

[79] Google, "Our latest quality improvements for Search", April 25, 2017, retrieved from: http://bit.ly/2sya4ro. Before the announcement, it was possible for a user to report the terms suggested by the autocomplete function if they violated Google's AutoComplete Policy, however, it was necessary to go to the help page, find and fill out a form (http://bit.ly/2C0rj95).

Note. Reports on outstanding results in the form of snippets can be made by clicking on the 'Comments' option and then pointing to 'The content is deceptive or incorrect'.

According to Google, this feedback is used to evaluate if the changes introduced in its search system are successful and if they should be applied to other users. Moreover, the company gave more information on how Google Search works and disclosed a few content policies for its autocomplete service.[80]

1.2. Changes in advertising services: Google Adwords, Google Adsense and YouTube

In March of 2017, an investigation by The Times revealed that YouTube ads of L'Oréal, The Guardian, Nissan, the BBC, among many others, appeared while playing videos with extremist, anti-Semitic or homophobic messages.[81] According to the research, those who post videos on YouTube generally receive up to $ 7.6 per thousand reproductions, and some videos had millions of views. After the disclosure, major advertisers such as PepsiCo, Starbucks, Walmart, AT&T, Verizon, Volkswagen, Johnson and Johnson, decided to withdraw their ads, in an event that was known in the media as the YouTube Ad Boycott.[82]

In response, Google published a statement acknowledging that, despite all efforts, its technological tools did not always detect bad ads or advertisers that violate its policies. A few days later, the company announced, among several measures, the strengthening of its policies on hate or offensive content, the creation of mechanisms for advertisers to control where their advertising appears, and the development of new tools to check questionable content.[83]

The problem of ads with undesirable content is the same as with fake news: advertisers do not want their products associated with misinformation. Before the controversy broke out, Google was already working to identify "bad ads": for example, products that falsely promise to help lose weight, illegal or counterfeit products or ads that carry malicious code or viruses.[84] Along the same lines, on November 2016, Google introduced changes to its Google AdSense

[80] Google, "Políticas de autocompletar", http://bit.ly/2EqUeRb, last access: December 14, 2017; Cómo ayudar a mejorar los resultados de búsqueda de Google", http://bit.ly/2o9oqci, last access: December 14, 2017; Google, "How Search works", http://bit.ly/2rudrP8, last access: December 14, 2017.

[81] Mostrous, Alexi, "YouTube hate preachers share screens with household names", *The Times,* March 17, 2017, retrieved from: http://bit.ly/2HfmUON.

[82] Solon, Olivia, "Google's bad week: YouTube loses millions as advertising row reaches US", *The Guardian,* March 25, 2017, retrieved from: http://bit.ly/2n41Ccw.

[83] Google, "Improving our brand safety controls", March 17, 2017, retrieved from: http://bit.ly/2uPIDdc; Google, "Expanded safeguards for advertisers", March 21, 2017, retrieved from: http://bit.ly/2EJTCJT; YouTube, "YouTube Partner Program overview", http://bit.ly/1bQpskt, last access: December 14, 2017.

[84] Google, "How we fought bad ads, sites and scammers in 2016", June 25, 2017, retrieved from: http://bit.ly/2jpTTWP; Google, "How we fought bad ads, sites and scammers in 2016", June 25, 2017, retrieved from: http://bit.ly/2jpTTWP.

policy - its ad network that appears on different Internet sites - and prohibited the publication of ads on websites that distort or conceal information about the publisher, its contents or the main purpose of the site.[85] This policy, however, does not refer to fake news but to false representation.[86]

In the latest report on the implementation of its policies to combat prohibited advertisements and sites, Google reported that after introducing said changes, actions were taken against 340 sites for false representation and other offenses.[87] Two hundred of these sites were definitively expelled from the company's advertising network. However, Google did not clarify how many of them were cases of fake news. The closest thing to the issue appears in an excerpt in the report where it reveals the blocking of more than 1,300 accounts of tabloid cloakers, portals that publish ads claiming to be news headlines but directing them to sites with advertising.

3. Twitter

Despite having 16% of the users Facebook has, Twitter has also become part of the controversy surrounding misinformation. According to the company's own statements at the hearings with the US Congress, more than 2,700 accounts associated with Russian agents moved 1.4 million tweets between September and November 2016.[88]

This is not a completely new problem for Twitter, which is constantly questioned due to false accounts and bots in their platform. According to the company, this type of accounts does not represent more than 5% of its users, but external sources assure that there are many more. For example, according to a study conducted by Alessandro Bessi and Emilio Ferrara, bots were responsible for a fifth of the tweets related to the US presidential elections.[89] Nonetheless, it is important to remember that Twitter allows automated accounts, and many of them fulfill a relevant service in terms of information (news channels, government services) and are not considered spam.

Twitter has acknowledged that it is impossible to determine the veracity of the published tweets and, like Facebook, has maintained that it does not want to be an arbiter of the truth.[90] If Twitter starts to evaluate information it would not only be undesirable, but impossible to implement in practice: at least one billion tweets per day pass through the platform. Thus, the focus of the company's response is to detect and remove accounts that, in an automated or manual manner, disseminate malicious content (spam, falsehoods or attacks, among others). These are some of the concrete actions announced by the company:

Reducing the visibility of tweets and possible spam accounts while investigating whether a violation actually occurred.

- Suspension of accounts once prohibited activity has been detected.

- Implementing measures to detect applications that abuse Twitter's public API[91].

- Twitter considers that any automatic detection system carries a high risk of false positives; that is, accounts that the algorithm considers may be violating the policies but in reality have a legitimate behavior. For example, an account of an activist who is tweeting constantly can be confused with a bot or with a person who is deliberately sending spam. On the other hand, misinformation in the platform is carried out through joint actions between groups. These are coordinated operations that comply with the rules and restrictions of the platform to achieve

---

[85] Google, "Prohibited content", http://bit.ly/2nZvWYe, last access: December 14, 2017.

[86] In fact, in January 2017, the Media Matters organization for America reported that Google removed this expression which was included before as an example of forbidden content. "Google Quietly Removes "Fake News" Language From Its Advertising Policy", Media Matters, January 12, 2017, retrieved from: http://bit.ly/2BZZY6U.

[87] Google, "How we fought bad ads, sites and scammers in 2016", January 25, 2017, retrieved from: http://bit.ly/2kclNnZ.

[88] Solon, Olivia and Siddiqui, Sabrina. "Russia-backed Facebook posts 'reached 126m Americans' during US election", October 31, 2017, retrieved from: http://bit.ly/2hoiMRe.

[89] Bessi, Alessandro and Ferrara, Emilio. "Social bots distort the 2016 US presidential election online discussion", November 7, 2016, retrieved from: http://bit.ly/2oZNst8.

[90] Twitter, "Our Approach to Bots & Misinformation", June 14, 2017, retrieved from: http://bit.ly/2HhmFCC.

[91] API is an application programming interface that allows an external software to use several functionalities of a service, in this case, Twitter

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

their purpose. "It is much more complicated to detect non-automated coordination", explains the company. "The risks of inadvertently silencing legitimate activity are much higher".[92]

V. Conclusion: the problem of the solution

The main objective of this document was to expose the measures that some intermediaries -mainly Facebook and Google- have set up to combat misinformation. After explaining the problem and placing it in the territory of each platform, we describe the proposed solutions, most of which are in the preliminary phase. In this final part of the text we detail some of the problems that these solutions entail and we make some recommendations. To carry out this analysis, we propose four points: i) the scale and time of the solution; ii) the impact; iii) the role of civil society, and iv) transparency. This analysis starts from a basic assumption that we described at the beginning and we reiterate now: it is neither possible nor desirable that the solution to fake news be automatic. It is a phenomenon with technological incorporation, which is social by definition.

This analysis focuses on organic content and not on the promoted one. The pieces of fake news that are promoted commercially present some challenges and different demands for the platforms. It would be reasonable to demand a higher level of monitoring and control in that case, since they receive a direct profit for advertising information for commercial purposes. Organic content is the one which users disseminate among their contact networks, and in contrast to commercial advertisements it includes spontaneous user information, manipulation and coordinated actions in the same degree. This is where we find the greatest challenges for freedom of expression and, therefore, where we want to direct our attention.

1. Scale and time

To analyze these two factors, it is necessary to return to the concept of affordances. As explained above, the response to misinformation must necessarily start with the configuration of the space where the problem occurs. This demonstrates a structural issue: some elements that enable misinformation could only be eliminated in the design of the service itself. These are limitations inherent to the architecture of the space. For example, if Facebook introduced a system of prior review of all content: if it were possible, it would eliminate in equal parts misinformation and a lot of legitimate content. But as Facebook is not going to modify that substantial aspect of its structure and space, the "offering" of this platform limits the scope of the solution.

Facebook and Google, YouTube and Twitter, "offer" a space where content can be disclosed without prior review. This means that, in the vast majority of cases, the strategies to deal with misinformation will be some form of subsequent control: labels, related articles, less visibility, etc. That type of control, as we saw, is not susceptible to total automation, and to that extent there is a challenge in terms of scale and time.

The scale refers both to the company's teams and to the work with external actors. Can the company replicate the initiatives against misinformation for all its users (more than two billion in the case of Facebook)? Does it have enough internal human teams to review all the questionable content? Are there enough fact-checkers for all the available information? The answer will invariably be negative. If the strategy to deal with misinformation requires a human effort - which is also desirable- it will not be possible to meet all the demand of the problem.

Last September, Facebook carried out several preventive actions to avoid misinformation during the German elections: suspension of dozens of accounts, direct work with local authorities and educational campaigns.[93] With the

---

92 Op. Cit., Twitter, "Our Approach to Bots & Misinformation". Informal translation.

93 Facebook, "Update on German Elections". September 27, 2017, retrieved from: http://bit.ly/2Buursp. See also: Locklear, Mallory. "Facebook axed 'tens of thousands' of accounts before German election". Engadget, September 27, 2017 retrieved from: http://engt.co/2k1Etdw.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

20

antecedent of the North American elections, and this time in the midst of the tense European regulatory context, the company could not sit idly by.

This type of efforts could hardly be deployed in markets with less relevance, as are the vast majority of Latin American countries. The scale of the solution, then, is mainly focused on key countries, and although some of the answers can be extended to other markets, they are thought and located for those priority scenarios (language, problem approach, solutions, and context). In the case of Colombia, as we have seen, the tool to report fake news has apparently not been available at any time. The fact-checking schemes, on the other hand, could be done in countries of the region (Share the Facts, we repeat, does not include Latin American organizations at this time), but they require local joint efforts that will not be easily applied to all of Latin America.

The limitation in scale is related to the time variable. The risk of misinformation in the public debate is even higher during election time. A false news or a rumor spread as truth can affect the outcome of an election. To that extent, a timely response to address this problem is desirable. A post-mortem action is relevant to understand the phenomenon, but not to face its immediate effects.

At key moments in a campaign, the responses described in this document may be too late. While building a lie is quick, verifying a fact requires time. It is essential to be aware of this limitation, not so much as an argument to justify more restrictive measures for public debate, but to understand the reality we face and the scope of the proposed solutions.

2. Impact

The scale and time of the solution influence its impact. On the one hand, a set of partial and isolated actions, many of them late, will hardly serve to combat misinformation in the public debate. On the other, these actions may have an effect contrary to what is sought.

On the possible undesired effects, the evidence is still unreliable. Nonetheless, some studies indicate that verification tools and warnings can have a negative impact. According to an investigation by a group of academics -also under review-, labeling fake stories as such does not necessarily change the user's perception of them. Furthermore, which is even more serious, if a user begins to see stories where there are warnings of possible falsehood, they may conclude that all those that do not have a warning are true, which, of course, is a wrong generalization.[94]

Regarding the first point, this document does not intend to disqualify actions such as fact-checking, using warnings or context articles. These are answers that seek to ponder the problem of misinformation guaranteeing the freedom of expression of users. Discarding restrictive or openly arbitrary measures hinders the solution of the problem, but prevents the creation of worse ones. From this perspective, it is also relevant to analyze the decisions to hide or remove contents under potentially arbitrary categories, such as "low quality" or not being a "reliable" source. These schemes based on reputations run the risk of favoring mainly mass and commercial media, to the detriment of voices that cannot access this type of alliance, but whose content is legitimate and relevant.

The measures to deal with misinformation are located in the context of the terms of service of the platform and, in particular, the moderation of content. Therefore, following the proposal of the Special Rapporteurship for Freedom of Expression of the Inter-American Commission on Human Rights, "in the design and setting of their terms of service and community rules, companies should not limit or restrict the freedom of expression disproportionately or unnecessarily".[95]

---

[94] See, Pennycook, Gordon, et. al. "Prior exposure increases perceived accuracy of fake news". Version of August 26, 2017; Pennycook, Gordon, et. al, "Assessing the effect of "disputed" warnings and source salience on perceptions of fake news accuracy". Version of September 15, 2017. These texts have not had peer review.

[95] Special Rapporteurship for Freedom of Expression of the Inter-American Commission on Human Rights, Standards for a Free, Open and Inclusive Internet OAS OEA/Ser.L/V/II CIDH/RELE/INF.17/17 March 15, 2017, Original: Spanish, Para. 98

3. The role of civil society

Most strategies to deal with misinformation include the participation of civil society. Whereas companies work with organizations to verify information, make warnings or provide more context. This participation is essential, as long as these initiatives are set regionally, they should include organizations that have legitimacy and knowledge to weigh information about the public debate. The user, however, is less important in this discussion. Except for Google feedback, it is not clear how the user feedback is taken into account to address misinformation in these services.

Finally, at this point it is necessary to take into account that civil society is also an actor in the production of misinformation: political parties, communication agencies and different interest groups are part of the problem which is currently expected to be solved directly by platforms. Concerted actions, for example, show an organized and systematic process to exploit the services for the benefit of an individual purpose and to the detriment of the general interest.

While this does not remove liability from intermediaries, it does highlight the need to seek answers in addition to those of a preventive and educational nature, in different economic, political and social sectors. In other words, the solution to this problem is not limited to the traditional actors of Internet governance.

4. Transparency

The development of this document had a constant difficulty: to understand if the large number of decisions and actions covered in the media and announced by the companies were being effectively implemented and to what extent. In fact, as these words are being written, measures that would reverse the changes announced recently and referenced in this text are being reported.[96] Ultimately, the conclusion on this point is that many of the measures are just announcements, and those that are implemented reach only a partial and temporary degree of application. In any case, it is not possible to determine it accurately, which shows a serious problem of transparency.

It is understandable and desirable that companies experiment with possible solutions. Misinformation must be faced with creative and innovative actions, and strategies cannot be set in stone. However, the lack of clarity for users and the general public obscures the understanding of the problem and prevents civil society from making prudent feedback.

Technology produces informative returns as its adoption becomes more widespread. Through this use - and the network effect - data is obtained which serves to improve and adapt this technology.[97] That social return must also be widespread, or at least have some degree of openness. Knowledge of how misinformation works and measures to address it are controlled by companies that are unwilling to share it with civil society. There is no clarity in the diagnosis or in the responses adopted, all of which adds to the existing questions against intermediaries for their lack of transparency in the way they provide their service. Transparency, to this extent, is not a subsequent unilateral action, but a joint process that is part of the solution.

---

[96] Cohen, David. "Facebook Is Opting for Context Over Labeling in Its Battle Against Fake News". *Adweek*, December 21, 2017. Retrieved from: http://bit.ly/2Ero77P.
[97] See MacKenzie, Donal. Knowing Machines. Essays on Techincal Change. MIT Press, 1998.

CELE
Centro de Estudios en Libertad de
Expresión y Acceso a la Información

22