



Dipartimento di Scienze Politiche

Cattedra di Digital Economy

**DISINFORMATION IN THE DIGITAL ERA:
A FOCUS ON THE COVID-19 PANDEMIC**

RELATORE:

Prof.ssa Cristina Alaimo

CANDIDATO: Noemi Bruno

Matr. 085712

ANNO ACCADEMICO 2019/2020

TABLE OF CONTENTS

INTRODUCTION.....	3
1. ORIGIN AND DEFINITIONS OF DISINFORMATION.....	5
1.1 Online news consumption on the web.....	5
1.3 Social media fruition.....	7
1.4 Fake news evolution.....	8
1.5 Classifications of disinformation.....	9
1.6 Reasons behind creation and diffusion of false information.....	11
1.7 The role of technology.....	14
1.8 Considerations about the context.....	15
2. HOW TO DETECT AND DEFEAT IT.....	17
2.1 Current challenges.....	17
2.2 Necessary contributions for detecting.....	18
2.3 Attempts to combat fake news.....	25
3. CASE STUDY: COVID-19 INFODEMIC.....	27
3.1 COVID-19 pandemic.....	27
3.2 The role of social networks.....	28
3.3 Novel coronavirus-related fake news.....	30
3.4 Multidimensional responses.....	33
3.4.1 Individual awareness.....	33
3.4.2 International commitment.....	34
3.4.3 Policy actions.....	35
3.4.4 Platforms answers.....	36
CONCLUSION.....	38

***ABSTRACT* 39**
***BIBLIOGRAPHY and SITOGRAPHY* 41**

INTRODUCTION

“The new source of power is not money in the hands of a few, but information in the hands of many.”

-John Naisbitt

Media have always influenced the perception of reality. However, the advent of the so-called new media has broadened the boundaries of what has become mediated and deeply transformed access to and, as a consequence, perception of reality. New media and the internet over the years have become to be dominated by few powerful players who have gained the ability to affect individual perception on a global scale. From the spread of social media, in particular, the level of global connectivity and interconnection has led to massive changes in the news market and its main actors. As a matter of fact, social media platforms have removed most of the limits that were imposed to news production and distribution and their space, time or costs, and this in turn has procured an increase in both consumption and production.

The amount of information in circulation is relentlessly growing and the rate it spreads is becoming faster. This is also due to the fact that information production is no longer a prerogative of information specialists: everyone can contribute to it. Hence news and information generation are not regulated anymore or guided by professional ethic, and their quality is considerably altered. All these changes have conditioned a gradual and increasing separation of news and information from reality. At the same time, the pervasiveness of the mediation offered by the digital world and its false immediacy implies a greater difficulty for individuals which cannot separate accurate information from verified news or sources. People, indeed, have trouble in clearly distinguishing online fabrication or conspiracy theories from fact-checked information and tend to trust in whatever they see online. As a result, they take as real both accidental mistakes made by algorithms behind web pages, as well as fake news purposely released by malevolent users. The additional risk lies in the fact that the continuous flow of information and its powerful digital infrastructures produces immediate and global-scale dissemination, conditioning what seems as an increase of more and more people believing in inaccurate information. Considering that beliefs have an impact on actions, disinformation has indeed the power to affect society to a great deal.

The purpose of this thesis is to describe the phenomenon of misinformation and illustrate the various means to combat it. The first chapter gives an overview of the various definitions of false news and their different nature, illustrating the many reasons behind their creation. For instance, news may have been created with satirical or provocative intent. They could constitute means to make propaganda or influence political affiliation. One of the most common reasons is indeed to generate a profit. Besides the motivations that can lead to fabricate news, the techniques used in their fabrication can also be varied. One of the most successful is to use the so-called *clickbait* titles: they are sensationalist titles fabricated to easily attract a wide audience, which often tends to reshare such articles without checking if they have any link to reality. Additionally, there are also several techniques which refer to the various ways in which content can be manipulated. For instance, a content that is not completely fabricated can be described as misleading, when an element has been omitted or manipulated. Another type of manipulation may concern the source, which can be modified to look reliable, or the context. The latter in particular is an essential element to evaluate the misleading effect of the news, as any news can become false or misleading if extrapolated from the context to which it belongs.

Given the number of typology and classifications, the second chapter analyses the verification process, which must be carried out at several levels. Firstly, it is necessary that users learn to detect some features that are commonly considered distinctive of fake news. Among these there are syntax and lexicon, structural elements such as URL, grammatical or spelling errors. The reliability of the source is also important, as well as the presence of references. Since the phenomenon has such a wide extension, methods have been developed that can help common users. First of all, there are the so-called professional *fact-checkers*, which are experts who assess the news in their area of competence, that have created websites and an international association for this purpose. Following the advancements in Artificial Intelligence (AI), a number of algorithms and systems have been developed. These are in fact able to replace humans in the phases of identification, verification and correction, on a massive scale leading to time and result optimization. Disinformation has developed so much that several measures have been taken on multiple levels: from international organisations to local governments, from online platforms to individual citizens. Such spread of action has characterized a phenomenon which has invested the whole world in recent months, and which has been called *infodemic*. The third and last chapter discusses some of the themes described in previous sections with the use of an illustrative example. The chapter is dedicated to the over-abundance of COVID-19-related news and information which have circulated mainly through social network, fed by several actors operating on different levels. During one of the most delicate moments in our recent

history, the consequences of *infodemic* risk becoming as serious as the pandemic to which they are related.

1. ORIGIN AND DEFINITIONS OF DISINFORMATION

1.1 Online news consumption on the web

The beginning and development of the web is intimately connected to the searching and sharing of information and news. The development of new technologies has always posed a threat to the existing media industries: first with radio, then TV and finally the Internet and the web. The latter, particularly, has caused epochal changes, eliminating costs as well as space and time barriers in news production and consumption and allowing a direct connection between advertisers and consumers. In fact, researches show that people's search for information has increased exponentially compared to the time when getting it was necessarily connected to newspapers, books or face-to-face communication, tools that differ profoundly from the Internet.

One of the main elements that characterizes this digital revolution is the transition from a linear business model to a multi-sided market¹, which is structurally different from existing models. First, the World Wide Web is built on the decentralized infrastructure of the internet which has been designed to avoid centralised control. Even during the newspaper era, it was possible to be in a situation where a certain article reflected a specific political point of view or idea. However, the publication of news was subjected to a given procedure which generated a particular degree of widespread trust among readers. On the Internet, instead, compliance with this procedure is not mandatory. Everyone can easily become the author of contents that will get to the attention of the reader through a variety of channels such as aggregators or emails, without undergoing prior editorial review. The only exception is represented by parallel sites to existing newspapers, which however constitute only a small percentage of the channels adopted by users for the consumption of news.

¹ Martens B. et al. (2018). The digital transformation of news media and the rise of disinformation and fake news. *JRC Technical Reports*. JRC 111529

In addition to direct and targeted news search, the main ways employed to acquire information are search engines and social media. Since technologies have completely modified both fabrication and distribution of news, it is no longer required to be professional journalists in order to produce news. Bloggers, vloggers, influencers can create them and assume the role of distributors, since anyone interested in their contents is only one click away from getting it within their websites. If, on the one hand there is a gain deriving from the vertiginous reduction of the costs of distribution, on the other hand, this corresponds to a loss in terms of reliability and quality of news and information. One of the main changes resulting from the digitalization of newspapers is the pace of publication. The fruition of articles can take place at any time, triggering a strong competitiveness among different websites. This can lead to two effects: on the one hand, in order to speed up the process, time is taken away to the verification of facts; on the other hand, for the purpose of prevailing over the competitors, news creators employ ad hoc instruments. For instance, they can choose to use expedients like *clickbait* titles or, in the case of already well-known newspapers, *freemium* access, which consists in providing a number of articles for free, while others can only be read by those paying a subscription fee.²

In order to optimize the additional cost of spending time in doing research, people are using more and more search engines, in which an algorithm has previously performed the job of sorting contents. It follows that users simply need to type a keyword in order to reach a temporary achievement called Search Engine Results Page, which shows just ten or twenty results. However, the choice of which contents are to be displayed and in which order is not regulated nor controlled by a single institution; on the contrary, the outcome is affected by search engine companies, consultants or content producers. As a consequence, it is not uncommon that they are biased.³ For example, after the 2016 U.S. presidential elections, if users were looking on Google for information about the winner, the top news link was from “70 news”, a fake news site with false numbers.⁴ In addition, an analysis revealed that up to 30% of presidential and senate candidates had their search results affected by potentially fake or biased contents.⁵ The problematic nature of controlling and adjusting these results is to be situated at the infrastructural level, since the algorithms behind them are inaccessible and opaque. In

² *Ibidem*

³ Introna L. & Nissenbaum H. (2000). Shaping The Web: Why The Politics Of Search Engines Matters. *The Information Society*. 16. 169-185. 10.1080/01972240050133634

⁴ Gray J., Bounegru L., Venturini T. (2020). ‘Fake news’ as infrastructural uncanny. *New Media & Society*. 22(2):317-341. doi:10.1177/1461444819856912

⁵ Metaxa D., & Torres-Echeverry N., (2017). Google's Role in Spreading Fake News and Misinformation. In *SSRN*. <https://ssrn.com/abstract=3062984> or <http://dx.doi.org/10.2139/ssrn.3062984>

particular, the ones behind search engine, when generating results, combine comprehensive data of different nature and from different contexts. First, in fact, since users are used to rely on the ranking that is placed before them, articles that are in the first places continue to be preferred, maintaining a prominent position. Moreover, the algorithm is able to detect each user's pre-existing preferences from their historical search activities. Such algorithms are therefore capable of performing tasks that once required human intervention, which is also the reason why they are called learning algorithms. They are not based on pre-specified instructions but on connections that change and get redefined each time data is added: since the resulting outcomes may be inexplicable even to those who created the algorithm, this characteristic is known as 'black-boxed performance'.⁶

In order to exercise a certain degree of control on the algorithm, a practice known as Search Engine Optimization is being employed; as opposed to it, however, there are some techniques able to make fabricated news appear as trending and included among the first stories shown by search engines.

1.3 Social media fruition

The channel that is becoming increasingly central for the fruition of news, especially within younger generations, is that of social media, since they are visited daily for several hours.

The reason behind this trend is that their use allows to optimize the cost of time to the maximum: while scrolling the Facebook homepage it is possible to stay in touch with friends, stay up to date on the events happening near, learn what is going on the other side of the world. All simultaneously and at any time and place.

Also, social media have an additional feature to their advantage: the links between users. While browsing social networks, user will find both the contents shared by their selected sources and friends and will be able to choose which to display first. Then, all these results are filtered by an algorithm which selects, among them, the more suitable contents for the user, who will also have the opportunity to comment on them. This process creates a dense and uninterrupted network of connections, bringing out a new type of news market that is multi-sided, thanks to the admission of these new digital platforms. In fact, they are able to encode data deriving from the activities that can be carried out by

⁶ Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28 (1), 62-70. <https://doi.org/10.1016/j.infoandorg.2018.02.005>

users on social media; once encoded, they assemble it and then extract value, profiting from it and bringing benefits to both readers and advertisers.⁷

Reading news has always been functional to the knowledge of facts that otherwise would have remained unknown to many. Nonetheless, it is not possible to reproduce a completely accurate representation of reality, hence writers have to implement a selection of information. Therefore, the use of social media puts the user in front of news resulting from two selection procedures: that of the writer who selected the fact and that of the algorithm that analysed users' interests to filter the results. Readers will thus have to be very careful in the evaluation of news, in order to prevent prejudices and previous opinions of those who made the first selection. But they too will be influenced by their own biases, which are fed in addition by the presence of only news in accordance with their opinions. Consequently, compared to the tools through which people traditionally learned the news, the use of social media may cause a lack of objectivity in the context of both production and consumption. In addition, the task of distribution has been assumed by online platforms: this has led to a drop in costs on both sides of the market and has made advertising the only source of income. The combination of these changes has had as immediate result a huge boost to the proliferation of false news.⁸

1.4 Fake news evolution

Fake news phenomenon existed long before the advent of the Internet. Clear examples are given by the *disinformation* of the World Wars, as well as the general practice of publishing unverified news during the Spanish War, that was known as *freak journalism* or *yellow journalism*. To better understand the differences between then and now, an example concerning the Belgian newspaper *Le soir* is particularly explicative.⁹ In 2017 *The Guardian* published an article about sites with domain names and features almost identical to news websites: the *doppelgänger* of *Le soir*, in particular, dealt with an alleged economic support from Saudi Arabia to the presidential candidate Macron. In 1943 an elaborate fake copy of the same newspaper circulated containing attacks against Nazi together with falsified advertisements and news. Clearly, the two events took place in two completely distinct moments in history, therefore they involved different infrastructure and

⁷ Alaimo C. & Kallinikos J. (2017). Computing the everyday: Social media as data platforms. *The information Society*, 33(4),175-191.

⁸ Martens B. et al. (2018). The digital transformation of news media and the rise of disinformation and fake news. *JRC Technical Reports*. JRC 111529

⁹ Gray J., Bounegru L., Venturini T. (2020). 'Fake news' as infrastructural uncanny. *New Media & Society*. 22(2):317-341.

doi:10.1177/1461444819856912

expedients. As a matter of fact, the 1943 copy had to be printed and infiltrated by volunteers inside the official Belgian kiosks: as a result, its spread was relatively restrained and easy to shut down, and those responsible were arrested. As regards the 2017 one, instead, it required both different skills and the help of some particularly credible channels which re-shared it. At the end, although these sites promptly removed the fabricated news, online platforms' infrastructures made it practically impossible to completely and definitively remove the news, and to track down those responsible for its spread.¹⁰

As a result, it can be said that the digital space appears to be the ideal place for the dissemination of *fake news*, which are basically facts interesting enough to attract the attention of the masses and that seem to be true, even if they have no evidence in their support in reality.

Thanks also to numerous political-electoral campaigns, this term has acquired great popularity, so much to be considered by Collins word of the year 2017 and included by the Reuters Institute among the trends for journalism in the same year.¹¹ Such popularity has led, however, to erroneously encompass under this expression also other types of content and to create numerous categorisations based on different factors.

1.5 Classifications of disinformation

One of the main classifications is the one established by First Draft¹², a nonprofit organization founded in 2015 with the purpose of providing guidance to fight deceiving information. Firstly, they recognize three concepts that can be summarized in the expression *information disorder*: disinformation, misinformation and malinformation.

The concept of *disinformation* implies the intention to deceive, confuse and create uncertainty “for the sake of it” or for economic or political reasons. Often, the sharing of such news transforms it in cases of *misinformation*: the fact is not true, but those who share it are unaware. Such diffusion may take place via any form of communication, but what accelerates it is the interpersonal one, especially if digital. In fact, quite often those who share such news tend to do so without prior verification, to conform to common thought, especially on social networks.

¹⁰ *Ibidem*

¹¹ Habgood-Coote J. (2018). The term “fake news” is doing great harm. *Theconversation.com*. <https://theconversation.com/the-term-fake-news-is-doing-great-harm-100406>

¹² Wardle C. (2019). Understanding Information Disorder. *Firstdraftnews.org*. https://firstdraftnews.org/wpcontent/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x76701

The opposite type, meaning the sharing of stories which are authentic but decontextualized and twisted, is called *malinformation*. There are two main reasons behind the spread of this genuine contents: on the one hand, they attract audience easily, on the other they obstacle Artificial Intelligence systems' fact-checking operations. This is one of the reasons behind the integration of *satire or parody* among the further seven categories identified by First Draft within these three types of information. Moreover, although many disagree in its inclusion as not negative in itself and actually constitutionally guaranteed as an art form in several systems, like the Italian one, there is a further issue. As a matter of fact, since the network of interconnections that the web generates is potentially infinite, the farther from the original sharing of the news, the more likely the perception of the real message is to be lost. An example of it is the site created by a Donald Trump's consultant as a parody of Joe Biden's official website, compared to which it was even indexed higher by Google at the launch of the election campaign in April 2019.

False connection refers instead to techniques used by journalists with the aim of attracting more readers, despite the fact that in the long term are likely to lead to the opposite result, fostering growing distrust of this category. Such practice is commonly known as *clickbait* and it is so widespread that it led to change the Facebook algorithm in order to place these kinds of articles further down in the ranking.

About contents, the organization distinguishes four different categories: *misleading, imposter, manipulated and fabricated content*.

The first one implies other types of deceptive techniques, like, for instance, to omit a fragment of something that can be a photo, a quote, a statistic. It can be hard to be detected, since it is not an obvious manipulation, but rather a small change that can impact the whole interpretation.

Imposter contents leverage on the assumption that people are inclined to believe in news published by an already known brand with no questions asked: the result is that those who want to spread false content, tend to copy their logos, as in the case of the lookalike *Le Soir*.

The difference between the last two is that one consists in an altered evidence, the other is completely false. For instance, a *manipulated content* is the video of the speech given by the U.S. House of Representatives Speaker Nancy Pelosi in May 2019, that was slowed down in order to make her look drunk; an example of a *fabricated* one, instead, was the alleged endorsement of Pope Francis toward Trump at the 2016 presidential elections.

Finally, cases of actually happened but represented in a *false context* stories can also occur.¹³

¹³ *Ibidem*

1.6 Reasons behind creation and diffusion of false information

Once these different types of subject characterizing the previously mentioned information disorder have been identified, it is necessary to understand the justifications behind their creation. In this regard, once again the First Draft author Claire Wardle, created a model based on ‘8 Ps’¹⁴: these false stories can be examples of *Poor journalism* or simple *Parody*, they can have intention to *Provoke*, demonstrate *Passion* for a certain theme, foster someone’s own *Partisanship*, exert some *Political influence or Power*, make pure *Propaganda*, or finally to *Profit*, as happened to certain American fake news geniuses. Hence it can be concluded that the virality of a story can be fostered by any of the actors constituting this hybrid contemporary media environment, and that they can be both in bad and in good faith.

In the first instance, circulation can be traced back to social groups and lobbies interested in influencing public opinion, or to proper disinformation campaigns concerning burning issues like immigration or health. On the other hand, a genuine sharing can be generated by the features of the web infrastructure, from which the excessively accelerated pace of newspapers publication and the endless interconnections between users derive. Given the importance of this latter mechanism, there are additional points of view proposing to shift attention from the perspective of the creator of the news and their intention to that of the one who is exposed to it.

The premise is that Information is such only when someone pays attention to it, so in its evaluation prior knowing and prejudices must be taken into account.¹⁵ Based on them, indeed, the individual will also judge the alleged intent and objectivity of the fact. For instance, when on November 18th of 2016 Chris Lamb published on the *Huffington Post* a satirical piece stating that Donald Trump racial intolerance could have gone as far as to order the removal of the Statue of Liberty, the news has been taken seriously and reported on several newspapers, also by the Italian *Corriere della Sera* and *La Repubblica*. This is a further evidence of the fact that the level of interconnection is so high that people from distant countries, with various backgrounds and degrees of education, could get in touch with each other’s ideas in an infinitesimal period of time.

¹⁴ Wardle C. (2019). Understanding Information Disorder. *Firstdraftnews.org*.

https://firstdraftnews.org/wpcontent/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x76701

¹⁵ Giglietto F., et al. (2019). ‘Fake news’ is the invention of a liar: How false information circulates within the hybrid news system. *Current Sociology*. 67(4):625-642. doi:10.1177/0011392119837536

Especially thanks to social media, this flow of communication can potentially be continuous and endless; therefore, in the process of spreading news, different interpretations are involved that can lead to various outlines.

From this perspective, the case in which both the creator and propagator share the news while being conscious of its falsehood, is called disinformation. This scenario can concern both the already-mentioned examples of sharing made in good faith like humoristic content, and propaganda. It may happen instead that whoever wrote the news has omitted or wrongly report something, while the one who shares it is aware of this error: this is misinformation propagated through disinformation. The opposite situation, disinformation spread through misinformation, occurs when someone propagates news while being sure of its truthfulness, without knowing that the author created a false content. Finally, a further circumstance is when both who creates and who disseminates the news think that it is true: the most common example is the propagation of conspiracy theories.¹⁶

The credibility of news is then further influenced by a series of cognitive biases that characterize the human psyche, pushing it to be more attracted by personalized messages. For example, the *bias blind spot* that encourages individuals to notice the mistakes of others but not their own, or the *confirmation bias* which leads people to believe in interpretations in accordance with their prior ideas.

More specifically, according to the *Center for Information Technology and Society*, the influence fake news has on society is fostered by four types of mental shortcuts.¹⁷

The first one is the tendency to stop at the appearance without trying to elaborate a more accurate assessment. It is in fact a common practice, especially on social media, to share an article after having read just the title, therefore ignoring all the points of view that characterize the analysis of news. This behaviour contributes not only to the dissemination of potentially inaccurate stories between groups of people who have direct contacts, but also to their overall popularity, creating a “socio-cognitive epidemic”.

Another likely consequence of fake news is the so-called *bandwagon effect*, meaning the inclination to trust something for the simple reason that everyone else seems to believe in it. Consequently, it is likely to create a chain of people sharing news without having read it.

The third mentioned bias is based on research conducted during the 2016 U.S. elections concerning the political orientation. It would seem in fact that Trump’s supporters were more likely than

¹⁶ *Ibidem*

¹⁷ “Why We Fall for Fake News”. *Centre for Information Technology and Society*. <https://www.cits.ucsb.edu/fake-news/why-we-fall> [Ultimo accesso: 18 settembre 2020]

Clinton's to share false stories articles, so much that several private individuals who tried to profit from *fake news* claimed they stop sharing pro-liber articles.

The last one is about a cognitive aspect of the persistence of *fake news*, called *belief echo*. Studies have demonstrated, indeed, that, if people once had faith in a fact, they will tend to persist in believing, even when facing denials and corrections about it. Even more when such revisions concern genuine warnings: the fact of having found in some cases the possibility of their existence, pushes people to believe more easily in what they read if they are absent.¹⁸

Furthermore, the current online communication is not purely textual but rather multimodal, meaning that it also relies on photos and videos. It is in fact undeniable that an image has a stronger and more direct communicative power than words. For this reason, the use of such tools can also be decisive in the dissemination of incorrect information, since the direct vision of something makes it inevitably more believable.

The manipulation of images is in fact one of the most used techniques among agents of disinformation, whereas it has even become more manageable, thanks to the invention of numerous software. From this perspective, there can be four types of disinformation: using real visuals accompanied by deceptive texts is called *decontextualization*, while *multimodal doctoring* is the case in which they are both falsified; *reframing* is the practice of modify photos or videos in order to stress a certain aspect, and *visual doctoring* to manipulate them distorting reality.¹⁹

However, what is certain is that, in order to understand the role and meaning of false information, it is not possible to disregard the online context in which it circulates, because the infrastructures and methods every online platform use to collect and analyse data can play a fundamental role in their creation and dissemination.

According to the *Global Disinformation Index*, in fact, disinformation must be considered as a process involving several actors driven by different reasons, but united by the fact of abusing the system that is behind online advertising. Among them, there are the politically-motivated ones: *state actors*, both governments and linked figures, and *grassroots trolls*, individuals or groups acting for a common cause.

¹⁸ Thorson E. (2016). Belief Echoes: The Persistent Effects of Corrected Misinformation. *Political Communication*, 33:3, 460-480, DOI: 10.1080/10584609.2015.1102187

¹⁹ Hameleers M., et al. (2020). A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication*. 37:2, 281-301, DOI: 10.1080/10584609.2019.1674979

Another purpose may be the financial one, which connects *private influence operators*, such as for-hire companies which want to disinform for their high-paying clients, and *pure rent-seekers*, whose only purpose is to generate traffic on their sites since this will create for them a profit.

These actors mutate the advertising world availing of four trends: the first is the one that sees online media as the new protagonists in the spreading of news; the second focuses on the investments of both brands and politicians in online advertising, which in 2016 surpassed that of television. This led to the third trend, an increment in the number of ad-tech companies. With respect to the last one, every time users enter a website, advertisers are able to collect more and more data about them, in order to target ads on the fly; at the same time an automated advertising auction is generated. Through this system, it is possible to spread any type of content: those who want to disseminate disinformation can employ the same tools to address it to a specific group, which may in turn interact with the message and spread it further. In the same way, these interactions will generate data which will be used again with the aim to customise advertisements to the audience.²⁰

Therefore, the production, spread and effects of *fake news* are a synthesis of a number of variables that can be cognitive, social, algorithmic.

1.7 The role of technology

From the creation perspective, a number of advanced techniques can be found.²¹

One of these is represented by the existence of technological systems capable of generating texts on any subject, in a form that replicates natural language. However, such devices have a vocabulary and a mastery of limited grammatical rules, so that it is possible to find semi correct sentences.

Other kinds of tools can be employed to counterfeit voices: they can be converted, modified or morphed, but there are several ways to discover a scam like this.

Related to videos, several potentially dangerous techniques are increasingly used. For instance, starting from a file audio, it is possible to create a credible video also from the point of view of

²⁰ Melford C., Fagan C. (2019). Cutting the Funding of Disinformation: The Ad-Tech Solution. *Disinformationindex.org*. https://disinformationindex.org/wp-content/uploads/2019/05/GDI_Report_Screen_AW2.pdf

²¹ Wang P., Angarita R., Renna I. (2018). Is this the Era of Misinformation yet: Combining Social Bots and Fake News to Deceive the Masses. In Companion Proceedings of the The Web Conference 2018 (WWW '18). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1557–1561. DOI: <https://doi.org/10.1145/3184558.3191610>

synchronisation of lips movements. Similarly, to extrapolate and transpose facial expressions or even to swap faces from one video to another is increasingly gaining popularity.

From the distribution perspective, distorted information may have either an internal source or an external one. In the case of Twitter, an example of the first one could be a citation or retweet, the other instead is an external link. Usually, there is an introduction from the outside of a news that is then shared, therefore internalized by the platform in question. Given this process, it is common to employ software-controlled profiles or pages in order to spread fabricated news; they can produce content as if they were humans and can also interact with them, although their action will always be influenced by the conditions given by the platform and by the human action that created them.

The role of this *social bots* in dissemination can be understood by analysing the tweets published during ten months in 2016 and 2017.²² Focusing on the dissemination of low-credibility contents, finding proved that they have the same potential of virality as verified facts. Differences has emerged, however, in the diffusion of these two types of news: not extremely realistic content was hardly shared as part of a conversation and almost always by the same accounts. Since the only plausible explanation was that such stories were automatically shared, two tools have been developed and used: the Hoaxy platform and the Botometer machine learning algorithm. Through them, it was possible to draw different conclusions. First, these accounts are used to operate during the first seconds of the news' appearance on the social media, meaning before it reaches a level of diffusion to be considered viral. Also, in posting links that lead to fake news, they often mention popular accounts with a spread audience, probably hoping for a re-sharing.

However, the chain of sharing that then makes a news viral, is carried out by accounts held by real people. A further finding deriving from another research during and right after the U.S. presidential campaign of 2016, is that these bots are able to falsify their geographical position, managing to reach a wider audience.²³

Repressing these automatic shares would therefore reduce at least to some extent the spread of false information, although the use of algorithms with this purpose could lead to censorship.

1.8 Considerations about the context

²² Shao C. et al. (2018). The spread of low-credibility content by social bots. *Nature Communications*. 9. 10.1038/s41467-018-06930-7.

²³ Shao C. et al. (2017). The spread of fake news by social bots. *Researchgate.net*.

https://www.researchgate.net/publication/318671211_The_spread_of_fake_news_by_social_bots/citations

Finally, it is necessary to consider that the circulation of *fake news* on such a large scale is taking place in a context of decreasing trust in news and social media, as the 2017 *Reuters Institute Digital News Report* demonstrates.²⁴ According to a research carried out through an online survey including around 70,000 respondents across 36 markets, in fact, many believe that a big incentive to the propagation comes from the contextual presence of algorithms and absence of rules. Nevertheless, the number of people who prefer consuming news through algorithmic selection is greater than those who rely on journalists, especially among young people. One of the justifications is that many perceive such technologies as more impartial.

As a consequence, there are repercussions deriving from the employment of social media, search engines and aggregators. In the first place, the offered amount and type of news are much wider and more diverse than they would be if news were only read on newspapers. It is also true, however, that people are more inclined to follow on social media for example a political spokesman of already internalized ideas, rather than one with opposite values. In fact, authors found out that media distrust is usually related to the fear of political biases.

As relate to different brands producing information, first of all, credit of news is much more commonly given to the platform than to them.

Another negative repercussion is that only few people are willing to pay for news and the main reason is the huge amount of information available for free, for everyone at any time; so many that individuals frequently avoid reading the news they are facing. This is because social media as well as aggregators involve frequent incidental exposure to news, thus increasing the probability that what is read is contrary to individual's beliefs, or incorrect. Moreover, it has been proved that credibility judgement also depends on the number of accessible information. In fact, when reading two pieces, individuals will be more inclined to deepen the one that does not reflect their prior point of view; if there are several information, instead, they will automatically tend to consider more reliable those conforming to their own ideas.²⁵ It is therefore essential to develop mechanisms enabling people to objectively assess the quality of information. Especially in order to maintain a democratic order that works properly. For this to happen, a strong collaboration is necessary: from governments to ordinary citizens, from journalists to owners of major platform.

²⁴ Newman N. et al. (2017). Reuters Institute Digital News Report 2017. *Reuters Institute for the Study of Journalism*. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf

²⁵ Fischer P., Schulz-Hardt S., Frey D. (2008). Selective exposure and information quantity: how different information quantities moderate decision makers' preference for consistent and inconsistent information. *J Pers Soc Psychol.* 94(2):231-44. doi: 10.1037/0022-3514.94.2.94.2.231. PMID: 18211174.

2. HOW TO DETECT AND DEFEAT IT

2.1 Current challenges

In March 2017, during the 28th anniversary of the World Wide Web, its founder and inventor Sir Tim Berners-Lee explained that he conceived it as a space that would allow anyone to share information overcoming geographical and temporal barriers. In recent times, however, three phenomena occurred in contrast to his initial idea: loss of governance over personal data, uncontrolled dissemination of information, non-transparent use of advertising.²⁶ Taking part in online life means agreeing to cede personal data in exchange for free services, often without even actually realizing it. It could be argued that those who owns data do not make it a weapon of destruction but uses it purely for profit. However, it must be considered that, by giving up possession, personal information could be sold to anyone and for any purpose. Furthermore, those who can access this huge amount of data, are able to perfectly frame user preferences and produce information based on it, making extremely easy to manipulate minds. In fact, for instance, it is not uncommon for politicians to increasingly use personalised advertising for a particular type of audience: in this way, the same campaign can be adapted to the preferences of the reference electoral basin, deeply affecting the democratic process. Once the importance of knowing people's preferences has been understood, digital market started to focus on advertising, to generate more and more earnings. There are several channels through which online advertising can take place. For example, a distinction must be made between *display advertising*, where advertisement banners are placed by web publishers on their websites, and *search engine marketing*, where ads are shown along with search word results. *Affiliate marketing* involves two organizations: one allows the other to use their website to advertise, and eventually revenues are shared. Finally, businesses can take advantage of *email* or *social media*, both to show a product or service to their customers and to attract new potential buyers.²⁷

²⁶ Three challenges for the web, according to its inventor (2017). *Web Foundation*. <https://webfoundation.org/2017/03/web-turns-28-letter/>

²⁷ Kapoor K., Dwivedi Y. K., & Piercy, N. C. (2016). Pay-per-click advertising: A literature review. *Marketing Review*, 16(2), 183-202. <https://doi.org/10.1362/146934716X14636478977557>

Such changes involving advertising environment have a significant impact on the spread of *fake news*: while the prior system was oriented towards control and quality, this one focuses on any type of content that is easily and virally shareable. In fact, nowadays companies remunerate relying not on news itself, but on the number of clicks and views it has generated. Usually, ad agencies avail of mediators which publish the news, so, based on the resulting views, advertisers are charged, and money are relocated to the intermediary; except in the case in which intermediary own the space of publication, as for Google and Facebook, that will withhold the entire sum.

Ultimately, in different ways and for different reasons, several actors make use of the publication of untrue news: as Berners-Lee claimed, the current structure of the internet is the outcome of all the actions of every single actor that makes or has been part of it.²⁸ From political figures using fake news to undermine the integrity of other actors or systems that are at odds with their policies, to news media that feed them to try to conform to the current logics of news propagation. Then, it must be considered that such propagation is promoted on one hand by many tools characterizing the current digital media, such as the previously mentioned ad system, on the other by civil society, often unconsciously.

2.2 Necessary contributions for detecting

In order to succeed in safeguarding the digital environment from the threat of disinformation, the contribution of everyone is necessary. Among the various theories developed for this purpose, some relate disinformation to news itself, others to the user. The former assumes that *fake news* has distinctive stylistic features, as the quality and quantity of used words, which are essential elements to be considered in the identification process.²⁹ In the paper *“Fake News” Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content* (Molina, M. D. et al, 2019). for example, authors considered some of the several types of content which are commonly summed

²⁸ Three challenges for the web, according to its inventor (2017). *Web Foundation*- <https://webfoundation.org/2017/03/web-turns-28-letter/>

²⁹ Zhou X. & Zafarani R. (2018). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities.

Researchgate.net

https://www.researchgate.net/publication/329388190_A_Survey_of_Fake_News_Fundamental_Theories_Detection_Methods_and_Opportunities

up by the term *fake news* and then compare them to authentic news, in order to avoid confusion and create systems that can automatically detect them.

Their codification is based on several points of view. In that of message and linguistic, inherence to reality, structural lexical and syntactical qualities, headline, visuals' truthfulness are considered. Another one is about verifying sources and intentions, and their reliability. Then, attention is given to the structure, in particular URL, About Us and Contact Us sections, and comments. Finally, it is important to evaluate the network, therefore author, metadata, social media shares and practices of personalization and customization. In fact, articles containing false information are likely to include errors concerning grammar, punctuation, spelling, vocabulary. As concerns references, they are usually not mentioned or not verified, while URLs are typically similar to those of known sources, but ending for example in *.com.co*. However, authors specify that, if taken separately, the various characteristics could mislead, as in the plausible case of real news containing errors.³⁰

This is just one of the many existing classifications considering a few aspects, but which is founded on a fundamental premise: the first step of the battle against disinformation is the awareness of common users.

The theories focusing on users, instead, investigate their behaviours and interactions, starting from the fact that those who contribute to the propagation of fake news are divided into normal and malevolent users. In the fight against disinformation it would be necessary to intervene differently on each other.³¹

Anyway, the numerous approaches developed to deal with disinformation could be very different, but they seem to have nearly common goals: identify misleading news spread on the internet; verify or assist in verifying the veracity of claims; disseminate any corrections in real time, so that they can reach an audience that is as wide as possible. In order to achieve these objectives, it is crucial to observe and study various forms of public discourse through progressive stages. The first step is that of identification, which consists in draw on media sources to then distinguish the concrete affirmations from those that need to be verified. As far as verification, however, the facts may not always be classified in true and false, so they must be compared with something else. A reasonable type of comparison is that with facts that have already been verified, but two limitations of this approach may be the cases where a sentence is paraphrased, or situations that involve slight changes

³⁰ Molina M. D. et al (2019). "Fake News" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. *American Behavioral Scientist*. <https://doi.org/10.1177/0002764219878224>

³¹ Zhou X. & Zafarani R. (2018). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *Researchgate.net*

in context, timing etc. Then, it is possible to resort to unstructured approaches based for example on the attitude of a document, meaning whether it goes to support or contradict the claim, or on its content or context, such as style, language, network of propagation. The result of all these procedures is correction, which may consist in reporting falsehood, give further details or share the revision.³²

In more detail, there are at least four types of approaches used to recognise *fake news*, based respectively on knowledge, style, propagation and source. The first one involves the use of a technique known as *fact-checking*, which can be carried out both by individuals in the flesh and through technology.³³

First of all, it should be pointed out that there is a difference between this term and that of verification: according to Bill Adair³⁴, the latter is the editorial technique used by journalists to verify the accuracy of a statement³⁵, while fact checking is a particular application of this technique. This procedure may be carried out either by experts or by ordinary people. Professional fact-checking involves the work of experts in different areas of knowledge, who time by time verify the news according to their experience. The result of this methodology will be surely very accurate, also because they may take measures themselves, such as make corrections, give warnings or directly censure the content in question. Their reliability is also due to the fact that they are a limited number of individuals; yet, this could also represent a limit in situations in which the news to be verified is in high quantity.³⁶ Nevertheless, as this type of approach has become increasingly used over time, expert-based fact-checking websites are becoming more and more common. One of the best known and detailed is *PolitiFact*, which from time to time focuses on a topic, evaluates all statements about it and provides credibility statistics.³⁷ There is even an association, the *International Fact-Checking Network*, which unites all the *fact-checkers* of the world and is responsible for monitoring trends, offering trainings, organizing programs and partnerships as well as conferences. In addition, it promotes the guidelines

³² Graves L. (2018). Understanding the Promise and Limits of Automated Fact-Checking. *Reuters Institute For The Study Of Journalism*. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves_factsheet_180226%20FINAL.pdf

³³ Zhou X. & Zafarani R. (2018). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *Researchgate.net*

³⁴ Founder of the Pulitzer Prize-winning website *PolitiFact*

³⁵ Silverman C., (n.d.). Verification and Fact Checking. *DataJournalism.com* <https://datajournalism.com/read/handbook/verification-1/additional-materials/verification-and-fact-checking>

³⁶ Collins B. et al. (2020). Fake News Types and Detection Models on Social Media A State-of-the-Art Survey. In: Sitek P. et al. (eds) *Intelligent Information and Database Systems. ACIIDS 2020. Communications in Computer and Information Science*, vol 1178. Springer, Singapore. https://doi.org/10.1007/978-981-15-3380-8_49

³⁷ Zhou X. & Zafarani R. (2018). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *Researchgate.net*

in the *International Fact-Checking Network Code of Principles*,³⁸ which is addressed to all those organizations specifically created for the purpose of fact-checking and that steadily fulfil this task with regard to declarations made by persons and institutions who play a leading role. According to it, such organizations must comply with certain principles, divided into further criteria. They must avoid any preferential treatment but, on the contrary, operate according to a principle of fairness; they shall ensure that anyone is able to follow their same verification procedures, and for this reason provide details about their sources, the followed methodology, their financing and their organizational structure; they disclose corrections, making sure that readers see them.³⁹

A different type of fact-checking approach relies on the effectiveness of the “wisdom of the crowds”, based on some studies, such as the one carried out by Pennycook G. and Rand D. G.⁴⁰

They divided 60 news websites into mainstream, hyperpartisan and known-for-spreading-fake-news, finding out that the accuracy of laypeople’s judgement was very close to that of specialists in the sector. This approach consisted in the use of groups of individuals, identifiable through marketplaces like *Amazon Mechanical Turk*, who carry out the above-mentioned news audit work. There are several disadvantages and limitations in using this strategy. First, as already corroborate, individuals cannot be totally reliable, both because they are easily manipulable and because of the bias they carry. However, there are concrete working examples, such as the site *Fiskkit*, in which users categorize articles and assess sentences that compose it.⁴¹ Nonetheless, most of the studies show that, from this point of view, human work cannot be considered effective.

In addition, as the amount of content to be checked tends to infinity, it has been tried to develop tools capable of automatically fulfilling this task, exploiting the advances made in the field of Artificial Intelligence (AI). These approaches employ in particular three techniques: *Information Retrieval* (IR), *Natural Language Processing* (NLP), and *Machine Learning* (ML) or network/graph theories. Most of knowledge-based approaches that adopt this perspective make use of a unified standard representation whose definitions are set out below. *Knowledge is composed by SPO, meaning triples that summarise the main content of the information and are made up of entities (Subject, Predicate and Object); triple verified as truth is a fact, and a set of facts is called Knowledge Base (KB).*

³⁸ The international Fact-Checking Network (2020). *Poynter.org* <https://www.poynter.org/ifcn>

³⁹ Commit to transparency — sign up for the International Fact-Checking Network’s code of principles (n.d.). *IFCN Code Of Principles*. <https://www.ifncodeofprinciples.poynter.org/>

⁴⁰ Pennycook G., Rand D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences* 116 (7) 2521-2526; DOI: 10.1073/pnas.1806781116

⁴¹ Zhou X. & Zafarani R. (2018). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *Researchgate.net*

*Knowledge Graph (KG) is the graphic representation of triples in a knowledge base, where entities are represented as nodes and relationships (predicates) are represented as edges.*⁴²

Therefore, the first thing to do is to identify these facts, either from a single-source or an open one: the first type of extraction will be more efficient, the second more complete. Once such facts have been obtained, they must be adjusted to form KB: for example, it is needed to eliminate repetitions or to check that reported information has not changed over time. The obtained triples must then be compared with true knowledge.⁴³ SPO is therefore one of the often-considered inputs that requires natural language understanding, because any type of declaration shall be summarised in this triplet. Different approaches, however, choose to focus on textual claims which can be numerical claims, entity and event properties, position statements, quote verification.

Other variables that influence the output are the evidences employed. Some methodologies do not use any beyond the input itself, therefore the veracity is dictated only by the way the statement is presented. Others make use of factors which, while not constituting genuine evidence, contribute to the accuracy of the classification, such as the creator of the news or the channel through which it has been propagated. From the point of view of actual evidences, through the KG is possible either to recognize the component in favour/against the claim or forecast its probability to be true. Obviously, this graph is limited because, on the one hand, it is assumed to include true facts relevant to the claim, on the other it cannot be taken for granted that something will not happen just because it is considered improbable.

A distinct methodology consists in relying on information arising from texts such as verified news, articles or journals. In this case, it is possible to consider different information: single phrases or the whole document; data coming from several documents in combination, as in the case of the Fact Extraction and Verification (FEVER); findings deriving from previously checked assertions.

An additional alternative is that of evaluating the reliability of news by analysing the behaviour of users about it, involving interactions, distribution etc. The simplest evaluation technique would then include only two outcomes: true or false, though in natural language adherence to reality includes different nuances. For this reason, some systems use a scale with different judgments, others employ ordinal classifications and so on.

Finally, automated fact checking can use different methods, most of which are supervised. Those involving evidence are *verification*, which responds to the fact that the evidence found is not always true, and the techniques to trace the implicit *common-sense reasoning* behind each statement. As

⁴² *Ibidem*

⁴³ *Ibidem*

involves *language detection*, there are systems able to track certain words that lead to classify the news either as *subjective* or *emotional*, for instance ‘reportedly’ or ‘liberal’, or as *deceptive*, mostly pessimistic words. Two other types of *detection* regard *rumours* and *clickbait*. The first one is an uncertain news becoming viral on social networks, typically characterised by language subjectivity; however, it is not taken for granted that the fact of being a rumour implies the falseness of a fact. The other one considers only the title and does not use any evidence. The last approach regards the *speaker profile*: one technique involves verifying whether a statement fits the creator, another one uses metadata such as affiliation, age and gender.⁴⁴

The above-mentioned method relying on style, on the other hand, assesses the veracity of news based on the intent which it was written with. To recognize it, it is necessary to consider the union of textual and visual features that allow to distinguish fake news from real ones. From the textual point of view, there are general features that consider lexicon, syntax discourse and semantic, and latent features, used for news text embedding. The employed approach can be either the one of traditional Machine Learning, where general features are manually selected and verified through supervised, semi-supervised or unsupervised models, or the Deep Learning one. In any case, a strong feature of this approach is that it allows to identify fake news before they spread; however, those who want to encourage the propagation can intervene by simply altering the style.

From a different perspective, in order to understand how news spreads, it is possible to rely on direct methods, such as *news cascade*, or indirect representation, like *self-defined graph*. News cascade is a tree structure where the root node is the initiator of the distribution. Self-defined graph is a method employing homogenous, heterogeneous or hierarchical networks. Different studies used both these approaches to compare fake and real news. Results showed that the first propagates faster, expands more and therefore has a greater potential to become viral,⁴⁵ besides producing more shares, engagement and dense networks.⁴⁶ In addition, it seems that these features refer to fake news regarding the political sphere more than those concerning any other subject.⁴⁷ One weakness of this methodology is that it allows to detect false news only after they have spread; nevertheless, it is more reliable than the previously-mentioned one.

⁴⁴ Thorne J., Vlachos A. (2018). Automated Fact Checking: Task formulations, methods and future directions. *Association for Computational Linguistics* <https://www.aclweb.org/anthology/C18-1283>

⁴⁵ *Ibidem*.

⁴⁶ Zhou X. & Zafrani R. (2019). Networkbased Fake News Detection: A Pattern-driven Approach. *CoRR*, arXiv:1906.04210

⁴⁷ Vosoughi Roy D., Aral S. (2018). The spread of true and false news online. *Science*. 359. 1146-1151. 10.1126/science.aap9559.

A further factor through which assess what can be read online, is the credibility of the sources: reference is made by who created and those who published it, as well as those involved in the distribution. The first two categories, in particular, would seem to tend to form homogeneous connections according to which they can be placed in certain groups. Those who have written more than two authentic stories are indeed defined true-news authors; in direct opposition there are fake-news authors; a third group then includes creators of both true and false stories. The same degree of homogeneity is found in the analysis of publishers. In this case, however, there are even more groups: Russian/conspiracy community; right-wing/conspiracy community; U.S. mainstream community; left-wing blog community; U.K. mainstream community.⁴⁸

Connected to the reliability of authors is the credibility of the site where news is published. As analysed in the first chapter, when searching for information via search engine, people tend to consult only the first results. For this reason, website owners use techniques aimed at inclusion in the first SERP⁴⁹: content spam, meaning the use of certain words; outcoming or incoming link spam; cloaking and redirection; click spam. Certain algorithms can help in identifying the use of such strategies. Through content-based spam detection, it was found for instance that, in these pages, URLs have multiple punctuation marks and digits, their content is duplicated and changes rapidly. Link-based algorithms recognize spam using graph information, trust propagation etc; other algorithms rely on click streams or user behaviour.⁵⁰

Finally, a further element that may prove to be crucial is the credibility of users who spread news, since they can do so with the intent of deceiving or not. In the first case, they are defined malicious users, which are the aforementioned *bot*; normal users instead can be vulnerable to fake news.⁵¹ The assumption that the opinions of users are to some extent a factor on which to rely, is at the base of a study consisting in aggregating the opinions of users on news, based on their engagement with it on social media.⁵² In particular, it analyses two tweets regarding the fake news of the alleged support of the Pope towards Trump, already mentioned in the first chapter. One tweet was shared with the aim

⁴⁸ Zhou X. & Zafarani R. (2018). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *Researchgate.net*

⁴⁹ *Search Engine Page*

⁵⁰ Zhou X. & Zafarani R. (2018). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *Researchgate.net*

⁵¹ *Ibidem.*

⁵² Yang S., et al. (2019). Unsupervised Fake News Detection on Social Media: A Generative Approach. *Proceedings of the AAAI Conference on Artificial Intelligence*. 33. 5644-5651. 10.1609/aaai.v33i01.33015644.

of denouncing its falsehood by a user who probably was particularly able in detecting untrue news; the other individual could have been deceived by the story or could have shared it while aware of its inaccuracy. In any case, it is very likely that those who came across the first tweet interpreted it as false, contrary to those who read the one shared by the other person. The authors of this study specify being aware of the lack of reliability of people's opinions: they often contrast with each other and are easily manipulable, especially since fake news is created with this intention. Nonetheless, in their opinion, the link created by news on social network are too complicate to be correctly evaluated by truth discovery algorithms, in particular, those based on the most common supervised approaches. In contrast to these, the method used by them starts from the premise that each tweet is characterised by its content and the engagement it generates. For reasons of simplification, only verified users' shares are considered in the study: the reason is that their popularity, on the one hand, generates much more interactions, on the other is supposed to be a sign of credibility and therefore ability in discerning truth from falsehood. Using truthfulness and credibility as latent variables, they create the algorithm; then, after some adjustments, they employ two well-known fake news datasets, *LIAR* and *BuzzFeed News*. Finally, when comparing it with benchmark algorithms that do not take into account users' engagement, results show that their algorithm has a better performativity.

2.3 Attempts to combat *fake news*

As already mentioned, disinformation is a threat that is fuelled and originates from many sides, hence verification, whether manual or automatic, is not sufficient. Action must be taken against those responsible, in order to discourage such practice. Nevertheless, regulating the world of information could be a dangerous task since it can be extremely easy to cross the line and result in censorship.

A practice that has become increasingly widespread, in fact, is that of punishing journalists who write stories which are authentic but unfavourable to those in power, by calling them *fake news*. Countries such as Philippines established by law the imprisonment of five years for those who are accused of distributing or publishing it. There is a great risk that strong powers, especially the most authoritarian

ones, will increasingly abuse it to their advantage. For this reason, too, it is necessary to act and provide answers on multiple levels.⁵³

As a matter of fact, some of the solutions with a non-regulatory character already proposed over time by various stakeholders, have been analysed by the European Commission in a 2018 report⁵⁴, although it should be noted that effects of some of them have not yet been fully assessed. These initiatives are first and foremost divided into positive and negative, based on their impact on fundamental rights, freedom of expression being the first. Subsequently, three subdivisions are identified among the good ones, founded respectively on transparency, trust-enhancement, and media and information literacy.

Basically, it is about actions that can be taken in favour of a greater awareness by all those actors who might be facing false news. From this point of view, among the most important actors there certainly are online platforms, although it should be pointed out that not all of them have given the same kind of response. In any case, such platforms can exercise their power on some specific factors: checking accounts and possibly deactivate the illegitimate ones; working on algorithms' reliability; disincentive the possible profit coming from untrue news; collaborate with other sources that have the same objective.

One of the most powerful tools they have at their disposal is the already mentioned advertising: in this context, platforms can ensure the reliability of the websites before allowing advertising to be published and consequently prevent it from being recognised by disinformation spreaders.

Also, from the monetary point of view, another fundamental action is to ensure the transparency of sponsorships. As an example, for some time now Instagram has made it mandatory for public figures to clearly report cases in which their posts are the result of a paid partnership. In the same way, since users read a large amount of news on social platforms, these should establish codes of conduct, whose compliance must be verified by the reader. In order to assure this, in fact, press organizations are oriented towards the use of credibility indexes.

Another delicate aspect is that several studies have shown that fake news has more influence on people's opinions than their corrections. On a psychological level, it is difficult to eliminate information one used to believe to be true, especially if there is no immediate alternative that

⁵³ West D. M. (2017). How to combat fake news and disinformation. *Center for Technology Innovation*
<https://www.brookings.edu/research/how-to-combat-fake-news-and-disinformation/>

⁵⁴ European commission (2018). A multi-dimensional approach to disinformation, Report of the independent High level Group on fake news and online disinformation. *Communications Network, Content and Technology* In Final report of the High level Group on fake news and Online Disinformation <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>

contradicts it. It is therefore important that the media that wants to spread the correction of a false story, would do so immediately. At the same time, to correct an information, the previous version must be repeated, and this is not positive considered that the mere exposure to an information increases the possibility of believing in it. This possibility, then, increases more when the person who receives the information already has a negative attitude towards the subject matter. As a result, trying to spread the correction of a story could be counterproductive.⁵⁵ However, a high degree of collaboration among all the actors involved is essential, and not only for the sake of truth. Disinformation is indeed one of the worst evils of modern society, because it has a strong impact on real life not only of individuals, but also of business and public authorities; briefly of society as a whole.⁵⁶

3. CASE STUDY: COVID-19 INFODEMIC

3.1 COVID-19 pandemic

Statistics show that *fake news* can spread six times faster than real news. It is also well known that during a time of crisis, fear drives people to irrationality. This irrationality pushes individuals to believe anything and consequently to spread news without the required attention and objectivity. Therefore, it can be recognized how this process is intensified when the crisis is caused by an uncontrollable and mostly little-known event.⁵⁷ This is precisely what is currently happening around the world, because of the pandemic caused by the new coronavirus strain SARS-CoV-2. Between the end of 2019 and the beginning of 2020 news about this novel virus and its circulation in China began to diffuse. Over time, it has started to spread more and more quickly throughout the whole world, thanks also to the globalization and the ease and speed of movements that it implicates. The number of countries involved, and infected individuals began to grow dramatically. As a

⁵⁵ Tsfati Y. et al. (2020). Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis. *Annals of the International Communication Association*, 44:2, 157-173, DOI: 10.1080/23808985.2020.1759443

⁵⁶ Meschi M., Eastwood D., Kanabar R. (2020). The real-world effects of ‘fake news’- and how to quantify them. *Tech Law for Everyone, SCL*. <https://www.scl.org/articles/12022-the-real-world-effects-of-fake-news-and-how-to-quantify-them>

⁵⁷ Kapetanovic A. (2020). “INFODEMIC – spreading fake news is almost as dangerous as spreading the virus”. *Regional Cooperation Council* <https://www.rcc.int/docs/507/majlinda-bregu-secretary-general-of-the-rcc-on-1st-edition-of-balkathon-and-future-plans-in-an-interview-for-albanian-business-magazine>

consequence, before acquiring real awareness of the characterization of this new virus, the World Health Organization declared that a pandemic was in progress.

To the paralysing fear that man, by his nature, has always felt towards the unknown, was added an extraordinary rate of the virus' speed, together with the inevitable lack of certainty from experts around the world. The result could only be an unprecedented panic. Citizens around the world have grown in great need of information about the symptoms, the real consequences, the actual severity, the behaviour to be adopted. The phenomenon that has been generated has been summed up by the WHO Director-General Tedros Adhanom Ghebreyesus in the term *infodemic*: an overabundance of pandemic-related information which put the citizens' confidence and stability to the test at a time when they need them most.⁵⁸ Based on the general engagement, expressed by the individual comments and outbursts taking place on social networks, many argue that this is the first real social media *infodemic*.⁵⁹ the Associate Director of the *IFCN* Cristina Tardàguila, has defined it “the biggest challenge fact-checkers have ever faced”.⁶⁰

3.2 The role of social networks

As previously analysed, social network has become one of the most popular tools in the fruition of news. The paradox is that people have relied on their ability to provide a large amount of information, but the resulting overabundance has ended up being “misleading, unsettling and confusing for a largely uninformed public”.⁶¹ Individual conduct is reasonably strictly dependent on the information that everyone is offered, therefore disinformation has had and continues having very serious consequences, on individual lives as well as on the general development of the pandemic. An

⁵⁸ 1st WHO Infodemiology Conference (2020). *World Health Organization. who.int* <https://www.who.int/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference>

⁵⁹ Ahmad A.R., Murad H.R. (2020). The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. *J Med Internet Res* 22(5):e19556 10.2196/19556

⁶⁰ Brennen J. S. et al. (2020). Types, Sources, and Claims of COVID-19 Misinformation. *Reuters Institute for the study of journalism* <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>

⁶¹ Ashrafi-Rizi H, Kazempour Z. (2020). Information Typology in Coronavirus (COVID-19) Crisis; a Commentary. *Arch Acad Emerg Med.* 8(1):e19. PMID: 32185370; PMCID: PMC7075270.

example is what happened in the Republic of Iran: the news that high-proof alcohol would kill the virus has led over 300 people to die and more than 1000 to get sick.⁶²

Unfortunately, the amount of dangerous fake news online is not only about baseless treatments, but also figures, causes, prevention, conspiracies. For this reason, although a relatively short period of time has elapsed, several datasets for COVID-19 misinformation have been published. Some have focused on users, others on content, conspiracy theories, bots. One among them, for instance, characterised COVID-19 misinformation communities by analysing tweets' content and interactions.⁶³ In particular, the author identified seventeen categories in which such tweets could be categorised, on the basis of which he distinguished two communities: one sharing content such as conspiracies or false treatment, and the other sharing corrections, true public health responses and so on. After deleting bots' tweets, he carried out a linguistic analysis through a program based on word counts, identifying precise lexical categories. In order to detect the use of narratives within the two communities, he starts from some assumptions: narrative structure is usually characterized by the use of function words, authenticity, low analytical thinking. Surprisingly, he found out that in the COVID-19 case informed users are those employing a more emotional and authentic language. This is an understandable result, because it indicates the use of personal and family stories with the aim of fighting conspiracies and false news. Then, he considered tone, formality and uncertainty. Resulting analysis showed a certain homogeneity in the negativity of both groups' tone, but some differences in the other two categories. In fact, misinformed users employ a more informal language (fillers, netspeak, swear words) and seem to be more confident; in addition, an upward polarisation and a higher presence of bots among them can be detected.⁶⁴ Another element to be considered can be the use of authority references, which is usually absent in misinformation news. It is instead very likely that articles written in order to correct misleading information provide further evidence of what they claim, which can derive from fact-checking organizations, institutions, and so on. For this reason, this element should be one of those considered by users in assessing whether to believe in a news item.⁶⁵

⁶² Baines D. & Robert Elliott R. J. R. (2020). Defining misinformation, disinformation and malinformation: An urgent need for clarity during the COVID-19 infodemic. Discussion Papers 20-06. *Department of Economics, University of Birmingham*.
RePEc:bir:birmec:20-06

⁶³ Memon S. A. (2020). Characterizing Misinformed Online Health Communities. 10.1184/R1/12988244.v1

⁶⁴ *Ibidem*

⁶⁵ McGlynn, J., Baryshevtsev M., Dayton Z. A. (2020). Misinformation more likely to use non-specific authority references: Twitter analysis of two COVID-19 myths. *The Harvard Kennedy School (HKS) Misinformation Review*

3.3 Novel coronavirus-related *fake news*

A further typology of analysis⁶⁶ aims at understanding which statements, types and sources of bad information prevail within the framework of this novel coronavirus. It is carried out on the basis of 225 *fake news* detected until the end of March by *First Draft News* fact-checkers combined with the *IFCN*⁶⁷ and *Google Fact Checking Tools*, and the relative interactions of individuals on social media (platforms; TV; new outlets; websites). Clearly, as previously stated, there are limits to the work that fact checkers can do: they must select certain types of information, they cannot access the spread of *fake news* via private channels, nor consider all existing channels among public ones. In this unquantifiable number of incorrect information on the COVID-19, the majority appears to be not completely false, but rather reframed; these manipulations can be explained on the basis of the aforementioned categories, identified by Claire Wardle.⁶⁸ A *misleading content*, for example, can be that of a tweet that recommended washing hands, an act whose usefulness has been certified, but at the same time supported the uselessness of hand sanitisers, as they would kill bacteria and not viruses. This post has been very shared for several reasons that have increased its credibility. First of all, it contains true information familiar to most people, such as that of washing hands or the difference between virus and bacteria. Furthermore, the author wrote the tweet calling himself “a tired scientist”, an expression which contributes to an increasing credibility of what has been said for two reasons. The first one is the alleged role of an expert in the field, while the second the exasperated tone of those who know the truth but have to see everyone behaving in the wrong way.⁶⁹ An example of *false context* could be instead that of a circulating photo which exposed the difference between a full shelf of vegan food and the others emptied, to emphasize that despite the critical situation no one wanted to buy that particular food. The photo was later found to be from another time and place than the one mentioned.

<https://misinforeview.hks.harvard.edu/article/misinformation-more-likely-to-use-non-specific-authority-references-twitter-analysis-of-two-covid-19-myths/>

⁶⁶ Brennen J. S. et al. (2020). Types, Sources, and Claims of COVID-19 Misinformation. *Reuters Institute for the study of journalism*
<https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>

⁶⁷ *International Fact-Checking Network*

⁶⁸ Wardle C. (2019). Understanding Information Disorder. *Firstdraftnews.com*.

https://firstdraftnews.org/wpcontent/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x76701

⁶⁹ Oremus W. (2020). The Simplest Way to Spot Coronavirus Misinformation on Social Media. *Onezero.medium.com*
<https://onezero.medium.com/the-simplest-way-to-spot-coronavirus-misinformation-on-social-media-4b7995448071>

Another situation highlighted within this analysis is that in which celebrities, although in a small number, shared incorrect information, inevitably generating an excessive resonance. On this theme, however, there is a certain degree of disagreement, as this analysis claims that celebrities and politicians are responsible for sharing 20% of misleading content ⁷⁰, while a *Reuters Institute for the Study of Journalism and the Oxford Internet Institute* report shows of a percentage equal to 69.

However, an example can be traced in particular to *fake news* that connects the spread of the virus to 5G. The British singer MIA published a video of people setting fire to fibre broadband installed near her house, claiming that “they should just turn it off ‘till the pandemic”. The British boxer Amir Khan argued on Instagram that coronavirus was man-made, and the lockdown was a cover to complete 5G technology ⁷¹.

Even more serious, however, is the role that some leading politicians have played in this respect. The BBC has produced a summary report in this regard. According to it, for instance, Donald Trump initially downplayed the severity of the virus, and then called it a pandemic after a few weeks; successively, he tried to trace the increase in demand for masks to a series of alleged thefts to the detriment of NY hospitals. Several Chinese figures, such as the foreign ministry spokesman Zhano Lijian, denied that the origin of the virus was their country, accusing the United States or Italy. With respect to the latter, the role of *fake news* spreader is attributable to Matteo Salvini, who publicly claimed that the virus was created in a Chinese laboratory. In Brazil, the amount of false information shared by the President Jair Bolsonaro was so vast that he has been censored by all major social platforms: Facebook, Instagram, Twitter, Youtube. Also, in India the entire world witnessed Subramanian Swamy declare that the strain present in his country was less virulent.⁷²

In terms of claims, the analysis carried out by Brennen et al.⁷³ has classified as the most common ones, primarily those concerning the solutions undertaken by the authorities: from local governments to international ones, such as the UN. Next, there are those concerning the spread of the virus geographically, but especially ethnically, with the intention of targeting greater faults to

⁷⁰ Brennen J. S. et al. (2020). Types, Sources, and Claims of COVID-19 Misinformation. *Reuters Institute for the study of journalism* <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>

⁷¹ Waterson J. (2020). Influencers among ‘key distributors’ of coronavirus misinformation. *The Guardian* <https://www.theguardian.com/media/2020/apr/08/influencers-being-key-distributors-of-coronavirus-fake-news>

⁷² Morris C. (2020). Coronavirus: False claims by politicians debunked. *BBC News*

⁷³ Brennen J. S. et al. (2020). Types, Sources, and Claims of COVID-19 Misinformation. *Reuters Institute for the study of journalism* <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>

determined categories. One of the best known international “blame-game”⁷⁴ is that between China and USA. One of the many mutual accusations, is the one made by the U.S. Secretary of State Mike Pompeo, who claimed that SARS-CoV-2 would have spread due to the little regulation of the laboratory in Wuhan, in which the first cases were confirmed. On the contrary, China has pointed the finger at the U.S. administration for its inability to deal with the situation and superficiality of the early days. But the false reports about China’s responsibility have been so numerous and so widespread that a strong racial resentment has begun to raise towards ordinary Chinese citizens. The consequences were physical and verbal assaults both to the detriment of ordinary people and doctors, as well as large losses affecting establishments of Chinese origin, from restaurants to shops, whose owners were forced to display on windows signs in which it was specified that they had not returned to their homeland for years. Such episodes have seen as subjects also Italians, as well as all those originating from countries in which the virus has gone gradually spreading. Some countries have also taken advantage of this situation to isolate minorities that governments ostracized. It is the case of the Muslim inhabitants of India: dehumanised, arrested, threatened, accused of carrying out a "COVID jihad".⁷⁵ This type of racist action, together with false reports which aim to portray democratic institutions as incapable of dealing with the situation, continues to proliferate in a coordinated way everywhere in the globe, and each country uses it to strike its main enemy. In Africa real hate campaigns against certain ethnic groups are taking place; Europe is the target of Syria that attacks its sanctions, as well as of Turkey and the Western Balkans, that insert the pandemic into the narrative that sees Europe turning its back on them.⁷⁶

Among the various studies undertaken to try to stem the seriousness of the phenomenon of disinformation in the context of the current pandemic, one linked to the *Institute of Public Health, College of Medicine and Health Sciences* of the United Arab Emirates University, rely on the usage of *Google Trends*. This tool allows both to evaluate the trends of the moment, and the changes of interest towards some specific themes. Their analysis focused in particular on the period from January 21st, 2020, to March 24th, 2020 in Italy. In detail, this tool can calculate the popularity of all the keywords that people search on Google, allowing a comparison of five groups of terms at one time.

⁷⁴ Rahn W. (2020). “Does US-China coronavirus blame game threaten scientific investigation?”. *Deutsche Welle*
<https://p.dw.com/p/3bx1P>

⁷⁵ Bezzi D. (2020). Arundhati Roy: Indian Muslims facing ‘genocidal climate’ amid pandemic. *openDemocracy.net*
<https://www.opendemocracy.net/en/arundhati-roy-indian-muslims-facing-genocidal-climate-amid-pandemic/>

⁷⁶ Eeas special report update: short assessment of narratives and disinformation around the covid-19 pandemic (2020). *EUvdSiSiNFO*
<https://euvsdisinfo.eu/eeas-special-report-update-short-assessment-of-narratives-and-disinformation-around-the-covid-19-pandemic/>

Results showed that the five trending terms related to *infodemic* were “novel coronavirus,” “China coronavirus,” “COVID-19,” “2019-nCoV,” and “SARS-COV-2”; the health-related were “face masks”, “amuchina” (hand-sanitizing gel), “symptoms of the novel coronavirus”, “health bulletins”, “vaccines for coronavirus”. Another finding was that in regions such as Campania, Umbria and Basilicata news related to the aforementioned racism against Chinese have been broadcasted.⁷⁷

3.4 Multidimensional responses

Given the scale of this *infodemic* and the seriousness of the consequences of *fake news* related to coronavirus, all the actors involved are collaborating more than ever to curb the phenomenon. As stated before, actions taken in response to misleading information have always involved the whole society on several levels. In this case, however, a more heavy and incisive response has taken place. This is due to the severity of the triggering cause and to its main and most frightening characteristic: that of affecting all citizens of the world, without distinction. Answers indeed come from the most important international organizations, as well as local governments, online platforms, private initiatives.

3.4.1 Individual awareness

To become such, news needs an audience and, for it to acquire notoriety, it is necessary that this audience share it and initiates large-scale dissemination. Even in this situation, therefore, the first solution must be sought in the awareness of the individual. For this to happen, however, people must be equipped with the necessary tools which allow them to identify false news. The channels through which most information passes must therefore be controlled. The aforementioned fact-checking organisations have coordinated within the *International Fact Checking Network*, to which all the actors operating on other levels have taken part, in some cases even contributing with donations.

⁷⁷ Rovetta A., Bhagavathula A. S. (2020). COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. *JMIR Public Health Surveill.* 6(2):e19374. doi: 10.2196/19374. PMID: 32338613; PMCID: PMC7202310

Between January and March, *IFCN* was able to detect and deny over 1,500 COVID-19-related online falsehoods, relying both on manual and automatic fact checking methodologies.⁷⁸

Being a topic on which not even the experts have too many certainties, however, some criticalities have occurred. For example, some have called false the information that masks are a sufficient protection, thing that turns out to be contradictory to the official provisions given by authorities.

3.4.2 International commitment

On the international level, the leading player is certainly the World Health Organisation because, although the current crisis is affecting all areas, the underlying concern is primarily related to science. First of all, WHO created the *Information Network for Epidemics*, which brings together technical and social media teams in identifying and responding to myths and *fake news* with counterblows corresponding to reality.⁷⁹ Moreover, from June 29 to July 21, 2020 it organised the *1st WHO Infodemiology Conference*, which also included experts in public health, data science, technology applications, media studied, marketing and so on. The aim was indeed to address the situation from a multidisciplinary point of view, so as to establish a common agenda in which to invest.⁸⁰ Among participants there was also another international actor: the United Nations. UN team is trying also to spread correct information on all channels at its disposal, such as radio and social networks. Then, UNESCO has created a rubric named “misinformation shredder”, in which there are content in local languages, as well as two policy briefs supported by the *International Center for Journalists*, which help journalists in publishing accurate information. UNESCO is then particularly engaged in initiatives concerning developing countries, which already have an environment at risk. For instance, they keep in contact with some Eastern African journalists in order to exchange information; the same

⁷⁸ Posetti J., Bontcheva K. (2020). DISINFODEMIC: Deciphering COVID-19 disinformation. *UNESCO.org*
<https://en.unesco.org/covid19/disinfodemic/brief1>

⁷⁹ Department of Global Communications (2020). 5 ways the UN is fighting ‘infodemic’ of misinformation. *unitednations.org*
<https://www.un.org/en/un-coronavirus-communications-team/five-ways-united-nations-fighting-%E2%80%98infodemic%E2%80%99-misinformation>

⁸⁰ 1st WHO Infodemiology Conference (2020). *World Health Organization. who.int* <https://www.who.int/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference>

thing occurs with a network of 25 radio communities that manage to reach the most marginalized communities.⁸¹

The European Union is also particularly committed to the cause: it tries to favour cooperation in detecting, preventing and countering attacks; it promotes the use of authoritative sources, as well as the online platforms' commitment, while supporting fact-checkers and researchers. It also creates the *EUvsDisinfo* project, for the purpose of make citizens more aware.⁸²

All these initiatives aim to mobilise the online community and create partnerships and campaigns.

3.4.3 Policy actions

It is easy to understand that the current one is a *fake news* epidemic that has no precedent, so the usual kind of answers may not be enough. From NGOs to think tanks⁸³, indeed, some have attempted to provide investigative responses that result in real disinformation campaigns and involve data that usually the mere fact checking does not consider. The actions that can be taken by politics are crucial, considered the power it has. It can play a decisive role in disseminating correct information, as well as in discouraging false news, through incentives and sanctions. In the first case, for instance, countries such as Canada, Belgium, Italy are engaging in media tax relief and funds in support of their category. On the other end, several States' responses consisted in providing themselves with the necessary powers to punish both publishers and sharers of false information.⁸⁴ At the same time, however, it is necessary to maintain a free and independent environment. In general, all governments should be at the forefront of strategic and transparent communication.⁸⁵ For this reason, clearly and directly conveying information may not be sufficient. A fundamental element is

⁸¹ Department of Global Communications (2020). 5 ways the UN is fighting 'infodemic' of misinformation. *unitednations.org*
<https://www.un.org/en/un-coronavirus-communications-team/five-ways-united-nations-fighting-%E2%80%98infodemic%E2%80%99-misinformation>

⁸² European Council (2020), Fighting disinformation. *consilium.europa.eu*
<https://www.consilium.europa.eu/en/policies/coronavirus/fighting-disinformation/>

⁸³ *Ibidem*

⁸⁴ *Ibidem*

⁸⁵ OECD Open Government Unit (2020). Transparency, communication and trust: The role of public communication in responding to the wave of disinformation about the new coronavirus. Tackling coronavirus (COVID-19): contributing to a global effort. *OECD*
https://read.oecd-ilibrary.org/view/?ref=135_135220-cvba4lq3ru&title=Transparency-communication-and-trust-The-role-of-public-communication-in-responding-to-the-wave-of-disinformation-about-the-new-coronavirus

to separate public and political communication, given the strong polarization that exists especially in some countries. For it to be effective, then, communication must not only be transparent, but also continuous. In many governments, such as Italy, there have been daily press meetings and conferences aiming to update and inform citizens about the new developments in both political and health fields. From this point of view, it is also very important to assess the degree of trust of the individual states' citizens towards politicians: if it is low, it would be more useful to entrust the dissemination of information only to scientists. Another measure taken to facilitate the direct transmission of information, has been the creation of channels and *chatbots* on messaging platforms, as well as dedicated telephone lines.

3.4.4 Platforms answers

It is now abundantly clear that social network, news media, search engines are the main information spreaders, therefore their responses could be decisive, considering also that their action can be undertaken on different levels. First, they have a degree of control over the content that is published under their domain. Nonetheless, as already stated, such aspect presents some criticisms when this power collides with censorship. In this aspect, it is explanatory the war that Trump has filed against Twitter, guilty of adding a disclaimer to a tweet of his which contained false accusations against democrats.

As regards the misinformation about SARS-CoV-2, this platform has adopted a new policy that provides to ban all tweets containing content that does not correspond to the truth or falsely attributed to authorities, while trying to verify reliable profiles.

Facebook, which refused to ban lies in political ads, ensures it instead for those trying to make a profit on non-existent treatments. Furthermore, it is assisting people in finding information through a *Coronavirus Information Center*; helping government health organizations in answering to frequently asked questions; providing citizens with a section in which to ask for or offer some kind of help to their neighbours.

As for the giant Google, every coronavirus-related article is accompanied by an "SOS Alert" banner, together with information coming from recognised organisations. In addition, it provides a database in which it is possible to find fact checked articles, as well as the main areas affected by the pandemic. Instagram has adopted measures like those of Facebook, introducing a hashtag that redirects directly to the *Centers for Disease Control and Prevention* website. All filters linked to the virus have been

banned, except those agreed with health organisations, like the one that encourages people to stay home.

Youtube announced that it will ban videos regarding the 5G theory.

Whatsapp is instead one of those platforms less facilitated because of its encrypted structure, but still it is employing AI in order to shut down spam accounts, as well as prevent forwarding the same message too many times.⁸⁶

As previously said, in content moderation there is a fine line between the protection of truth and the suppression of freedom of expression. In this *infodemic* situation, in particular, human rights and media observer are concerned about the implications of government behaviour. There are in fact examples of disproportionate reactions, such as the arrest of some users in Cambodia and Indonesia. Even more serious is the episode of the Egyptian journalist or the Chinese ophthalmologist. The first one criticized the government and was arrested under the pretext of a conviction for spreading *fake news*; while he was still in custody, he died for contracting the virus. The doctor instead was admonished for spreading rumours about the virus, only to receive an apology after he died because of COVID-19.

Furthermore, as the global director of research at the *International Center for Journalists* Julie Posetti said, the efficiency of the laws that many countries are approving to stem the phenomenon, has not even been tested. What is certain, however, is the high probability that fundamental rights will be damaged. In her opinion, instead, much more should be done to provide accurate, timely and consistent information.⁸⁷

In addition to content moderation, however, platforms have the possibility to resort to the use of those AI's tools previously debated. Those employed during the pandemic rely, in addition to the verification of language, also on that of visuals, geolocation, advertising transparency. In this regard, however, given the particular extent of the phenomenon, some of the most important platforms have found themselves forced to warn users that they will no longer be able to guarantee a second review made by a human being, in case of appeal.

⁸⁶ Skopeliti C., John B. (2020). Coronavirus: How are the social media platforms responding to the 'infodemic'? *Firstdraftnews.org*
<https://firstdraftnews.org/latest/how-social-media-platforms-are-responding-to-the-coronavirus-infodemic/>

⁸⁷ Hand J. (2020). 'Fake news' laws, privacy & free speech on trial: Government overreach in the infodemic? *Firstdraftnews.org*
<https://firstdraftnews.org/latest/fake-news-laws-privacy-free-speech-on-trial-government-overreach-in-the-infodemic/>

Finally, from an economic perspective, the first type of action would be to eliminate and demonetize advertising on virus issues; allowing now only those of governments and authorities, the limit is that it is also prevented from giving visibility to those intent on information campaigns.⁸⁸

CONCLUSION

The digital revolution has undeniably had a profound impact on every aspect of society. One of the most affected areas is certainly that of information. Networks connecting the world through the Internet have made a new revolutionary rapid way of communicating possible. In fact, a whole series of real-world barriers have been overcome thus stimulating the consumption of news. Information circulation has indeed become much simpler and more immediate, and mostly costs-free. Consuming information online is advantageous not only from the economic point of view. Compared to the past, in fact, there is a considerable saving also in terms of time and effort. However, like any innovation, it has entailed negative aspects. As a matter of fact, the quality of information has been affected by the opportunity for anyone to produce it. In addition, it has become possible to spread content in an infinitesimal time lapse, reaching simultaneously a wide audience in various places in the world. This lowering of quality combined with the rate of transmission have made disinformation one of the most serious problems of recent years.

The first two chapters presented the main features characterising *fake news*, which have been further illustrated through a brief analysis of the *infodemic* phenomenon. In the current situation of global pandemic, inaccurate information of all kinds has spread faster than the virus itself, causing particularly serious consequences. It is therefore crucial to acquire awareness of the ways and forms in which *fake news* can present itself, in order to be able to identify it.

It has been discussed how news can be written with different intentions: irony, profit, provocation, simple will to deceive. Being able to distinguish those categories is essential to approach online information in a beneficial way. Because of the overabundance of information by which users are surrounded, the loss of the ability to judge is understandable. For this reason, interconnections and development of new technologies, especially those related to Artificial Intelligence, represent pivotal

⁸⁸ Posetti J., Bontcheva K. (2020). DISINFODEMIC: Deciphering COVID-19 disinformation. *UNESCO.org*
<https://en.unesco.org/covid19/disinfodemic/brief1>

tools during the verification process. This proves that the possibilities offered by technology, and especially by the Internet, are endlessly providing modern society with all the means necessary to make a right use of them. Finally, they are proven to play a fundamental role during this current period of uncertainty caused by the ongoing global pandemic.

The reflection that this thesis sought to convey starts from the consideration that *fake news* is not a recent phenomenon; what has changed are the dynamics that characterise the news sector. Their continuous evolution is in fact increasingly directed towards a less attention to professionalism and reflection. Instead, the growing need to express opinions and the will to do so in the immediate are privileged.

Given the dimensions of the phenomenon and the space in which it performs, simple censorship would be impossible and counterproductive. Any attempt to answer must be based on transparency and cooperation, but in order to win this battle, it will not be enough to intervene on the news itself. On the contrary, it will be necessary to start with culture and a re-education of citizens based on awareness and empowerment.

ABSTRACT

L'avvento delle nuove tecnologie digitali e la loro diffusione nella società ha gradualmente reso possibile quella che da molte voci è stata definita come una vera e propria rivoluzione basata sull'informazione. I dati procurati dalle continue interazioni di milioni di individui online, in particolare, si sono rivelati come il motore di una nuova economia e hanno guadagnato il potere di influenzare l'intera società. Gli sviluppi più recenti delle tecnologie digitali vanno tutti nella direzione di ricerca di metodi e tecniche che possano ancora ampliare la quantità dei dati prodotti per poter sfruttare la conoscenza che ne deriva rendendola la principale fonte di potere economico. Uno dei settori maggiormente colpiti da tale rivoluzione è sicuramente quello delle notizie. Il Web ha eliminato tutta una serie di ostacoli che tradizionalmente hanno caratterizzato la produzione e fruizione di notizie: barriere geografiche, temporali, economiche. Di conseguenza è diventato possibile a tutti non solo creare notizie, ma assumere contemporaneamente i ruoli di produttore e distributore. Questo però ha anche fatto sì che il lettore si sia ritrovato sommerso da un'enorme quantità di notizie, spesso anche senza cercarle. La mancanza di ostacoli ha rivelato i propri risvolti negativi: dal lato dell'offerta si è verificato un crollo della qualità dovuto alla mancanza di rispetto

delle norme sia tecniche che etiche; dal lato della domanda, si è persa la capacità di giudizio. Questi due processi insieme hanno portato alla nascita delle *fake news*, appellativo con cui si definiscono le notizie inventate, ingannevoli o distorte. Data la diffusione su larga scala del fenomeno, tale definizione viene usata in realtà per descrivere informazioni talvolta anche molto diverse tra loro. Una classificazione molto utile nella comprensione delle diverse sfumature esistenti, è quella operata dall'organizzazione nonprofit *First Draft*. Per prima cosa, bisogna distinguere tre tipi di “disturbo dell'informazione”: *disinformation* implica la volontà di ingannare e confondere il lettore; *misinformation* descrive un caso di condivisione consapevole di una notizia falsa; *malinformation* è invece una notizia vera estrapolata dal proprio contesto di appartenenza. Sulla base di queste tre tipologie, la classificazione in questione individua poi l'esistenza di ulteriori sottocategorie. È possibile per esempio che il contenuto di una notizia sia stato totalmente inventato, oppure che sia soltanto leggermente manipolato. O ancora, che l'elemento che ne determina la falsità sia un'errata interpretazione dell'intento con le quali sono state scritte. Le ragioni possono essere infatti molteplici: la volontà di ironizzare su un avvenimento, provocare, esercitare un'influenza politica, fare propaganda, creare un profitto. In ogni caso, qualunque sia la ragione dietro la creazione delle *fake news*, ciò che le rende preoccupanti è la loro diffusione, che dimostra dinamiche nuove e allarmanti e sembra avvenga in gran parte accidentalmente ad opera di utenti comuni. Per evitarlo è infatti necessario comprendere e analizzare i metodi di individuazione di questo tipo di notizie. Data la pericolosità e diffusione del fenomeno, infatti, essi sono stati oggetto di numerosi studi. Ricerche recenti si sono focalizzate sulle caratteristiche che accomunano gran parte delle notizie di questo tipo. Ad esempio, gli aspetti sintattici e lessicali, errori di punteggiatura, tentativi di emulare siti web noti attraverso piccole modifiche nell'URL, mancanza di citazioni. Tutte caratteristiche che, se considerate singolarmente potrebbero non avere alcuna implicazione sulla natura della notizia, perché simili errori possono verificarsi anche nella pubblicazione di notizie vere, ma se valutate e individuate correttamente, possono decisamente aiutare l'utente nell'individuazione delle *fake news*. Data l'enorme quantità di informazioni fuorvianti presenti sul web, l'intervento umano non può però essere sufficiente. A questo proposito, l'apporto della tecnologia è fondamentale. Negli anni, infatti, i progressi fatti nel campo dell'Intelligenza Artificiale sono stati applicati a questo problema. Il risultato è stata la creazione di un gran numero di algoritmi che si sostituiscono all'uomo nelle fasi di identificazione, verifica e correzione. In particolare, questi sistemi utilizzano tre tecniche: recupero delle informazioni; elaborazione del linguaggio naturale; apprendimento automatico. Attraverso queste tecniche poi, possono esserci diversi approcci alla verifica: si può scegliere come elemento determinante il contenuto stesso della notizia, il suo stile, la sua diffusione o ancora la credibilità della

fonte che lo ha pubblicato. L'automatizzazione di tali processi ha ovviamente portato enormi benefici alla rapidità del processo; tuttavia, non mancano errori dovuti al fatto che esistono delle sfumature individuabili soltanto dalla capacità umana. Data la delicatezza e serietà del fenomeno della disinformazione, tentativi di risposta sono stati dati da attori operanti su svariati livelli. Un caso esemplificativo di questa compattezza di azione è la situazione in cui si trova attualmente il mondo intero. Dall'inizio del 2020, infatti, l'intera umanità si è trovata costretta a combattere una duplice battaglia: quella della pandemia da SARS-CoV-2 e quella dell'*infodemia* che ha generato. Il volume di informazioni circolanti sul Web ha raggiunto proporzioni mai viste nel corso della storia dell'umanità, così come la serietà delle conseguenze. Le *fake news* collegate al COVID-19 riguardano infatti tutti gli ambiti coinvolti, andando dalle cospirazioni, agli articoli che annunciano false cure fai-da-te. Anche gli attori colpevoli della loro distribuzione non sono più soltanto utenti vulnerabili o con cattive intenzioni, ma si annoverano tra essi anche celebrità e politici di tutto il mondo. La portata del fenomeno ha raggiunto livelli tali da generare la reazione di diversi attori, come organizzazioni internazionali, governi nazionali, piattaforme online e gli stessi individui, che unendo le proprie forze hanno cercato di contrastare il fenomeno. La speranza è che tale unità d'azione, unita alla capacità di sfruttare al meglio la tecnologia e i suoi progressi, possano generare una vittoria nella battaglia contro un fenomeno sempre più pervasivo e dai rivolti quanto meno minacciosi.

BIBLIOGRAPHY and SITOGRAPHY

1st WHO Infodemiology Conference (2020). *World Health Organization*. *who.int*
<https://www.who.int/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference>

Ahmad A.R., Murad H.R. (2020). The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. *J Med Internet Res* 22(5):e19556
10.2196/19556

Alaimo C. & Kallinikos J. (2017). Computing the everyday: Social media as data platforms. *The information Society*, 33(4),175-191

Ashrafi-Rizi H, Kazempour Z. (2020). Information Typology in Coronavirus (COVID-19) Crisis; a Commentary. *Arch Acad Emerg Med*. 8(1):e19. PMID: 32185370; PMCID: PMC7075270.

Baines D. & Robert Elliott R. J. R. (2020.) Defining misinformation, disinformation and malinformation: An urgent need for clarity during the COVID-19 infodemic. Discussion Papers 20-06. *Department of Economics, University of Birmingham*. RePEc:bir:birrec:20-06

Bezzi D. (2020). Arundhati Roy: Indian Muslims facing ‘genocidal climate’ amid pandemic. *openDemocracy.net* <https://www.opendemocracy.net/en/arundhati-roy-indian-muslims-facing-genocidal-climate-amid-pandemic/>

Brennen J. S. et al. (2020). Types, Sources, and Claims of COVID-19 Misinformation. *Reuters Institute for the study of journalism* <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>

Collins B. et al. (2020). Fake News Types and Detection Models on Social Media A State-of-the-Art Survey. In: Sitek P. et al. (eds) Intelligent Information and Database Systems. *ACIIDS 2020. Communications in Computer and Information Science*, vol 1178. Springer, Singapore. https://doi.org/10.1007/978-981-15-3380-8_49

Commit to transparency — sign up for the International Fact-Checking Network's code of principles (n.d.). *IFCN Code Of Principles*. <https://www.ifcncodeofprinciples.poynter.org/>

Department of Global Communications (2020). 5 ways the UN is fighting 'infodemic' of misinformation. *Unitednations.org* <https://www.un.org/en/un-coronavirus-communications-team/five-ways-united-nations-fighting-%E2%80%98infodemic%E2%80%99-misinformation>

Eeas special report update: short assessment of narratives and disinformation around the covid-19 pandemic (2020). *EUvdSiSiNFO*. <https://euvsdisinfo.eu/eeas-special-report-update-short-assessment-of-narratives-and-disinformation-around-the-covid-19-pandemic/>

European commission (2018). A multi-dimensional approach to disinformation, Report of the independent High level Group on fake news and online disinformation. *Communications Network, Content and Technology* In Final report of the High level Group on fake news and Online Disinformation <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>

European Council (2020). Fighting disinformation. *consilium.europa.eu* <https://www.consilium.europa.eu/en/policies/coronavirus/fighting-disinformation/>

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28 (1), 62-70. <https://doi.org/10.1016/j.infoandorg.2018.02.005>

Fischer P., Schulz-Hardt S., Frey D. (2008). Selective exposure and information quantity: how different information quantities moderate decision makers' preference for consistent and inconsistent information. *J Pers Soc Psychol.* 94(2):231-44. doi: 10.1037/0022-3514.94.2.94.2.231. PMID: 18211174.

Graves L. (2018). Understanding the Promise and Limits of Automated Fact-Checking. *Reuters Institute For The Study Of Journalism*.
https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves_factsheet_180226%20FINAL.pdf

Gray J., Bounegru L., Venturini T. (2020). ‘Fake news’ as infrastructural uncanny. *New Media & Society*. 22(2):317-341. doi:10.1177/1461444819856912

Giglietto F., et al. (2019). ‘Fake news’ is the invention of a liar: How false information circulates within the hybrid news system. *Current Sociology*. 67(4):625-642. doi:10.1177/0011392119837536

Habgood-Coote J. (2018). The term “fake news” is doing great harm. In *theconversation.com*.
<https://theconversation.com/the-term-fake-news-is-doing-great-harm-100406>

Hameleers M., et al. (2020). A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication*. 37:2, 281-301, DOI: 10.1080/10584609.2019.1674979

Hand J. (2020). ‘Fake news’ laws, privacy & free speech on trial: Government overreach in the infodemic? *Firstdraftnews.org* <https://firstdraftnews.org/latest/fake-news-laws-privacy-free-speech-on-trial-government-overreach-in-the-infodemic/>

Introna L. & Nissenbaum H. (2000). Shaping The Web: Why The Politics Of Search Engines Matters. *The Information Society*. 16. 169-185. 10.1080/01972240050133634.

Kapetanovic A. (2020). “INFODEMIC – spreading fake news is almost as dangerous as spreading the virus”. *Regional Cooperation Council* <https://www.rcc.int/docs/507/majlinda-bregu-secretary->

general-of-the-rcc-on-1st-edition-of-balkathon-and-future-plans-in-an-interview-for-albanian-business-magazine

Kapoor K., Dwivedi Y. K., & Piercy, N. C. (2016). Pay-per-click advertising: A literature review. *Marketing Review*, 16(2), 183-202. <https://doi.org/10.1362/146934716X14636478977557>

Martens B. et al. (2018). The digital transformation of news media and the rise of disinformation and fake news. *JRC Technical Reports*. JRC 111529

McGlynn, J., Baryshevtsev M., Dayton Z. A. (2020). Misinformation more likely to use non-specific authority references: Twitter analysis of two COVID-19 myths. *The Harvard Kennedy School (HKS) Misinformation Review* <https://misinforeview.hks.harvard.edu/article/misinformation-more-likely-to-use-non-specific-authority-references-twitter-analysis-of-two-covid-19-myths/>

Melford C., Fagan C. (2019). Cutting the Funding of Disinformation: The Ad-Tech Solution. *Disinformationindex.org*.
https://disinformationindex.org/wp-content/uploads/2019/05/GDI_Report_Screen_AW2.pdf

Memon S. A. (2020). Characterizing Misinformed Online Health Communities. 10.1184/R1/12988244.v1

Meschi M., Eastwood D., Kanabar R. (2020). The real-world effects of ‘fake news’- and how to quantify them. *Tech Law for Everyone, SCL*. <https://www.scl.org/articles/12022-the-real-world-effects-of-fake-news-and-how-to-quantify-them>

Metaxa D., & Torres-Echeverry N., (2017). Google's Role in Spreading Fake News and Misinformation. Available at SSRN: <https://ssrn.com/abstract=3062984> or <http://dx.doi.org/10.2139/ssrn.3062984>

Molina M. D. et al (2019). “Fake News” Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. *American Behavioral Scientist*. <https://doi.org/10.1177/0002764219878224>

Morris C. (2020). Coronavirus: False claims by politicians debunked. *BBC News* <https://www.bbc.com/news/av/52299689>

Newman N. et al. (2017). Reuters Institute Digital News Report 2017. *Reuters Institute for the Study of Journalism*. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf

OECD Open Government Unit (2020). Transparency, communication and trust: The role of public communication in responding to the wave of disinformation about the new coronavirus. Tackling coronavirus (COVID-19): contributing to a global effort. *OECD* https://read.oecd-ilibrary.org/view/?ref=135_135220-cvba4lq3ru&title=Transparency-communication-and-trust-The-role-of-public-communication-in-responding-to-the-wave-of-disinformation-about-the-new-coronavirus

Oremus W. (2020). The Simplest Way to Spot Coronavirus Misinformation on Social Media. *onezero.medium.com* <https://onezero.medium.com/the-simplest-way-to-spot-coronavirus-misinformation-on-social-media-4b7995448071>

Pennycook G., Rand D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences* 116 (7) 2521-2526; DOI: 10.1073/pnas.1806781116

Posetti J., Bontcheva K. (2020). DISINFODEMIC: Deciphering COVID-19 disinformation. UNESCO.org <https://en.unesco.org/covid19/disinfodemic/brief1>

Rahn W. (2020). Does US-China coronavirus blame game threaten scientific investigation?. *Deutsche Welle* <https://p.dw.com/p/3bx1P>

Rovetta A., Bhagavathula A. S. (2020). COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. *JMIR Public Health Surveill.* 6(2):e19374. doi: 10.2196/19374. PMID: 32338613; PMCID: PMC7202310

Shao C. et al. (2018). The spread of low-credibility content by social bots. *Nature Communications.* 9. 10.1038/s41467-018-06930-7.

Shao C. at al. (2017). The spread of fake news by social bots. *Researchgate.net.* https://www.researchgate.net/publication/318671211_The_spread_of_fake_news_by_social_bots/citations

Silverman C., (n.d.). Verification and Fact Checking. *DataJournalism.com* <https://datajournalism.com/read/handbook/verification-1/additional-materials/verification-and-fact-checking>

Skopeliti C., John B. (2020). Coronavirus: How are the social media platforms responding to the ‘infodemic’? *firstdraftnews.org* <https://firstdraftnews.org/latest/how-social-media-platforms-are-responding-to-the-coronavirus-infodemic/>

The international Fact-Checking Network (2020). *poynter.org* <https://www.poynter.org/ifcn/>

Thorne J., Vlachos A. (2018). Automated Fact Checking: Task formulations, methods and future directions. *Association for Computational Linguistics* <https://www.aclweb.org/anthology/C18-1283>

Thorson E. (2016). Belief Echoes: The Persistent Effects of Corrected Misinformation. *Political Communication*, 33:3, 460-480, DOI: 10.1080/10584609.2015.1102187

Three challenges for the web, according to its inventor (2017). *Web Foundation* <https://webfoundation.org/2017/03/web-turns-28-letter/>

Tsfati Y. et al. (2020). Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis. *Annals of the International Communication Association*, 44:2, 157-173, DOI: 10.1080/23808985.2020.1759443

Vosoughi Roy D., Aral S. (2018). The spread of true and false news online. *Science*. 359. 1146-1151. 10.1126/science.aap9559

Wang P., Angarita R., Renna I. (2018). Is this the Era of Misinformation yet: Combining Social Bots and Fake News to Deceive the Masses. In Companion Proceedings of the The Web Conference 2018 (WWW '18). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1557–1561. DOI:<https://doi.org/10.1145/3184558.3191610>

Wardle C. (2019). Understanding Information Disorder. *firstdraftnews.org*.
https://firstdraftnews.org/wpcontent/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x76701

Waterson J. (2020). Influencers among ‘key distributors’ of coronavirus misinformation. *The Guardian* <https://www.theguardian.com/media/2020/apr/08/influencers-being-key-distributors-of-coronavirus-fake-news>

West D. M. (2017). How to combat fake news and disinformation. *Center for Technology Innovation*
<https://www.brookings.edu/research/how-to-combat-fake-news-and-disinformation/>

Why We Fall for Fake News. *Centre for Information Technology and Society*.
<https://www.cits.ucsb.edu/fake-news/why-we-fall> [Ultimo accesso: 18 settembre 2020]

Yang S., et al. (2019). Unsupervised Fake News Detection on Social Media: A Generative Approach. *Proceedings of the AAAI Conference on Artificial Intelligence*. 33. 5644-5651.
10.1609/aaai.v33i01.33015644

Zhou X. & Zafarani R. (2018). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *researchgate.net*.
https://www.researchgate.net/publication/329388190_A_Survey_of_Fake_News_Fundamental_Theories_Detection_Methods_and_Opportunities

Zhou X. & Zafrani R. (2019). Networkbased Fake News Detection: A Pattern-driven Approach, *CoRR*_arXiv:1906.04210