



---

## **Misinformation: A Threat to Digital Governance**

**By Ludovica Vecchio**



## Publication Details

**Published:** December 4, 2020

**Publisher:** Invictus Corporation Ltd.

**Department:** Emerging Technologies and Law

**Author:** Ludovica Vecchio

**Editors:** Ajatshatru Bhattacharya, Harshita Bhattacharya

© INVICTUS CORPORATION LTD. and the author 2020

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Produced, published and distributed by INVICTUS CORPORATION LTD., The Hague, The Netherlands

*Website:* <https://invictuscorp.org/>

The registered company address is: New World Campus, Spaarneplein 2, 2515 VK The Hague, The Netherlands

---



## Abstract

---

At least 800 people have died globally as a direct consequence of misinformation about COVID-19.<sup>1</sup> In Hungary, at the same time, the new *Enabling Act* has empowered the government to detain and jail anyone for up to 5 years for spreading misinformation about the COVID-19.<sup>2</sup> These two cases represent the extreme consequences of different approaches adopted by governments to tackle misinformation. On the one hand, a liberal approach threatening the right to life; on the other, a rigorous regime, threatening freedom of expression. What is the best approach? As it is often the case, the answer lies somewhere in the middle.

Whilst strict policing to ensure the truthfulness of the information may be *prima facie* desirable, freedom of expression is a defining element of all democratic societies. Governments should strive to balance the preservation of human rights and civil liberties, and the increasingly pressing necessity to contain misinformation. The latter can only be successfully achieved via a joint venture with tech companies, in an effort to deploy emerging technologies as a means to combat fake news. Therefore, this publication will be dedicated to evaluating whether States can indeed rely on tech companies to tackle misinformation and still comply with their fundamental rights obligations.

---

<sup>1</sup> Alistair Coleman, "'Hundreds dead' because of Covid-19 misinformation" *BBC* (12 August 2020) <https://www.bbc.com/news/world-53755067>.

<sup>2</sup> Vlagyiszlav Maksimov, 'Hungary's state of emergency law is a 'blank cheque' to Orban, critics say' *Euractiv* (23 March 2020) <https://www.euractiv.com/section/justice-home-affairs/news/hungarys-state-of-emergency-law-is-a-blank-cheque-to-orban-critics-say/>.



## Table of Contents

---

1. Introduction .....	5
2. Interplay between Freedom of Information and Misinformation .....	11
2.1. Freedom of Information .....	11
2.1.1. History .....	11
2.1.2. Definition .....	13
2.2. Fake News .....	14
3. Governmental Efforts against Misinformation .....	18
3.1. Legislative Efforts to Combat Misinformation .....	19
3.2. Balancing Freedom of Expression and Anti-Misinformation Measures .....	23
3.3. Existing Solutions .....	25
3.4. Digital Literacy .....	26
4. Technological Innovations to Combat Misinformation .....	28
4.1. Automated Fact-Checking .....	30
4.2. Fact-checking on Facebook and Twitter .....	32
4.2.1. Facebook .....	32
4.2.2. Twitter .....	34
4.3. The Shortcomings of Automated Fact-Checking .....	35
5. Conclusion .....	38
6. Bibliography .....	40



# 1. Introduction

---

*'Disinformation can cost lives'*

Ursula von der Leyen, President of the European Commission<sup>3</sup>

The aforementioned statement was made in relation to the cases of the People's Republic of China and the Russian Federation using *'the global coronavirus crisis to spread false reports and other online disinformation'*.<sup>4</sup> For instance, Russia stated that the European Union (EU) was *'on the verge of collapse'* due to its national governments' inadequate responses to the ongoing pandemic.<sup>5</sup> It should be noted that the dangers of misinformation go far beyond the pandemic. Take for instance, the recent presidential elections in the United States of America (US).<sup>6</sup> In October 2020, Facebook and Instagram reported taking down over 150 million posts containing false political information.<sup>7</sup> The main subject of such misinformation has been mail-in ballots which threatened election participation and now threatens *'the legitimacy of the outcome of the election across the political spectrum'*.<sup>8</sup> Experts state that with the advancement of technology, we have reached an unprecedented proliferation of misinformation.<sup>9</sup> This particular example showcases how misinformation has been destabilising *'the effective functioning of the democratic process'*.<sup>10</sup> It undermines the peoples' trust in the government; it questions the sanctity of freedom of speech, the defining characteristic of democracy; it overall threatens the very foundations of a democratic society. Moreover, the potential damage of misinformation extends beyond democratic processes to one's life: reportedly, at least 800 people have died globally due to misinformation about the COVID-19.<sup>11</sup>

The phenomenon of a rise in "fake news" originated from the significant changes that the information society has undergone in the past two decades. The traditional journalism industry is experiencing a crisis as digitalisation advances *'citizen journalism'*, which takes place largely on social media.<sup>12</sup> There

---

<sup>3</sup> Mark Scott, 'Russia and China push "fake news" aimed at weakening Europe: report' (*Politico*, 1 April 2020) <https://www.politico.eu/article/russia-china-disinformation-coronavirus-covid19-facebook-google/>.

<sup>4</sup> *ibid.*

<sup>5</sup> *ibid.*

<sup>6</sup> 3 November 2020.

<sup>7</sup> Agence France-Press, 'Facebook says it rejected 2.2m ads for breaking political campaigning rules' (*The Guardian*, 18 October 2020) <https://www.theguardian.com/technology/2020/oct/18/facebook-says-it-rejected-22m-ads-seeking-to-obstruct-voting-in-us-election>.

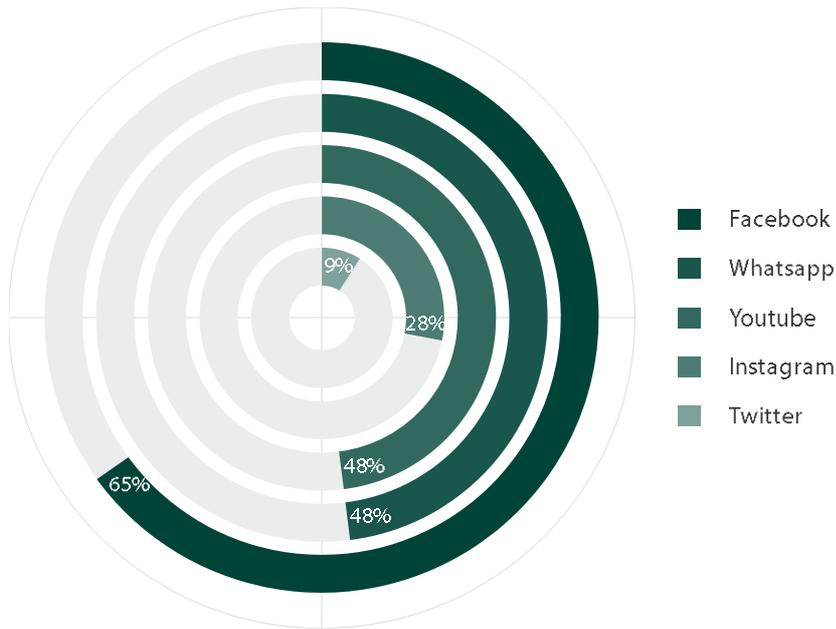
<sup>8</sup> Yochai Benkler et al, 'Mail-In Voter Fraud: Anatomy of a Disinformation Campaign' (*Berkman Klein Center*, 1 October 2020) <https://cyber.harvard.edu/publication/2020/Mail-in-Voter-Fraud-Disinformation-2020>.

<sup>9</sup> Davey Alba and Joseph Plambeck, 'What You Need to Know About the Election and Disinformation' (*NYT*, 16 November 2020) <https://www.nytimes.com/2020/11/16/us/election-disinformation.html>.

<sup>10</sup> Philip Napoli, *Social Media and the Public Interest* (Columbia University Press, 2019) 81.

<sup>11</sup> Alistair Coleman, "'Hundreds dead" because of Covid-19 misinformation' *BBC* (12 August 2020) <https://www.bbc.com/news/world-53755067>.

<sup>12</sup> Napoli (n 10) 91.



Graph 1: Percentage of people using social platforms

are 3.5 billion people online, approximately half the world population, and about 70% of these people are Facebook users.<sup>13</sup> This data showcases the widespread and pervasive nature of social media which is further appreciable by considering the graph above which outlines the percentage of users of the most “popular” social media, namely Facebook, Whatsapp, Youtube, Instagram, and Twitter.

On the one hand, social networking services facilitate civic participation, providing users, *viz.* citizens, with a platform where to voice their opinion, ultimately creating a more diverse and accessible public debate.<sup>14</sup> On the other, the phenomenon of ‘*digital journalism*’ has undermined what many consider genuine journalism to the extent that it has been described as ‘*parasitic*’.<sup>15</sup> This shift in trend is attributable *inter alia* to the high costs of traditional journalism as opposed to those of digital journalism. Analysts contend that the print newspaper business has been in long-term decline, considering that numerous newspaper providers have experienced a year-on-year decline in sales.<sup>16</sup> In contrast, an online subscription model significantly reduces production costs and improves accessibility.<sup>17</sup>

<sup>13</sup> Esteban Ortiz-Ospina, ‘The rise of social media’ (*Our world in data*, 18 September 2019) <https://ourworldindata.org/rise-of-social-media>.

<sup>14</sup> Patrick Leerssen, ‘Cut Out by the Middle Man: The Free Speech Implications of Social Network Blocking and Banning in the EU’ (2015) 6 *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 99.

<sup>15</sup> Napoli (n 10) 91.

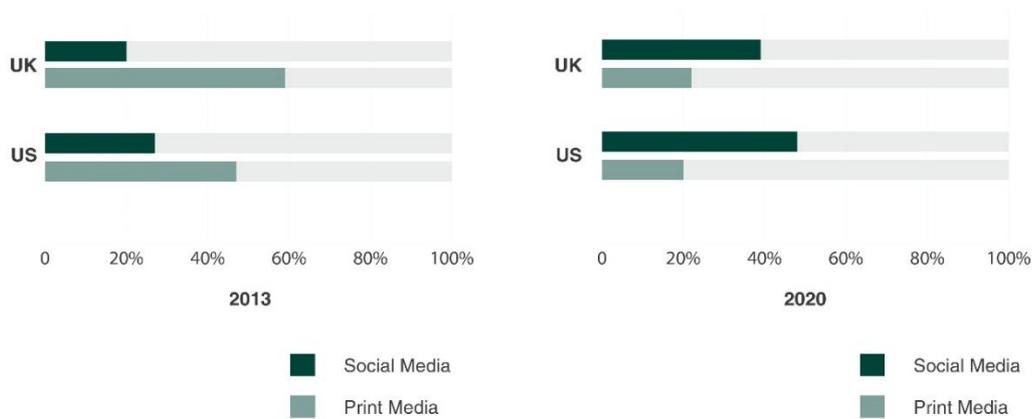
<sup>16</sup> BBC, ‘ABC figures: Newspapers will no longer have to publish sales’ *BBC* (21 May 2020)

<https://www.bbc.com/news/entertainment-arts-52754762..>

<sup>17</sup> *ibid.*



Shift in reliance on social media for news vs print media from 2013 to 2020 in the US and UK

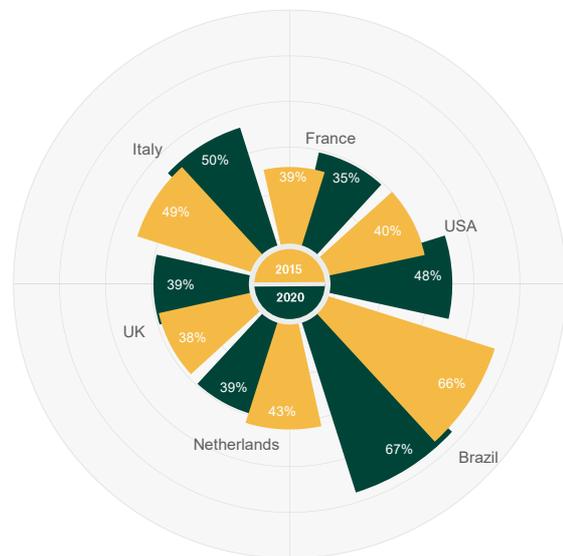


Graph 2: Traditional print media compared to social media for news

More outstandingly, ‘social media have surpassed print newspapers as a news source’: 20% of US adults reportedly resort to social media for news as opposed to 16% turning to newspapers instead.<sup>18</sup> Similarly, in the United Kingdom (UK), reliance on print media as a source of news has decreased from 59% to 22% between 2013 and 2020, with social media gaining ground: from 20% to 39%.<sup>19</sup>

However, whilst social media have rendered ‘the sharing of information ubiquitous and nonstop’, ‘where will that information come from in the first place when all the reporters have been let go?’<sup>20</sup>

The digitisation of the news industry via social media has significantly lowered the barriers of ‘verification and balance’ for entry to such industry and in particular, for the distribution of news.<sup>21</sup> As a result, we have experienced a rise in fake news. Even though there is no agreed-



Graph 3: People using social media for news as of January 2015 compared to 2020

<sup>18</sup> Elisa Shearer, ‘Social media outpaces print newspapers in the U.S. as a news source’ *Pew Research Center* (10 December 2018) <https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/>.

<sup>19</sup> Nic Newman et al, *Reuters Institute Digital News Report 2020* (Reuters Institute and University of Oxford, 2020).

<sup>20</sup> Douglas McLennan and Jack Miles, ‘A once unimaginable scenario: No more newspapers’ *Washington Post* (21 March 2018) <https://www.washingtonpost.com/news/worldpost/wp/2018/03/21/newspapers/?noredirect=on>.

<sup>21</sup> *ibid*; Francisco J. Pérez-Latre, Alfonso Sánchez-Tabernero, Gustavo García-Mansilla “Barriers to entry” in the media industry: the launching of El Español and Netflix’, University of Navarra (2-5 June 2016).



upon definition of this phenomenon, some defining characteristics are discernible. These include online distribution and strong presence on social media, the deliberateness of the action, i.e. *'intentionally or knowingly [making] false statements of fact'*, and its *'systemic dimension'*.<sup>22</sup> The latter refers to how the content is presented, i.e. in a way to appear as actual news, and it is what renders fake news appealing to the audience, and allows and facilitates the manipulation and persuasion of such audience.<sup>23</sup> In this regard, fake news is rather *'parasitic on traditional news media'*, typically mimicking their *'look and feel'*.<sup>24</sup> Emerging technologies have been a fundamental factor in this sense as they have allowed the sophistication of the systemic production of fake news.<sup>25</sup> The term fake news is also used as a claim to dismiss legitimate news.<sup>26</sup> Therefore, it amounts to *'a catch-all term with multiple definitions'*, *'a cluster concept'* characterised by *'a taxonomy of interrelated phenomena'*.<sup>27</sup> In particular, three ramifications have been identified: disinformation, misinformation, and malinformation, outlined in greater detail in Chapter 1. Overall, fake news amounts to one of *'the grand challenges'* of the 21<sup>st</sup> century representing one of the *'the most troubling'* effects of social media and algorithms on the *'media ecosystem'*.<sup>28</sup>

Considering the threats posed by fake news outlined thus far, it is in the best interest of governments to introduce some measure to control it. However, such control poses the risk of violating other rights, such as freedom of expression.<sup>29</sup> Therefore, it is necessary to enforce a checks and balances regime to preserve the competing interests. *Art 4.1 of the International Covenant on Civil and Political Rights* (ICCPR) lists freedom of expression among the derogable rights.<sup>30</sup> Therefore, in circumstances of absolute necessity, a State is not bound by it so long as it acts through proportionate and temporary measures. As such, if countries were to implement measures to tackle the fake news phenomenon, they would have to balance between the importance of freedom of expression and *'of the internet as a forum for public debate'* on the one hand, and *'the pressing social need'* for containing

---

<sup>22</sup> Axel Gelfert, 'Fake News: A Definition' (2018) 38(1) *Informal Logic* 84,96,109.

<sup>23</sup> *ibid.*

<sup>24</sup> *ibid* 91.

<sup>25</sup> Tarlach McGonagle, 'The development of freedom of expression and information within the UN: leaps and bounds or fits and starts?' in Tarlach McGonagle and Yvonne Donders (eds) *The United Nations and Freedom of Expression and Information* (CUP 2015).

<sup>26</sup> Gelfert (n 22).

<sup>27</sup> *ibid* 94.

<sup>28</sup> *ibid* 81.

<sup>29</sup> International Covenant on Civil and Political Rights (adopted on 16 December 1996, entered into force on 23 March 1976) 999 UNTS 171 (ICCPR) Art 19; European Convention on Human Rights (adopted 4 November 1950, entered into force 3 September 1953) ETS 5 (ECHR) Art 10.

<sup>30</sup> ICCPR Art 4.



misinformation on the other.<sup>31</sup> Any resulting restriction will have to be prescribed by law, and necessary in a democratic society.<sup>32</sup>

A significant hurdle in striking this balance is the aforementioned lack of an agreement on a common and comprehensive definition of fake news.<sup>33</sup> Moreover, given the significant involvement of tech companies in this matter, a legal definition of fake news is necessary to empower governments to hold them responsible. Indeed, human rights treaties do not have a direct effect on the legal relationships between private parties but rather on States and their citizens.<sup>34</sup> Therefore, private companies effectively operate in a human rights vacuum. Governments, accordingly, take advantage from this “loophole”, adopting *‘a soft power approach’* delegating onto tech companies the responsibility to *‘shape private policy indirectly’* through the so-called *‘regulation by raised eyebrow’*, *‘the invisible handshake’*.<sup>35</sup> The ultimate result is the *‘re-centralisation of power around (...) the social network service providers’*.<sup>36</sup>

Tech companies have thus been acquiring extensive powers and are *‘uniquely positioned to delimit the topics and set the tone of public debate’*, through *inter alia* content removal.<sup>37</sup> In this context, Mark Zuckerberg, Facebook Chief Executive Officer, has considered it necessary to clarify that his social network shall not be the *‘arbiter of truth of everything that people say online’*.<sup>38</sup> However, some may argue that whilst it is not such arbiter *de jure*, it definitely is *de facto*. Therefore, it is high time that governments regain the control of this increasingly concerning issue and embark upon a joint venture with tech companies, and society at large, in the containment of fake news.

This publication starts with outlining the notion of fake news. To that end, Chapter 2 will evaluate the merits of freedom of information, and of fake news. It will note how the former has been gaining greater prominence, whilst it originally was “just” the contiguity of freedom of expression. After framing the indefinability of fake news, the publication will highlight in more detail the issues this phenomenon has been causing. Continuing, Chapter 3 will consider governments’ responsibility to tackle fake news and the inherent human rights implications, especially in terms of freedom of expression. It will subsequently identify a possible solution to such implications, by resorting to

---

<sup>31</sup> ICCPR Art 19.3; ECHR Art 10.2; Leerssen (n 14) 102.

<sup>32</sup> ECHR Art 10.2.

<sup>33</sup> Gelfert (n 22).

<sup>34</sup> Leerssen (n 14) 101.

<sup>35</sup> *ibid* 108.

<sup>36</sup> *ibid* 99.

<sup>37</sup> *ibid* 100.

<sup>38</sup> Lisa Eadicco, ‘Satirical websites are testing Facebook’s policy on not being the ‘arbiter of truth’ by running false headlines claiming Mark Zuckerberg is dead or abusive’ (*Business Insider*, 29 May 2020) <https://www.businessinsider.com/fake-headlines-mark-zuckerberg-satire-test-facebook-policies-2020-5?IR=T>.



derogation and expanding the scope of freedom of information to include a right to receive truthful information. Finally, Chapter 4 will chart the tech-based initiatives which have been developed in light of the increasing need to automate fact-checking. The research will show that an adequate description of fake news becomes even more important when considering Ursula von der Leyen's claim that urges tech companies to be more active in tackling online disinformation, and their consequent use of Artificial Intelligence (AI) as a tracking tool and practice of content removal.<sup>39</sup> Indeed, whilst the automation of such a process is desirable as it may be immune from bias and efficient, the clearest instructions ought to be given to the machine in order to ensure fairness and human rights protection in the process. Holistically, this publication will outline the dynamics of a joint venture between national governments and tech companies and what obstacles it may encounter.

---

<sup>39</sup> *ibid.*



## 2. Interplay between Freedom of Information and Misinformation

---

News media generally amount to *'the source of much of what we take ourselves to know'*.<sup>40</sup> They provide the readers with a lens onto the world, which goes beyond their backyard, or their neighbourhood. Additionally, they can have a commercial function, facilitating purchases and consumerism through ads, and an entertainment function too.<sup>41</sup> Their pervasive nature thus renders them an ever-present element in our daily lives. However, it is of particular relevance to this publication to consider the impact that news media have had on two specific fundamental rights, namely, freedom of expression and information. News media indeed constitute a fertile soil for misinformation which has been proliferating, especially since 2016. This publication will evaluate the interplay between the containment of this phenomenon and the preservation of freedom of expression and information. To this end, this Chapter will define both freedom of information, and fake news, orderly.

### 2.1. Freedom of Information

The Internet of Things (IoT) and the spread of digitalisation have lowered the barriers to the online flow of information to an unprecedented level. As a result, social media networks have shifted from being gatekeepers of information to gateways for accessing external links and resources. They thus facilitate civic participation, and, consequently, civic information. Therefore, it is arguable that emerging technologies have been advancing freedom of information.

#### 2.1.1. History

The protection of freedom of information indicates the acknowledgement of *'the significance of people's sources of news'*.<sup>42</sup> At the international level, this right was initially granted to facilitate the cross-border flow of information, lowering the existing barriers between States.<sup>43</sup> Freedom of information was thus sanctioned under *Art 19.2 ICCPR*, and includes the *'freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice'*. At the national level, freedom of information

---

<sup>40</sup> Gelfert (n 22) 87.

<sup>41</sup> *ibid.*

<sup>42</sup> Erwin D Canham, 'International Freedom of Information' (1949) 14(4) *Law and Contemporary Problems* 584.

<sup>43</sup> *ibid.*



is commonly framed in narrower terms that is in terms of access to the information retained by public authorities.<sup>44</sup> For the purposes of this publication, however, I will follow the former approach.

It can be deduced from international provisions, such as the aforementioned *Art 19.2 ICCPR*, and *Art 10 European Convention on Human Rights* (ECHR), that freedom of expression and freedom of information are contiguous rights. In particular, the latter functions as a sort of extension of the former. Accordingly, when considering the relationship between the two, the United Nations (UN) Rapporteur on Freedom of Information, Salvador P Lopez defined freedom of information as ‘*little more than the aggregate of the more familiar antecedent principles of freedom of thought, freedom of expression and freedom of the press*’.<sup>45</sup> Freedom of expression is thus more frequently the object of legal claims, whilst freedom of information is more peripheral.<sup>46</sup>

This relatively marginal relevance of freedom of information can be further appreciated by considering the UN Conference on Freedom of Information, which resulted in, *inter alia*, the draft Convention on Freedom of Information.<sup>47</sup> Whilst the Conference was initially considered ‘*a great success*’, and the Convention ‘*the keystone*’ of its structure, the latter was never opened for signature and ratification.<sup>48</sup> The Convention eventually disappeared from the agenda of the UN General Assembly (UNGA) in the 1970s.<sup>49</sup> Additionally, the UN Conference on Freedom of Information was described as one of the most neglected UN chapters.<sup>50</sup> Similarly, in 1947, the UN Economic and Social Council (ECOSOC) had founded a Sub-Commission on freedom of information which was then disbanded four years later: ‘*a major setback*’ in UN history.<sup>51</sup>

This publication argues that irrespective of this rather daunting picture of general neglect for freedom of information, this right shall become increasingly relevant, and experience an expansion to include the freedom to “true” information. In order to conceive such a development, the following sub-section will outline a definition of freedom of information as it stands today.

---

<sup>44</sup> In the US, under the FOIA, any person has access to federal agency records and information, with 9 exemptions. In the UK, the right to access information held by public authorities is granted under the Freedom of Information Act 2000. In Germany, the *Gesetz zur Regelung des Zugangs zu Informationen des Bundes* (issued on 5 September 2005, last amended 7 August 2013) provides an unconditional right to access official federal information. In the Netherlands *Wet Openbaarheid van Bestuur*.

<sup>45</sup> UN ECOSOC, ‘Report by the Rapporteur on Freedom of Information S.P. Lopez 1953/12’ (1953).

<sup>46</sup> McGonagle (n 25).

<sup>47</sup> UN Draft Convention on Freedom of Information (adopted 13 February 1950) UN Doc E/RES/278(X).

<sup>48</sup> McGonagle (n 25) 11.

<sup>49</sup> *ibid.*

<sup>50</sup> *ibid* 10

<sup>51</sup> *ibid* 13.



### 2.1.2. Definition

Art 19.2 ICCPR describes freedom of information as entailing the freedom to both receive and impart information. The UNGA stated that *'freedom of information is (...) a touchstone of all the freedoms to which the UN is consecrated'*.<sup>52</sup> It is *'basic'* to democratic societies, ensuring the transparency and accountability of governments.<sup>53</sup> Overall, three main reasons for the importance of freedom of information are identifiable: it is *'an aspect of human dignity, (...) the best means of ascertaining the truth, and (...) a fundamental underpinning of democracy'*.<sup>54</sup>

Freedom of information generally entails the freedom to access information, and its violation occurs upon such information being withheld.<sup>55</sup> However, this publication argues that in the era of fake news, this right should include access to authentic information. The enforcement of such an extension shall be prudent as excessive regulation of the sort of information that can be shared may, in turn, violate freedom of expression, and legitimise censorship. Additionally, the regulation of the divulgence of information carries a significant risk of manipulation of such information which, in turn, would lead to the manipulation of the public. Whilst there have been instances of information manipulation for "good reasons", such as the deployment of media to advance peace and combat fascism during the Cold War, this manipulation has been increasingly entailing the spread of misinformation.<sup>56</sup>

Above we have noted how freedom of information largely entails access to government information.<sup>57</sup> In other words, freedom of information generally concerns governments' transparency. Emerging communication technologies have played a *'crucial instrumental role'* in its advancement.<sup>58</sup> Yet, by significantly increasing the speed of information distribution, they also worsened the issue of misinformation. Therefore, on the one hand, emerging technologies have had a positive impact on advancing democracy and freedom of information. On the other, they have fostered an increasingly worrying phenomenon. Indeed, misinformation has now attained such relevance that, as noted by Milanovic, it could aid the realisation cyberattacks.<sup>59</sup>

---

<sup>52</sup> UNGA Res 59/1946 (14 December 1946) UN Doc A/RES/59; UN Special Rapporteur Hussain also noted that lack of access to information undermines peoples' ability to exercise other freedoms: UN ECOSOC 'Question of The Human Rights Of All Persons Subjected To Any Form Of Detention Or Imprisonment' (14 December 1994) UN Doc E/CN.4/1995/32.

<sup>53</sup> Toby Mendel, 'Freedom of Information as an Internationally Protected Human Right' *Article 19* <https://www.article19.org/data/files/pdfs/publications/foi-as-an-international-right.pdf>; McGonagle (n 26).

<sup>54</sup> *ibid.*

<sup>55</sup> See Mendel (n 53); *Leander v Sweden* [1987] 9 EHRR 433; *Gaskin v UK* [1989] 12 EHRR 36; *Guerra v Italy* [1998] 26 EHRR 357.

<sup>56</sup> McGonagle (n 25) 12.

<sup>57</sup> Mendel (n 53).

<sup>58</sup> McGonagle (n 25) 4.

<sup>59</sup> Marko Milanovic and Michael N Schmitt, 'Cyber Attacks and Cyber (Mis)information Operations during a Pandemic' *EJIL: Talk!* (29 March 2020) <https://www.ejiltalk.org/cyber-attacks-and-cyber-misinformation-operations-during-a-pandemic/>.



We live in the so-called ‘post-truth’ era, where ‘objective facts are less influential in shaping public opinion than appeals to emotion and personal belief’.<sup>60</sup> ‘Truth is no longer dictated by authorities but networked by peers’.<sup>61</sup> Freedom of information is now arguably misconceived as the freedom to seek information from those news outlets which better align with one’s own views. The ultimate result is that individuals have become more ‘vulnerable to accepting and acting on misinformation’.<sup>62</sup> Therefore, it is paramount to regulate and contain it. This publication argues that this shall be done by, *inter alia*, extending the scope of the right to information to include a right to truthful information. However, as with any other issue, in order to contain it, it is fundamental to identify its boundaries first. The next section will thus consist of an attempt to define misinformation.

## 2.2. Fake News

The term “fake news” has become relatively common and frequent on social media, news outlets, and on the internet at large. Politicians even use it as a tool to discredit their opposition or news media organisations.<sup>63</sup> Overall, fake news is used ‘for managing the mass perception’ of a specific issue.<sup>64</sup>

The expression “fake news” gained global prominence only around 2016, in the wake of the US presidential elections, and the Brexit referendum.<sup>65</sup> In 2017, the UN Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples’ Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information issued the Joint Declaration on Freedom of Information and Fake News, Disinformation and Propaganda.<sup>66</sup> This soft-law document provides that States ought to tackle fake news, and outlines some General Principles and Guidelines which they should follow in doing so.<sup>67</sup> Whilst this international effort is commendable,

---

<sup>60</sup> Oxford Dictionaries identified ‘post-truth’ as the word of the year 2016: Janna Anderson and Lee Rainie, ‘The future of fake news, according to experts’ *World Economic Forum* (1 November 2017)

<https://www.weforum.org/agenda/2017/11/this-is-what-experts-think-will-happen-to-fake-news>.

<sup>61</sup> Richard Gray, ‘Lies, propaganda and fake news: A challenge for our age’ BBC (1 March 2017)

<https://www.bbc.com/future/article/20170301-lies-propaganda-and-fake-news-a-grand-challenge-of-our-age>

<sup>62</sup> Anderson (n 60).

<sup>63</sup> Tarlach McGonagle, ‘“Fake news”: False fears or real concerns?’ (2017) 35(4) *Netherlands Quarterly of Human Rights* 203; Jim Acosta, ‘How Trump’s ‘fake news’ rhetoric has gotten out of control’ *CNN* (11 June 2019)

<https://edition.cnn.com/2019/06/11/politics/enemy-of-the-people-jim-acosta-donald-trump/index.html>; Bianca Santana, ‘Jair Bolsonaro accused me of spreading ‘fake news’. I know why he targeted me’ *The Guardian* (22 June 2020)

<https://www.theguardian.com/commentisfree/2020/jun/22/jair-bolsonaro-fake-news-accusation-marielle-franco>.

<sup>64</sup> Plamen Atasanov, ‘Fake News between Artificial Intelligence and Credibility’ (PhD thesis, Goce Delcev University 2018) 1020.

<sup>65</sup> World Economic Forum, ‘Digital Wildfires in a Hyperconnected World’ *World Economic Forum*

<http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/>.

<sup>66</sup> Joint Declaration on Freedom of Information and “Fake News”, Disinformation and Propaganda (3 March 2017) UN Doc FOM.GAL/3/17 (Joint Declaration).

<sup>67</sup> *ibid*.



there is one underlying obstacle to its implementation: the absence of an agreed definition of fake news, which rather *'means different things to different people'*.<sup>68</sup> It is however possible to identify some types of fake news. Table 1 provides an overview of some of these.<sup>69</sup>

Types of Fake News	Description
Spoof or Parody	No intention to cause harm, but the news is satirical. For example: A news published by The Onion (now deleted) carried the headline <i>'Facebook apologised for giving Mark Zuckerberg a platform'</i> .
Deceiving Content	Content where facts and details are not correctly interpreted. For example: Some fake websites pretending to be famous news agencies and post fake content. They can be identified through URL checking websites.
Imposter Content	Authentic sources are imitated with wrong information for personal benefit. It is difficult to uncover this kind of content without any prior knowledge of the subject. For example: Multiple websites in 2018, reported that it was 'against the law' for people to bathe and do laundry on the same day in California. Although, the actual bill outlined the conservation guides for water districts, municipalities and water agencies and fining them if they fail to meet conservation goals till 2027.
Made-Up Content	New information is 100% false to mislead the reader. For example: A fake news surfaced that read <i>'Twitter users discovered that national trafficking hotline was run by Hillary Clinton'</i> .
False Relation	Headlines, titles or imagery do not support the content. For example: Even famous publishers like New York Daily post news that might not have any relation together. They published an article where the headline claims that sugar is as addictive as drugs like cocaine.
Wrong Context	Original content is shared with false information or facts. For example: Meme creators use old photos and publish them with the wrong context and false information
Exploited Context	True information is manipulated with edited content. For example: an image surfaced where the current Russian president, Putin was seen pulling the former US president, Obama's tie. In the real image, they were both just speaking.

Table 1: Types of Fake News

<sup>68</sup> Already in 2013, the World Economic Forum's Global Risk Report was titled "Digital Wildfires in a Hyperconnected World" and warned against the raising dangers of misinformation on social media: see n 67; Robert Booth, 'Journalists to use "immune system" software against fake news' *The Guardian* (8 August 2017) <https://www.theguardian.com/technology/2017/aug/08/fake-news-full-fact-software-immune-system-journalism-soros-omidyar>.

<sup>69</sup> Subject Guides: Fake News: Types of Fake News (2020). Available at: <https://guides.lib.uiowa.edu/c.php?g=849536&p=6077637> (Accessed: 3 December 2020).



Based on these types, it is possible to identify three main categories of fake news, namely disinformation, misinformation and malinformation. The Oxford Dictionary defines disinformation as *'false information which is intended to mislead, especially propaganda issued by a government organisation to a rival power or the media'*; and misinformation as *'false or inaccurate information, especially that which is deliberately intended to deceive'*.<sup>70</sup> The Reuters Institute characterises malinformation as the reconfiguration of existing true content.<sup>71</sup> Whilst all three do amount to different facets of the fake news phenomenon, this publication will use the term “misinformation” for consistency purposes.

In 2013, the World Economic Forum's Global Risk Report was titled “Digital Wildfires in a Hyperconnected World” warning against the rising dangers of misinformation on social media.<sup>72</sup> Misinformation comes from *'sources that spread falsehoods by manipulating their consumers' emotions and tapping into deeply held partisan beliefs'*.<sup>73</sup> It ultimately functions as a tool for the audience to choose what to believe in, often leading them to engage only with content that is self-reinforcing.<sup>74</sup> This is the so-called *'confirmation bias'* phenomenon, i.e. people perceive and interpret information in accordance with their beliefs and expectations.<sup>75</sup> Their newsfeed is then flooded only with the content of a similar sort, leading to *'an exaggerated echo chamber effect'*, which, if not precisely false, is still far from a comprehensive and reliable account of the truth.<sup>76</sup>

It is now appropriate to trace a couple of fundamental distinctions. Fake news shall not be equated with false news since the former is often *'invented (...) from thin air'* with the intent to deceive and to gain from its re-transmission.<sup>77</sup> On the other hand, false news is frequently the result of genuine mistakes in the collection or communication of information and is not spread with any deceiving intention. It is also crucial to differentiate misinformation from satirical content. In this regard, Dr Axel Gelfert distinguishes them on the basis that the latter *'pretty clearly'* guides the audience towards differentiating the fake and the real, so that in the end, *'the moderately attentive viewer [is] better informed about the world'*.<sup>78</sup> Misinformation's purveyors, on the other hand, spread it *'with the*

---

<sup>70</sup> Oxford Dictionary, *Disinformation* <https://www.lexico.com/definition/disinformation>; Oxford Dictionary, *Misinformation* <https://www.lexico.com/definition/misinformation?s=t>.

<sup>71</sup> J Scott Brennan et al, *Types, sources, and claims of COVID-19 misinformation* (Reuters Institute and University of Oxford, 7 April 2020) <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>.

<sup>72</sup> World Economic Forum (n 68).

<sup>73</sup> McGonagle (n 25) 93.

<sup>74</sup> Gray (n 61).

<sup>75</sup> Petter Bae Brandtzaeg, Asbjørn Følstad & María Ángeles Chaparro Domínguez, 'How Journalists and Social Media Users Perceive Online Fact-Checking and Verification Services' (2018) 12(9) *Journalism Practice* 1109, 1112.

<sup>76</sup> Gray (n 61).

<sup>77</sup> Gelfert (n 22) 96.

<sup>78</sup> *ibid* 92.



*intention to deceive*'.<sup>79</sup> It is done to convince the masses of something convenient to the source of such news, or something that generates more "clicks" and data traffic, and thus remuneration.<sup>80</sup> As noted by Bakir and McStay, online circulation is a '*core feature*' of misinformation.<sup>81</sup>

Nowadays, social media function as '*gateways*' to unprecedented amounts of (mis)information, and have lowered the barriers for accessing the news production industry.<sup>82</sup> Social media have thus played a '*transformative role*' facilitating the ability to both access and disseminate misinformation.<sup>83</sup> As a result, they have magnified the issue, to the extent that it is now a global socio-political concern. Accordingly, in 2017, '*the breakdown of trusted sources of information*' was identified by a panel of experts as one of '*the grand challenges we face in the 21<sup>st</sup> century*'.<sup>84</sup> In the following years, up to 2020, this challenge did not resolve but only worsened. In addition to the threats that this phenomenon has been posing to democracy, it has recently led to even more severe results.<sup>85</sup> Amidst the ongoing COVID-19 pandemic, misinformation about its origin, viable solutions, and the measures adopted has surged. This fake news has been said to '*spread faster and more easily than the virus*'.<sup>86</sup> In this regard, the American Journal of Tropical Medicine and Hygiene has reported that at least 800 people have died globally due to such misinformation.<sup>87</sup> The World Health Organisation (WHO) has referred to an "infodemic" running parallel to the actual pandemic.<sup>88</sup> It is defined as an '*overabundance of information*' during an epidemic, hampering people's search for accurate and reliable information.<sup>89</sup>

Overall, people should not underestimate the threat posed by the spread of misinformation.<sup>90</sup> '*Misinformation is not like a plumbing problem you fix. It is a social condition, like crime, that you must constantly monitor and adjust to*'.<sup>91</sup> It is, however, fundamental to determine who shall be responsible for this monitoring, the following chapter will thus be dedicated to the identification of the most appropriate actor.

---

<sup>79</sup> *ibid* 96.

<sup>80</sup> Napoli (n 10).

<sup>81</sup> *ibid* 96.

<sup>82</sup> Moran Yemini, 'The New Irony of Free Speech' (2018) 20 Colum Sci & Tech L Rev 119; Gelfert (n 23).

<sup>83</sup> Joint Declaration.

<sup>84</sup> Gray (n 61).

<sup>85</sup> *ibid*; Atanasov (n 64).

<sup>86</sup> UN Department of Global Communications, 'UN tackles "infodemic" of misinformation and cybercrime in COVID-19 crisis' UN <https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%99infodemic%E2%80%99-misinformation-and-cybercrime-covid-19>.

<sup>87</sup> Coleman (n 11).

<sup>88</sup> 'UN tackles "infodemic"' (n 86).

<sup>89</sup> WHO, '1st WHO Infodemiology Conference' WHO <https://www.who.int/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference>.

<sup>90</sup> Gray (n 61).

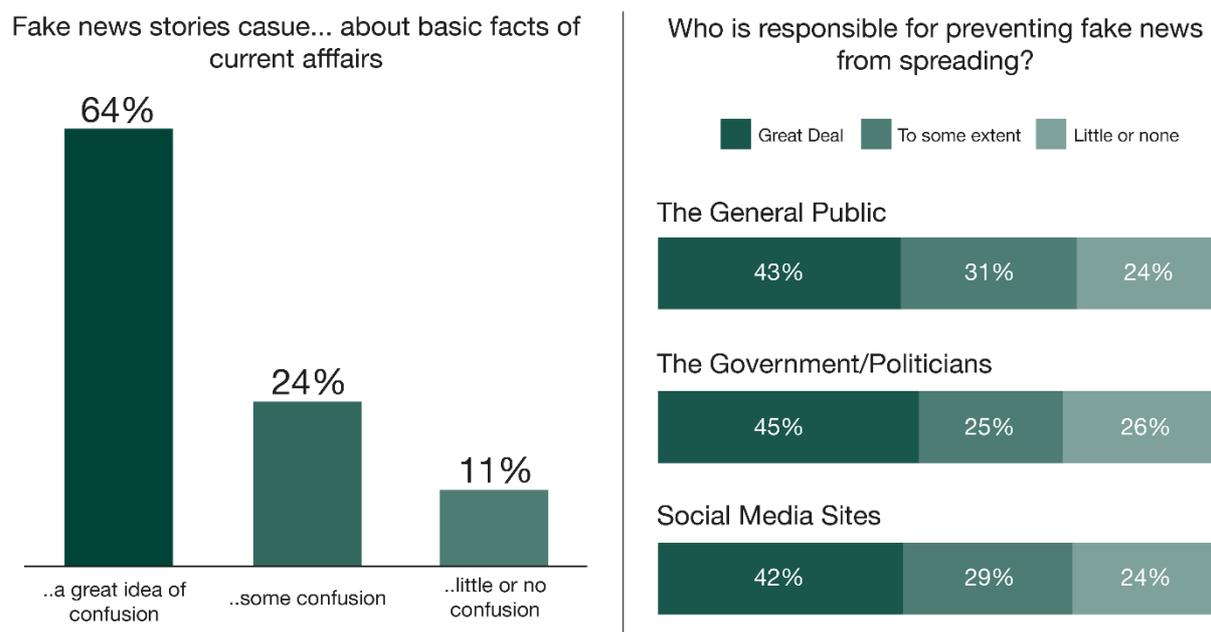
<sup>91</sup> Anderson (n 60).



### 3. Governmental Efforts against Misinformation

Having considered the definition of fake news, it is now important to evaluate who should be held responsible for the circulation of misinformation. However, beforehand, it is crucial to establish the legal nature of the sources posting or subsequently spreading the misinformation. Broadly, these sources can be grouped into three categories, namely: State actors, non-State actors, and the general public.<sup>92</sup> The umbrella term of “Non-State actors” (hereinafter: “Service Providers”) in this publication primarily refers to corporate actors such as but not limited to media companies, social media network providers (for instance Facebook, Instagram, LinkedIn and Twitter) and other non-governmental actors that are sufficiently organised to support, enable or spread misinformation independently or in collaboration with the other outlined actors.

Referring back to the accountability associated with sharing misinformation, in 2016, the Pew Research Center surveyed US citizens on the public perception on the responsibility of sources concerning fake news.<sup>93</sup>



Graph 4: 'Fake News Stories Are a Problem - But Who's to Blame?' (Source: Statista)

<sup>92</sup> Marko Milanovic, 'Viral Misinformation and the Freedom of Expression: Part II' *EJIL: Talk!* (13 April 2020) <https://www.ejiltalk.org/viral-misinformation-and-the-freedom-of-expression-part-ii/>.

<sup>93</sup> Jeff Dunn, 'Americans aren't sure who to blame for the 'fake news' problem' *Business Insider* (23 December 2016) <https://www.businessinsider.com/fake-news-survey-pew-facebook-chart-2016-12>; and Pew Research Center, *Newspapers Fact Sheet* (Pew Research Center, 9 July 2019) <https://www.journalism.org/fact-sheet/newspapers/>; The Independent ceased as a print edition in 2016: BBC, 'Independent to cease as print edition' *BBC* (12 February 2016) <https://www.bbc.com/news/uk-35561145>.



The data suggests that the participants held State actors and Service Providers equally responsible for the promotion or distribution of misinformation. Considering that a discussion on business and human rights warrants the assessment of entirely different (corporate) regimes that fall outside the scope of this research, the publication focuses only on the government's role in addressing the dangers of misinformation.

This chapter will first evaluate the role of governments in tackling misinformation. Second, it will consider governments' obligations regarding the preservation of freedom of expression, and the implications this has in relation to their role in combatting misinformation.

### 3.1. Legislative Efforts to Combat Misinformation

The advent of IoT and social media has rendered the use of the internet a fundamental component of our lives.<sup>94</sup> As a result, it is affecting, *inter alia*, the way governments work, '*altering the fabric of social, political and economic institutions*'.<sup>95</sup> Whilst some of these changes have had a positive impact such as enhancing governmental transparency and public participation, others have been detrimental to the governmental institutions. Misinformation falls into the latter category as it has been undermining the public trust in governmental actors and institutions. One could argue that the need to restore this trust is what motivates governments' work towards the suppression of this phenomenon.

However, vesting governments with the power to determine the truthfulness and accuracy of any underlying content poses a serious threat to democracy as it constitutes a pretext for State censorship.<sup>96</sup> For example, the *Enabling Act*, a recently adopted Hungarian Law, imposes a 1-5 years-long prison sentence onto those individuals spreading misinformation regarding the COVID-19 pandemic.<sup>97</sup> Whilst the prosecution of these people is *prima facie* desirable, this law constitutes a dangerous precedent both within Hungary and the EU at large.<sup>98</sup> However, this is certainly not a *sui generis* practise as there are many other instances in which States have adopted similar policies both within and outside of the EU. Even "the country of civil liberties", namely France, is now turning its back on its own legacy. The *Law on the Fight against the Manipulation of Information* provides for '*the immediate removal of fake news*' and the suspension of television which '*deliberately disseminate*'

---

<sup>94</sup> Julia Schwanholz and Todd Graham, 'Digital Transformation: New Opportunities and Challenges for Democracy?' in Julia Schwanholz et al (eds) *Managing Democracy in the Digital Age: Internet Regulation, Social Media Use, and Online Civic Engagement* (Springer 2017).

<sup>95</sup> *ibid* 1.

<sup>96</sup> Matthew Fraser, 'Debate: The legal fight against "fake news" must not veer into censorship' *The Conversation* (11 June 2018) <https://theconversation.com/debate-the-legal-fight-against-fake-news-must-not-veer-into-censorship-98049>.

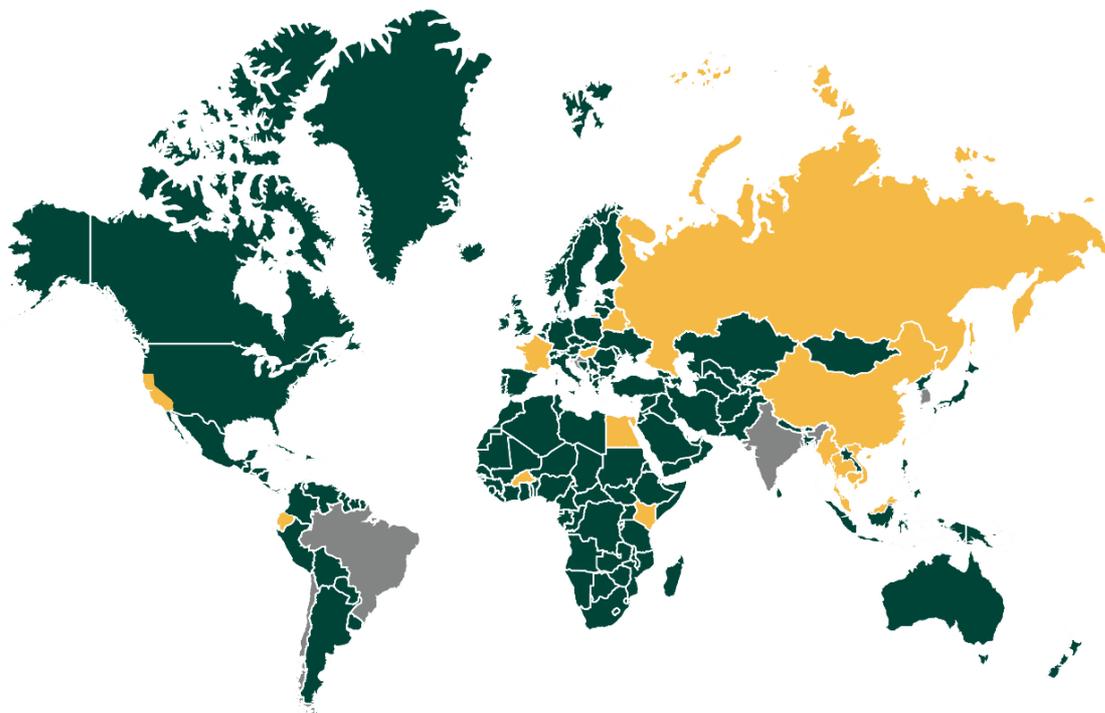
<sup>97</sup> Vlagyislav Maksimov, 'Hungary's state of emergency law is a 'blank cheque' to Orban, critics say' *Euractiv* (23 March 2020) <https://www.euractiv.com/section/justice-home-affairs/news/hungarys-state-of-emergency-law-is-a-blank-cheque-to-orban-critics-say/>.

<sup>98</sup> *ibid*.



misinformation during election campaigns.<sup>99</sup> The violation of the law could lead to 1-year imprisonment and €75,000 fine.<sup>100</sup>

## Law implemented by countries to tackle fake news



### Implemented Law

Belarus	Law on Mass Media, as amended in 2018	France	Law on the Fight against the Manipulation of Information
Burkina Faso	Law on the Fight against the Manipulation of Information	Hungary	Enabling Act
California, US	Senate Bill No. 830. CHAPTER 448. An act to add Section 51206.4 to the Education Code	Kenya	Computer and Cybercrimes Bill, 2017
Cambodia	May 2018 "Prakas" (regulation)	Malaysia	Anti-Fake News Act 2018
China	Cybersecurity Law 2016 and Provisions for the Administration of Internet News 2017	Myanmar	2014 News Media Law of Myanmar
Ecuador	2013 Communication Law, amended in 2019	Russia	March 2020, amendments to Article 207 of Criminal Code
Egypt	Law No.175/2018 - Combating Information Technology Crime; Law 180 of 2018 Regulating the Press, Media, and the Supreme Council for Media Regulation; Penal Code, Law No. 58 of 1937, and Its Amendments, ART 80C	Singapore	Protection from Online Falsehoods and Manipulation Law 2019
		Taiwan	Communicable Disease Control Act
		Thailand	Computer Crime Act
		Vietnam	Cyber Security Law

### Yet to be Implemented

Brazil	20 draft bills in 2018	India	Proposed amendments to 2000 IT Act
Chile	Ley Bolsonaro (Bolsonaro Law) 2018	South Korea	Pending law amendments
Croatia	Lex Facebook		

<sup>99</sup> LOI n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information; Michael-Ross Fiorentino, 'France passes controversial "fake news" law' *Euronews* (22 November 2018)

<http://euronews.com/2018/11/22/france-passes-controversial-fake-news-law>.

<sup>100</sup> *ibid.*



With regards to an example outside the EU, the Republic of Ecuador offers an interesting case study. In 2019, the Ecuadorian government amended *the General Regulation of the Organic Law of Communication* to include a prison sentence of 5-7-years for causing ‘*financial panic*’ by ‘*publishing, spreading or divulging false news*’.<sup>101</sup> Similarly, to combat the dissemination of fake news, the Arab Republic of Egypt introduced *Law Regulating the Press, Media, and the Supreme Council for Media Regulation* (SCMR Law), the *Law on Combating Information Technology Crime* (Anti-Cybercrime Law), and *Law No. 58 of 1937 and its amendments on the Penal Code*.<sup>102</sup> The *SCMR Law* prohibits the publication and broadcasting of misinformation to press, media, and websites, including personal websites, blogs, and social media accounts with at least 5,000 subscribers. Any content violating this prohibition can be censored under the *SCMR Law*. The *Anti-Cybercrime Law* penalises fake news as ‘*unlicensed online activity and content violation*’.<sup>103</sup> Those websites infringing this law will be blocked, and individuals suspected of breaching the law can be subject to a travel ban.<sup>104</sup> The *Penal Code* instead imposes imprisonment onto those who disclose fake or tendentious news in times of war, if it harms military preparations, creates social panic or weakens the toleration and endurance of Egypt.

The implications that these laws have for freedom of expression are appreciable by considering the case of the Al Jazeera reporter who was arrested and detained for more than two years in Egypt on allegations of broadcasting misinformation to spread chaos.<sup>105</sup> This case is one example of how ambiguities in the definition of fake news could result in governments enforcing arbitrary laws. Whilst this sets a dangerous precedent, there are also additional widespread concerns over governmental actors and institutions becoming the “Ministry of Truth”. Indeed, they could deliberately exercise their power to remove any content or discredit it as false, in an effort to contain the underlying disputed information. Such powers are especially problematic where the government has a history of exercising pressure to remove critiquing content or content that is not in line with its views, as it is the case, for

---

<sup>101</sup> 2009 Reglamento General a La Ley Orgánica de Comunicación (as amended in 2019); Attila Mong, ‘Countering fake news while safeguarding free speech’ *DW* (14 March 2017) <https://www.dw.com/en/countering-fake-news-while-safeguarding-free-speech/a-37924278>; US Department of State, *2019 Country Reports on Human Rights Practices: Ecuador* <https://www.state.gov/reports/2019-country-reports-on-human-rights-practices/ecuador/>.

<sup>102</sup> Law No. 180 of 2018 Regulating the Press, Media, and the Supreme Council for Media Regulation (SCMR Law); Law No. 175/2018 on Combating Information Technology Crime (Anti-Cybercrime Law); Penal Code Law No 58 of 1937 and Its Amendments, Art 80C.

<sup>103</sup> Kilian Bälz and Hussam Mujally, ‘The New Egyptian Anti-Cybercrime Law Regulates Legal Responsibility for Web Pages and Their Content’ *Lexology* (19 September 2018) <https://www.lexology.com/library/detail.aspx?g=90440972-f53e-46dd-b225-7f7cbdea7d73>.

<sup>104</sup> Anti-Cybercrime Law Art 7,9.

<sup>105</sup> Committee to Protect Journalists, ‘Egypt arrests Al-Jazeera producer on fake news charge’ *CPJ* (27 December 2016) <https://cpj.org/2016/12/egypt-arrests-al-jazeera-producer-on-fake-news-cha/>; AlJazeera News Agencies, ‘Egypt: Court orders release of journalist Mahmoud Hussein’ *AlJazeera* (23 May 2019) <https://www.aljazeera.com/news/2019/05/egyptian-court-orders-release-mahmoud-hussein-190523100017910.html>.



instance, in Iran and Bahrain.<sup>106</sup> Ultimately, anti-misinformation laws would effectively legitimise State censorship. These concerns are echoed by the EU Commissioner for the Digital Single Market, Andrus Ansip who famously stated, *'fake news is a bad thing, but according to my understanding, the Ministry of Truth is also bad'*.<sup>107</sup>

With the improvements in information technology, these concerns have only increased. Indeed, whilst States themselves can police and monitor social media networks, they often rely on Service Providers or other third-party information verification services, given their higher expertise.<sup>108</sup> As a result, these Service Providers can effectively determine the legitimacy of information, viewpoints and communications, thus establishing what reaches the public.<sup>109</sup> In other words, the government is delegating the crucial task of determining the contents of public information to third parties so that not the government itself but private actors might ultimately become the “arbiters of truth”.

We have yet to reach a balance between combatting misinformation and adequately protecting fundamental rights like freedom of expression. This is in part due to the lack of a guiding regulatory framework for Service Providers on the subject matter, and partly due to the absence of transparent oversight when containing misinformation.

---

<sup>106</sup> Human Rights Watch, Freedom Of Expression On The Internet <https://www.hrw.org/legacy/wr2k/Issues-04.htm#TopOfPage>.

<sup>107</sup> Mong (n 101).

<sup>108</sup> Anton Troianovski, 'Fighting False News in Ukraine, Facebook Fact Checkers Tread a Blurry Line' *NYT* (26 June 2020) <https://www.nytimes.com/2020/07/26/world/europe/ukraine-facebook-fake-news.html>.

<sup>109</sup> *ibid.*



### 3.2. Balancing Freedom of Expression and Anti-Misinformation Measures

*'Freedom of expression is the basis of human rights, the source of humanity and the mother of truth. To block freedom of speech is to trample on human rights, to strangle humanity and to suppress the truth'*

- Liu Xiaobo, Nobel Laureate<sup>110</sup>

Echoing the words of Liu Xiaobo, freedom of expression is essential to democracy and good governance.<sup>111</sup> It is often considered to be a *'multiplier or meta right'*, enabling the enjoyment of other rights, such as political participation, cultural rights or rights to assembly and association.<sup>112</sup>

However, it should be noted that freedom of expression is not an absolute right and derogation from it can thus be justified in a democratic society impaired by misinformation.<sup>113</sup> However, as required by international human rights law, this derogation should not exceed what is strictly necessary for the containment of misinformation.<sup>114</sup> The "state of emergency" argument invoked by many governments has however precluded the guarantee of any respect for such requirements. For example, the Council of Europe human rights chief, Dr Dunja Mijatović has raised concerns as to Hungary's *Enabling Act* which *'would grant sweeping powers to the government to rule by decree without a clear cut-off date and safeguards'*.<sup>115</sup> In this regard, Dr Marko Milanovic noted that the unwarranted removal of information is as harmful as misinformation itself.<sup>116</sup>

Whilst it should not be *'unjustified'*, States' intervention in the suppression of misinformation is recognised as legitimate.<sup>117</sup> On this basis, during the pandemic, many governments have derogated from freedom of expression in order to combat misinformation, and to protect *'the public against social harms'*.<sup>118</sup> However, these derogations have led to a number of violations of the right, due to their arbitrary nature.<sup>119</sup> To prevent any further violation, all derogations to fight misinformation shall be enforced in accordance with the principles of legitimacy, proportionality, and necessity.<sup>120</sup> In order

---

<sup>110</sup> Michael O'Flaherty, 'International Covenant on Civil and Political Rights: interpreting freedom of expression and information standards for the present and the future' in Tarlach McGonagle and Yvonne Donders (eds) *The United Nations and Freedom of Expression and Information* (CUP 2015) 55.

<sup>111</sup> *ibid* 56.

<sup>112</sup> *ibid* 59; ICCPR Arts 25,27,22.

<sup>113</sup> ICCPR Art 19.3.

<sup>114</sup> ICCPR Art 4.

<sup>115</sup> Maksimov (n 97).

<sup>116</sup> *ibid*.

<sup>117</sup> Milanovic (n 59).

<sup>118</sup> *ibid*; Michele Collazzo and Alexandra Tyan, 'Emergency Powers, COVID-19 and the New Challenge for Human Rights' (2020) 20(48) IAI Commentaries 1.

<sup>119</sup> IPI tracker on press freedom violations linked to COVID-19 coverage, 'COVID-19: Number of Media Freedom Violations by Region' IPI <https://ipi.media/covid19-media-freedom-monitoring/>.

<sup>120</sup> ECHR Art 15.



to achieve this, it is important that fake news is clearly defined. This would enable the authorities to effectively test the derogation against these three elements. Having established the requirements for a legitimate derogation, I will now evaluate the routes that governments can undertake to tackle the actual misinformation phenomenon.

Historically, international human rights law has been attempting to strike a balance between the regulation of misinformation and freedom of expression.<sup>121</sup> Already, in 1983, the Human Rights Committee had issued a General Comment on Art 19 referring to *'the development of modern mass media'*, and warning against the risk that the control of news media would pose to freedom of expression.<sup>122</sup> Similarly, in 1999, Human Rights Watch warned against the use of *'rating and blocking systems'* online and how they *'could be used by regulators to stifle expression'*.<sup>123</sup> Therefore, the necessity to appropriately balance the competing interests has been emphasised for long.

The approach adopted by the Service Providers, i.e. flagging fake content as opposed to removing it, is an adequate route to handling misinformation. Alternatively, the expansion of the scope of freedom of information to include a right to receive truthful information would also facilitate the resolution of the tension between freedom of expression and the fight against misinformation. Indeed, the derogation from a right in order to protect another will be more legitimate, and more regulated.<sup>124</sup> As a result, it will inevitably comply with the criteria for derogation set out by most human rights instruments: necessity and proportionality.<sup>125</sup> This approach is founded on *Art 19 Universal Declaration of Human Rights* (UDHR): *'Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of the frontier'*.<sup>126</sup> Misinformation arguably amounts to interference with the right to hold opinions and to seek information. The Human Rights Committee also issued a decision which arguably recalls the approach proposed in this publication. In *Mavlonov and Sa'di v Uzbekistan*, it found that the right to receive journalistic information was *'a corollary of the specific function of a journalist and/or editor to impart information'*.<sup>127</sup>

Having determined the freedom of expression implications of the fight against misinformation, and having proposed alternative solutions for striking a balance between the two, this publication will now

---

<sup>121</sup> Marko Milanovic, 'Viral Misinformation and the Freedom of Expression: Part I' *EJIL: Talk!* (13 April 2020) <https://www.ejiltalk.org/viral-misinformation-and-the-freedom-of-expression-part-i/>.

<sup>122</sup> OHCHR, 'General Comment No. 10: Article 19 (Freedom of opinion and expression)' UN Doc HRI/GEN/1/Rev.9 (1983) (GC 10).

<sup>123</sup> HRW Report (n 106).

<sup>124</sup> *Robert W Gauthier v Canada* (1999) 65<sup>th</sup> Session of Human Rights Committee 633.

<sup>125</sup> *ibid.*

<sup>126</sup> Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III)) (UDHR) Art 19.

<sup>127</sup> *Rakhim Mavlonov and Mr. Shansiy Sa'di v Uzbekistan* (2009) 95<sup>th</sup> Session of Human Rights Committee 1334.



demonstrate the necessity for such alternatives by evaluating the inadequacy existing legal avenues available to the victims of misinformation.

### 3.3. Existing Solutions

There are two main routes which can be undertaken by the victims of misinformation, namely defamation laws and the so-called “memory laws”.<sup>128</sup> The former provides victims with monetary compensation which is not a practical and satisfactory remedy.<sup>129</sup> Moreover, it requires non-contractual harm between private individuals, restricting its applicability to a limited set of scenarios. The Joint Declaration, accordingly, states that defamation laws are ‘*unduly restrictive*’ in the fight against fake news.<sup>130</sup> Memory laws, on the other hand, criminalise the denial of historical events, such as the Holocaust, again denoting the necessity for very specific circumstances.<sup>131</sup>

This publication thus respectfully disagrees with the argument of the UK human rights organisation “Article 19” that the existing laws are sufficient and that any further regulation to tackle misinformation would have a chilling effect upon the media.<sup>132</sup> Instead, the existing laws refer to specific scenarios, and establish a very high threshold for the victims of misinformation, often leaving them remediless. The necessity for a human rights approach to misinformation is thus motivated by the inadequacy of the existing solutions, in addition to being more appropriate in striking a balance between the competing interests. Governments would be able to legitimately enforce this new approach, with the aid of emerging technologies, and they would necessarily do so within the boundaries set by other human rights provisions, thus partly defeating the aforementioned concerns. However, the phenomenon of misinformation is so pervasive that the fight against it must come from multiple fronts. The following section will thus consider one of these additional fronts, one of crucial importance: the divulgation of the so-called digital literacy, i.e. the alphabetisation of the masses to how the media work.

---

<sup>128</sup> *Robert Faurisson v France* (1993) 58<sup>th</sup> Session of Human Rights Committee 550; UNHRC ‘General Comment 34 on Article 19: Freedoms of opinion and expression’ (12 September 2011) UN Doc CCPR/C/GC/34 GC 34 [49]; NPR Staff, ‘What Legal Recourse Do Victims Of Fake News Stories Have?’ *NPR* (7 December 2016) <https://www.npr.org/2016/12/07/504723649/what-legal-recourse-do-victims-of-fake-news-stories-have?t=1598346365344>; Milanovic (n 59).

<sup>129</sup> NPR Staff (n 128).

<sup>130</sup> Joint Declaration s 2(b).

<sup>131</sup> *Faurisson* (n 128); GC 34 [49]; Milanovic (n 59).

<sup>132</sup> Teemu Henriksson, ‘Article 19: legal restrictions to “fake news” could be used to limit press freedom’ *WAN IFRA* (29 November 2016) <https://blog.wan-ifra.org/2016/11/29/article-19-legal-restrictions-to-fake-news-could-be-used-to-limit-press-freedom>.



### 3.4. Digital Literacy

Historically, fully independent media have been praised and sought. However, those who have been praising and seeking did not envision the possibility of such media sharing misinformation, and most importantly did not and could not foresee the damage that this phenomenon would have produced. With literacy and access to information expanding through the internet revolution, an increasing amount of people have been seizing a growing amount of information. It goes without saying that not all of them have the intellectual abilities to critically judge all this information. It is, therefore, paramount to educate the users.

To this end, Nina Jankowicz, a global fellow at the Wilson Center's Kennan Institute in Washington DC, stated that we could defeat misinformation through tech-based solutions coupled with the diffusion of digital literacy.<sup>133</sup> Even the aforementioned Joint Declaration provides that States ought to advance digital literacy.<sup>134</sup> Moreover, fact-checking will also benefit from this and expand further as it is documented that 'those with high political knowledge' perceive it with greater favour.<sup>135</sup> In the words of Mr Justice Brandeis, *'the remedy to be applied is more speech, not enforced silence'*.<sup>136</sup>

Some countries have already begun following this path, introducing digital literacy in schools.<sup>137</sup> In Sweden, children are educated to build digital competencies and learn how to distinguish between reliable and unreliable sources.<sup>138</sup> Similarly, in Taiwan, children are developing media literacy, being taught how to identify and combat misinformation.<sup>139</sup> Italy has also initiated an experiment to train students to recognise misinformation and conspiracy theories online.<sup>140</sup>

In conclusion, there is an overall consensus around the need for digital literacy. This publication demonstrates that the spread of digital literacy is indeed the long-term investment that is necessary, and which will eventually defeat misinformation. Digital literacy will also provide a solution to the freedom of expression implications entailed in a more interventionist approach. However, for more

---

<sup>133</sup> Peter Talbot, 'Twitter Removes Thousands Of Accounts For Manipulating Its Platform' *NPR* (20 September 2019) <https://www.npr.org/2019/09/20/762799187/twitter-removes-thousands-of-accounts-for-manipulating-their-platform?t=1597910217677> .

<sup>134</sup> Joint Declaration s 3(e).

<sup>135</sup> Brandtzaeg (n 75) 1112.

<sup>136</sup> *Whitney v People of State of California* [1927] 274 U.S. 357 [44].

<sup>137</sup> Lee Roden, 'Swedish kids to learn computer coding and how to spot fake news in primary school' *The Local SE* (13 March 2017) <https://www.thelocal.se/20170313/swedish-kids-to-learn-computer-coding-and-how-to-spot-fake-news-in-primary-school>; Nicola Smith, 'Schoolkids in Taiwan Will Now Be Taught How to Identify Fake News' *Time* (7 April 2017) <https://time.com/4730440/taiwan-fake-news-education/>; Jason Horowitz, 'In Italian Schools, Reading, Writing and Recognizing Fake News' *NYT* (18 October 2017) <https://www.nytimes.com/2017/10/18/world/europe/italy-fake-news.html>.

<sup>138</sup> Roden (n 137).

<sup>139</sup> Smith (n 137).

<sup>140</sup> Horowitz (n 137).



immediate and necessary results, automation and emerging technologies constitute an effective and efficient response to the infodemic. In this regard, automated fact-checking should not lead to re-limiting the sources of information, compelling us to go backwards; instead, it should leverage the abundance of information to assess the credibility of a news item.<sup>141</sup> The next chapter is dedicated to evaluating the merits of the deployment of emerging technologies for the purpose of containing misinformation.

---

<sup>141</sup> Anderson (n 60).



## 4. Technological Innovations to Combat Misinformation

---

From the case studies analysed in Chapter 3, it becomes evident that a legislative effort without the support of adequate technologies will hinder the fight against the creation and dissemination of misinformation. Once citizens stop trusting the information divulged by the State actors, they might effectively *'switch off'*, thus impairing one of the pillars of democracy: public participation.<sup>142</sup>

In the post-truth era, fact-checking technology developed by Service Providers has become fundamental for the preservation of democracy and people's engagement with civil society. Moreover, adequate fact-checking ensures the accountability of those making the claims.<sup>143</sup> Overall fact-checking can be, and has been, compared to *'an immune system'* and, during a pandemic, there is, in my opinion, no more effective simile.<sup>144</sup>

Traditionally, news spreads through newspapers or television, which means that the information written either on publication or script, undergoes several editorial stages before being shared. The advent of digitalisation and the IoT has provided us with a means to directly access an unprecedented amount of information, at an unparalleled speed and easiness, yet with fewer checks. Our smartphones send us notifications for "Breaking News" with a significant frequency when, just 20 years ago, only events of the gravity of 9/11 led to the interruption of our daily activities. Accordingly, those journalists who work with online news publishing have declared to have *'insufficient time for verification and fact-checking due to a fast-paced publishing environment'*.<sup>145</sup>

This acceleration of news divulgation has been occurring in parallel with the advent and uptake of social media, which empowered all users to produce and share content giving rise to the so-called *'citizen journalism'*.<sup>146</sup> Moreover, social media have provided users with tools to share such information with just one click, be it a *Re-Tweet* on Twitter, or *Share* on Facebook. The effortlessness of this action gave it an exponential frequency: for instance,

---

<sup>142</sup> Full Fact, 'Automated Fact Checking' *Full Fact* <https://fullfact.org/automated>.

<sup>143</sup> Booth (n 68).

<sup>144</sup> *ibid.*

<sup>145</sup> Brandtzaeg (n 75) 1113.

<sup>146</sup> *ibid*; Napoli (n 10).



there are about 6000 tweets per second.<sup>147</sup> As a result, false or inaccurate content has been spreading at a faster and broader rate.<sup>148</sup> It is in this context that misinformation fostered, and came to obtain the weight and magnitude it currently retains. According to reports, the diffusion of misinformation '*typically outpaces*' that of the content to fact-check it.<sup>149</sup> Indeed, the current speed of data traffic is physically impossible to reach for manual fact-checkers.<sup>150</sup>

Fact-checking is a long multi-stage process consisting of first, the collection of information, second, the identification of a claim in the public domain, and lastly, the analysis of the claim's accuracy through an authoritative source of information.<sup>151</sup> Verifying the accuracy of online content and the identity of content-producers has become exceedingly more difficult.<sup>152</sup> The traditional methods of fact-checking have become outdated, and can be easily '*circumvented by social media*'.<sup>153</sup> Since the sources of information are bound to change with time, so shall change the assessment of their reliability. Hence this publication considers how emerging technologies can function as a way to successfully contain and eventually defeat the issue of misinformation.

Whilst the interplay between business and human rights is outside the scope of this publication, this chapter provides a comprehensive overview of how some of the Service Providers are combatting misinformation on their platforms. It further outlines the measures adopted by these providers and the underlying technologies which State actors could benefit from. Moreover, it also highlights avenues for a collaborative effort in the fight against misinformation. Holistically, this chapter will first outline the mechanics of automated fact-checking, considering some specific examples. Second, it will consider how tech giants like Facebook and Twitter have been implementing this technology. Thirdly, it will evaluate the current shortcomings of automated fact-checking, to conclude that the *status quo* of a combination AI-powered and human-led fact-checking is the preferable route.

---

<sup>147</sup> 'Twitter Usage Statistics' *Internet Live Stats* <https://www.internetlivestats.com/twitter-statistics/> accessed on 27<sup>th</sup> November 2020.

<sup>148</sup> Brandtzaeg (n 75).

<sup>149</sup> *ibid* 1112.

<sup>150</sup> 'A typical fact-check has been reported to take a day', which is unsustainable given the speed of circulation of information nowadays: Lucas Graves, *Understanding the Promise and Limits of Automated Fact-Checking* (Reuters Institute and University of Oxford, February 2018) 4.

<sup>151</sup> Matt Burgess, 'Google is helping Full Fact create an automated, real-time fact-checker' *Wired* (17 November 2016). <https://www.wired.co.uk/article/automated-fact-checking-full-fact-google-funding>; Brandtzaeg (n).

<sup>152</sup> McGonagle (n 25).

<sup>153</sup> Gray (n 61).



#### 4.1. Automated Fact-Checking

Digitalisation has significantly increased the speed of diffusion of misinformation. Hence, pre-existing fact-checkers had to be digitalised and automated with Full Fact and BBC being just two examples.<sup>154</sup> In 2017, it was possible to register 114 active automated fact-checkers.<sup>155</sup> In 2018, the Reuters Institute and the University of Oxford published a fact-sheet on automated fact-checking.<sup>156</sup> They defined it as having three main aims: first, the identification of false or questionable claims on the media; second, the verification of such claims; and third, the delivery of any correction to such claim in an instantaneous and spread manner.<sup>157</sup> The claims to be checked are identified, and prioritised, through a combination of natural language processing and Machine Learning (ML).<sup>158</sup>

In recent years, many companies have been developing technologies for automated fact-checking. This trend arguably points towards an increase in the demand for such technologies due to the aforementioned faster spread of fake news. The examples considered hereunder originate from the UK, the US, and the Argentine Republic, aiming at providing a comprehensive picture of some of the most prominent fact-checking start-ups around the world.

Full Fact is a UK-based fact-checking firm.<sup>159</sup> It recently developed a technology which automatically scans transcripts from the Parliament and the BBC, checking them against official statistics.<sup>160</sup> In particular, it functions through two tools: “Live” and “Trends”.<sup>161</sup> The latter traces and records fake claims as well as their source, in order to identify those users which persistently share misinformation.<sup>162</sup> It also shows the extent to which a particular claim spreads on the web.<sup>163</sup> “Live”, on the other hand, identifies already fact-checked claims,

---

<sup>154</sup> Automated Fact Checking, *Full Fact* <https://fullfact.org/automated/>; BBC Reality Check, *BBC* [bbc.com/news/reality\\_check](https://www.bbc.com/news/reality_check).

<sup>155</sup> Brandtzaeg (n 75).

<sup>156</sup> Graves (n 150).

<sup>157</sup> *ibid* 2.

<sup>158</sup> *ibid* 3.

<sup>159</sup> Full Fact (n 142).

<sup>160</sup> Daniel Funke, ‘Automated fact-checking has come a long way. But it still faces significant challenges’ *Poynter* (4 April 2018) <https://www.poynter.org/fact-checking/2018/automated-fact-checking-has-come-a-long-way-but-it-still-faces-significant-challenges/>.

<sup>161</sup> Full Fact (n 142).

<sup>162</sup> *ibid*.

<sup>163</sup> Matt Burgess, ‘Google is helping Full Fact create an automated, real-time fact-checker’ *Wired* (17 November 2016). <https://www.wired.co.uk/article/automated-fact-checking-full-fact-google-funding>.



and automatically validates the information in real-time.<sup>164</sup> It even earned the title of *'the most effective approach to automatic verification'*.<sup>165</sup> Full Fact is also investing in Research and Development (R&D) on how AI and ML can help to further automated fact-checking.<sup>166</sup>

Another service of this kind was developed by IBM through Watson: Watson Angles.<sup>167</sup> This tool checks online information against a collection of about 55 million news articles.<sup>168</sup> It collects data from these articles to assemble basic facts regarding the relevant topic, providing context, a time-frame as well as key quotes.<sup>169</sup> Watson Angles is dedicated to "everyone" rather than to journalists only, thus differing from Full Fact.<sup>170</sup> The ultimate aim is indeed *'to overcome the boundaries that separate readers and viewers from journalists'*.<sup>171</sup>

In Argentina, the fact-checking outlet Chequeado has developed Chequeabot.<sup>172</sup> The bot is an AI tool which functions through *'natural language processing and ML'*.<sup>173</sup> It automatically identifies *'fact-checkable claims'* from 25 national media outlets, and checks them against an ever-expanding database of previously verified facts.<sup>174</sup> In this way, the bot covers the whole country rather than focusing only on the capital, Buenos Aires.<sup>175</sup> Moreover, the bot checks news in Spanish, partly countering the English-language bias characterising the grand part of current automated fact-checkers.<sup>176</sup>

Having outlined a few examples of automated fact-checking, this publication will proceed onto considering the deployment of analogous technologies by, Facebook and Twitter, respectively.

---

<sup>164</sup> Full Fact (n 142); *ibid.*

<sup>165</sup> Graves (n 150) 4.

<sup>166</sup> Full Fact (n 142).

<sup>167</sup> Alan Greenblatt, 'What does the future of automated fact-checking look like?' *Poynter* (4 April 2016) <https://www.poynter.org/fact-checking/2016/whats-does-the-future-of-automated-fact-checking-look-like/>.

<sup>168</sup> *ibid.*

<sup>169</sup> *ibid.*

<sup>170</sup> *ibid.*; FullFact developers emphasised that, at least initially, the general public shall not be provided with fact-checking technology but only journalists: Booth (n 68).

<sup>171</sup> Greenblatt (n 167).

<sup>172</sup> Chequeado, 'Chequeabot' <https://chequeabot.chequeado.com/desgrabador/>.

<sup>173</sup> Daniel Funke, 'In Argentina, fact-checkers' latest hire is a bot' *Poynter* (11 January 2018) <https://www.poynter.org/fact-checking/2018/in-argentina-fact-checkers%C2%92-latest-hire-is-a-bot/>.

<sup>174</sup> *ibid.*; Funke (n 160); Chequeado (n 172).

<sup>175</sup> Funke (n 173).

<sup>176</sup> *ibid.*; Funke (n 160).



## 4.2. Fact-checking on Facebook and Twitter

An increasing number of Social Network Services are developing built-in algorithms to fact-check, flag and potentially remove fake news.<sup>177</sup> Given the prominence that Facebook and Twitter have acquired in many fake news scandals,<sup>178</sup> they are the chosen case studies for this publication.

### 4.2.1. Facebook

Facebook has been increasingly acting to combat viral misinformation.<sup>179</sup> It has been partnering with both governments and the academia to eradicate the so-called “deepfakes”.<sup>180</sup> These are characterised by ‘*edited or synthesised*’ content which would mislead the average user.<sup>181</sup> Additionally, Facebook’s independent third-party fact-checkers can still review content that is not classified as “deepfake”.<sup>182</sup> This content is detected through feedback from users, or by the fact-checkers themselves.<sup>183</sup> Upon the content being signalled and identified as fake, it is accordingly labelled, and banned from running as an ad.<sup>184</sup> Moreover, a notification is sent to those users who had shared that content.<sup>185</sup> As a result, there is a significant reduction in its distribution on the News Feed.<sup>186</sup>

The vastness of the pool of Facebook registered users (about 2.2 billion), and the daily posts exceeding 1 billion make human fact-checking physically impossible.<sup>187</sup> Therefore, Facebook relies on AI to supplement its big team of human fact-checkers.<sup>188</sup> During his trial before Congress in April 2018, Mark Zuckerberg referred to AI as the solution to digital

---

<sup>177</sup> Brandtzaeg (n 75).

<sup>178</sup> Julia Carrie Wong and Sam Levine, ‘Twitter labels Trump’s false claims with warning for first time’ *The Guardian* (27 May 2020) <https://www.theguardian.com/us-news/2020/may/26/trump-twitter-fact-check-warning-label>; Nicholas Confessore, ‘Cambridge Analytica and Facebook: The Scandal and the Fallout So Far’ (*NYT*, 4 April 2018) <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>.

<sup>179</sup> Facebook, ‘Fact Checking on Facebook’ <https://www.facebook.com/business/help/2593586717571940>; Agence France-Press (n 7).

<sup>180</sup> Monika Bickert, ‘Enforcing Against Manipulated Media’ *Facebook* (6 January 2020) <https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/>.

<sup>181</sup> *ibid.*

<sup>182</sup> *ibid.*

<sup>183</sup> Fact Checking on Facebook (n 179).

<sup>184</sup> *ibid.*; Bickert (n 180).

<sup>185</sup> Fact Checking on Facebook (n 179).

<sup>186</sup> Bickert (n 180).

<sup>187</sup> Malek Murison, ‘Facebook using machine learning to fight fake news’ *Internet of Business* (25 June 2018) <https://internetofbusiness.com/facebook-machine-learning-fake-news/amp/>.

<sup>188</sup> Kamalika Some, ‘Facebook Uses AI To Fight Coronavirus Misinformation And Fake News’ *Analytics Insight* (13 June 2020) <https://www.analyticsinsight.net/facebook-uses-ai-fight-coronavirus-misinformation-fake-news/>.



misinformation.<sup>189</sup> That same year, Tessa Lyons, Facebook product manager, stated that ‘ML helps us identify duplicates of debunked stories’, and their source.<sup>190</sup>

AI outpaces human fact-checking, both in terms of the amount of content it processes and its ability to distinguish ‘images that appeared similar to a human eye’.<sup>191</sup> The latter operates through a ‘similarity detector’ which identifies ‘the difference between images that look graphically similar but carry different information’.<sup>192</sup> Facebook AI was trained to categorise images by itself with 3.5 billion hashtagged Instagram photos.<sup>193</sup> Moreover, Facebook’s detector collects data to build a database which then scans the Social Network for content regarding the same data and flags it as suspicious.<sup>194</sup> In September 2019, Facebook launched the “Deepfake Detection Challenge” for experts to submit their deepfake detection models, fostering R&D for new tech tools.<sup>195</sup>

The flagging, as opposed to the removal, of content, allows the user to be informed about the fake nature of a specific post, preventing them from being misled on other platforms.<sup>196</sup> It reveals Facebook’s more far-reaching project to eradicate misinformation from the internet at large, rather than only from the platform, and arguably to further digital literacy. Thus far, the flagging can be said to have been successful: reportedly, these labels keep off 95% of the users from clicking it.<sup>197</sup>

In April 2020, Facebook detected and removed 50 million posts containing fake news on COVID-19.<sup>198</sup> As of summer 2020, the labelling began to apply to the information shared by State-controlled accounts which ‘if not necessarily false or unreliable, at least worth considering the origin of’.<sup>199</sup> The timing was arguably motivated by the forthcoming US

---

<sup>189</sup> Samuel Woolley, ‘We’re fighting fake news AI bots by using more AI. That’s a mistake’ *MIT Technology Review* (8 January 2020) <https://www.technologyreview.com/2020/01/08/130983/were-fighting-fake-news-ai-bots-by-using-more-ai-thats-a-mistake/>; Sheera Frenkel, ‘Facebook Says It Deleted 865 Million Posts, Mostly Spam’ (*NYT*, 15 May 2018) <https://www.nytimes.com/2018/05/15/technology/facebook-removal-posts-fake-accounts.html>.

<sup>190</sup> Tessa Lyons, ‘Increasing Our Efforts to Fight False News’ *Facebook* (21 June 2018) <https://about.fb.com/news/2018/06/increasing-our-efforts-to-fight-false-news/>; Murison (n 44).

<sup>191</sup> Some (n 188).

<sup>192</sup> *ibid.*

<sup>193</sup> Chris Middleton, ‘Facebook trained AI with 3.5 billion hashtagged Instagram photos’ *Internet of Business* <https://internetofbusiness.com/facebook-training-ai-instagram/>.

<sup>194</sup> Some (n 188).

<sup>195</sup> Facebook, ‘Deepfake Detection Challenge Dataset’ *AI Facebook* (25 June 2020) <https://ai.facebook.com/datasets/dfdc/>.

<sup>196</sup> *ibid.*

<sup>197</sup> Some (n 188).

<sup>198</sup> *ibid.*

<sup>199</sup> Devin Coldewey, ‘Facebook adds labels identifying state-controlled media’ *TechCrunch* (4 June 2020) <https://techcrunch.com/2020/06/04/facebook-adds-labels-identifying-state-controlled-media/>.



elections and the desire to prevent a repetition of the 2016 Cambridge Analytica scandal. Accordingly, Facebook has *enforced 'a new, cumbersome approval process for political advertisers'*.<sup>200</sup> It extinguished false claims and implemented *'a virality circuit-breaker'* so that fact-checkers had the time to assess *'suspicious stories'*.<sup>201</sup> As of October 2020, Facebook *'rejected 2.2 million ads for breaking political campaigning rules'*.<sup>202</sup> Overall, as the New York Times' headlines recite "On Election Day, Facebook and Twitter Did Better by Making Their Products Worse".<sup>203</sup>

#### 4.2.2. Twitter

Twitter anticipated Facebook in enforcing its anti-disinformation policy against Trump. In May 2020, Twitter had already taken action against the President Trump's tweets *'labelling them with a warning sign'*.<sup>204</sup> Trump was also one of the targets of Twitter's election-related measures: it *'disabled sharing features on tweets containing misleading information about election results, including some from'* his account, in addition to banning political ads.<sup>205</sup>

Generally, Twitter's policy has been slightly more aggressive than Facebook's, as the chirping social network proceeds to suspend those accounts spreading misinformation permanently.<sup>206</sup> This more intrusive approach is due to Twitter's higher political relevance, in comparison to Facebook's, as it is appreciable from the more frequent media reference to the former.<sup>207</sup> The basis for suspending the accounts is Twitter's policy on platform manipulation, which is defined as the artificial amplification, or suppression of information, misleading the users and disrupting their experience.<sup>208</sup> In addition to suspending the accounts, Twitter has been archiving those tweets fostering misinformation for research

---

<sup>200</sup> Kevin Roose, 'On Election Day, Facebook and Twitter Did Better by Making Their Products Worse' *NYT* (5 November 2020) <https://www.nytimes.com/2020/11/05/technology/facebook-twitter-election.html>.

<sup>201</sup> *ibid.*

<sup>202</sup> Agence France-Press (n 7).

<sup>203</sup> Roose (n 200).

<sup>204</sup> Wong and Levine (n 177).

<sup>205</sup> Roose (n 200).

<sup>206</sup> Talbot (n 133).

<sup>207</sup> See for example: Richard Luscombe, "'Chill!': Greta Thunberg recycles Trump's mockery of her as he tries to stop votes' *The Guardian* (5 November 2020) <https://www.theguardian.com/environment/2020/nov/05/greta-thunberg-donald-trump-twitter-chill>; Yelena Dzhanova, 'Joe Biden and Kamala Harris changed their Twitter bios to reflect their White House victories' *Business Insider* (7 November 2020) <https://www.businessinsider.com/biden-harris-change-twitter-bios-to-reflect-white-house-victories-2020-11?IR=T>.

<sup>208</sup> Twitter, 'Platform manipulation and spam policy' *Twitter* (September 2019) <https://help.twitter.com/en/rules-and-policies/platform-manipulation>.



purposes.<sup>209</sup> Another feature of Twitter’s fact-checking approach is the recent inclusion of a “get the facts” tag for the user to obtain more information.<sup>210</sup>

In 2019, Twitter acquired Fabula AI to tackle fake news.<sup>211</sup> This UK-based start-up can process and analyse vast datasets to find evidence of network manipulation, and identify patterns.<sup>212</sup> Moreover, the tool has developed ‘a *truth-risk score*’ for individual tweets, whose trustworthiness will then be assessed by Twitter itself.<sup>213</sup> The algorithm also identifies the pattern with which fake news spreads, which is reportedly different from that of “real” news.<sup>214</sup> At the moment of writing, Fabula AI is still at the developmental stage, and Twitter has only deployed it internally.<sup>215</sup>

Overall, it is arguable that Twitter has been slightly more hesitant in digitalising the fact-checking procedures, and it reportedly still relies mostly on human fact-checkers.<sup>216</sup> It is yet relatively soon to claim with certainty which Social Network adopted the best strategy.

### 4.3. The Shortcomings of Automated Fact-Checking

The examples mentioned above showcase that automated fact-checking is an ever-expanding field. However, currently, there are only a few tools which ‘*help assure readers that the media they are seeing online [originated] from a trusted source*’ and has not been altered.<sup>217</sup> Moreover, this technology still presents several shortcomings. Accordingly, Samuel Wolley, from the Oxford Internet Institute, argues that using AI to fight fake news AI bots is a mistake.<sup>218</sup> Indeed, one of the defining features of AI is that it keeps learning and self-

---

<sup>209</sup> Talbot (n 133).

<sup>210</sup> Elizabeth Culliford and Katie Paul, ‘With fact-checks, Twitter takes on a new kind of task’ *Reuters* (31 May 2020) <https://www.reuters.com/article/us-twitter-factcheck-idUSKBN2360U0> .

<sup>211</sup> Ryan Daws, ‘Twitter’s latest acquisition tackles fake news using AI’ *AI News* (4 June 2019) <https://artificialintelligence-news.com/2019/06/04/twitter-acquisition-fake-news-ai/> .

<sup>212</sup> *ibid.*

<sup>213</sup> *ibid.*

<sup>214</sup> Natasha Lomas, ‘Fabula AI is using social spread to spot “fake news”’ *TechCrunch* (7 February 2019) <https://techcrunch.com/2019/02/06/fabula-ai-is-using-social-spread-to-spot-fake-news/>

<sup>215</sup> *ibid.*

<sup>216</sup> Culliford (n 210).

<sup>217</sup> Natasha Lomas, ‘Microsoft launches a deepfake detector tool ahead of US election’ *TechCrunch* (2 September 2020) [https://techcrunch.com/2020/09/02/microsoft-launches-a-deepfake-detector-tool-ahead-of-us-election/?utm\\_medium=TCnewsletter&tpcc=TCdailynewsletter](https://techcrunch.com/2020/09/02/microsoft-launches-a-deepfake-detector-tool-ahead-of-us-election/?utm_medium=TCnewsletter&tpcc=TCdailynewsletter).

<sup>218</sup> Woolley (n 189) ; Bobby Allyn ‘Nearly Half Of Accounts Tweeting About Coronavirus Are Likely Bots’ *NPR* (20 May 2020) <https://www.npr.org/sections/coronavirus-live-updates/2020/05/20/859814085/researchers-nearly-half-of-accounts-tweeting-about-coronavirus-are-likely-bots>.



improving.<sup>219</sup> Therefore, combatting AI with similar technology would generate a sort of vicious cycle whereby both sides improve reciprocally, finding new ways to circumvent the other. In addition to this more overarching counterargument to automated fact-checking, some more specific and practical shortcomings are identifiable.

First, in order to automate fact-checking, it is necessary to collect sufficient data to enable artificial intelligence to be intelligent, and machine learning to learn.<sup>220</sup> The collection of data and the formation of a database is hard to actuate, especially with regards to fake news given the limited time-frame in which fact-checking is necessary and useful. “News” inherently entails real-time information about the most recent updates, requiring some time for the collection of reliable data on it. Moreover, fake news regards many different topics, and arises in *‘a plethora of different contexts’*, significantly hampering AI and ML data collection.<sup>221</sup>

Second, the verification stage requires contextualisation and (human) judgement.<sup>222</sup> For instance, the computer will need a certain sensitivity to determine whether the information was expressed in a *‘fair and balanced way’*.<sup>223</sup> Moreover, misinformation often concerns controversial facts whose checking process is even more complicated and “sensitive”.<sup>224</sup> AI may encounter difficulties in unpacking the *‘fuzzy logic and idiom’* people use in speech.<sup>225</sup> Thus far, automated fact-checking can identify only *‘declarative statements’* leaving out *‘implied claims’*.<sup>226</sup> The “translation” of claims to a “language” that is understandable to a database is indeed one of the biggest challenges faced by AI researchers. Computers will have to learn how to decipher, understand, and fact-check nuance and rhetoric.<sup>227</sup> Another language-related hurdle is English bias.<sup>228</sup> Indeed, most fact-checking start-ups are developed in English so that the technology cannot detect fake news in other languages.

---

<sup>219</sup> Lomas (n 217).

<sup>220</sup> Greenblatt (n 167).

<sup>221</sup> Woolley (n 188).

<sup>222</sup> Graves (n 150).

<sup>223</sup> Lora Kolodny, ‘Full Fact aims to end fake news with automated fact checking tools’ *TechCrunch* (17 November 2016) <https://techcrunch.com/2016/11/17/full-fact-scores-google-grant-to-put-an-end-to-fake-news/?guccounter=1> ; Graves (n 103).

<sup>224</sup> Brandtzaeg (n 75).

<sup>225</sup> Booth (n 69).

<sup>226</sup> Daniel Funke, ‘Here’s how close automated fact-checking is to reality’ *Poynter* (28 February 2018) <https://www.poynter.org/fact-checking/2018/heres-how-close-automated-fact-checking-is-to-reality/>.

<sup>227</sup> Greenblatt (n 167).

<sup>228</sup> Funke (n 226).



However, all these flaws can be overcome via R&D. Yet, as noted by both Mark Zuckerberg and Alessandro Moschitti, a computer scientist at the Qatar Computing Research Institute and the University of Trento, *'the technology is there, but we also need resources'*.<sup>229</sup> These include financial resources but also access to reliable official databases.<sup>230</sup>

Overall, although it accelerates the process, automated fact-checking will have to be coupled with human supervision for the foreseeable future.<sup>231</sup> Therefore, automated fact-checking will have more of an enhancing function for the work carried out by humans.<sup>232</sup> AI will just be responsible for mechanical and repetitive tasks, whilst humans retain control.<sup>233</sup> In the long run, news outlets websites will have in-built widgets for fact-checking purposes, as a way to facilitate and smooth fact-checking for users before they share information.<sup>234</sup> However, due to the rather vague definition of fake news, it is arguable that the enforcement of such a regime is currently unfeasible and legally controversial.

---

<sup>229</sup> Greenblatt (n 167).

<sup>230</sup> Funke (n 226).

<sup>231</sup> Graves (n 150); AI is still wrong 15% of the times: Middleton (n 193).

<sup>232</sup> *ibid.*

<sup>233</sup> Some (n 188).

<sup>234</sup> Greenblatt (n 167).



## 5. Conclusion

---

In the era of digitalisation, emerging technologies have been pervading many aspects of civil society. This development has had several positive outcomes, such as easier access to, and distribution of information. Social media has enhanced the public debate, providing a platform for voicing previously unheard opinions, and ultimately advancing fundamental rights such as freedom of expression. However, this increase in access to information has led to a proportionate increase in the dissemination of misinformation.

The lack of a legislative effort in containing and addressing misinformation on social media platforms has inevitably led to the users being more vulnerable to its dangers. It is perhaps because of this that we are now in a “post-truth era”, where rapid proliferation of misinformation constitutes a significant social challenge undermining the trust in governmental institutions, and thus reducing public participation. In parallel to destabilising democracies, purveyors of misinformation, be it State actors, or Service Providers, directly or indirectly infringed upon other fundamental rights such as but not limited to the right to life and to health. Acknowledging these dangers, the WHO has labelled the overabundance of contradictory information surrounding COVID-19 an “infodemic”.<sup>235</sup> The containment of misinformation needs to become a pressing issue for governments across the globe.<sup>236</sup>

This publication analysed the interplay between fundamental rights and proliferation of misinformation and highlighted how the latter might threaten the (re)instatement, and implicit legitimisation, of censorship by State actors or Service Providers. Addressing this, the publication advances a thesis whereby expanding freedom of information to include a right to true information may eliminate this threat.

One of the core hurdles in the regulation of fake news is, as noted in Chapter 2, the absence of an agreed-upon definition of this phenomenon. However, for clarity purposes, this publication has been referring to one of its ramifications instead: misinformation as ‘false or inaccurate information, especially that which is deliberately intended to deceive’.<sup>237</sup> Building on this, the third chapter focused on the responsibility of governments *vis-à-vis* misinformation. The publication advances the idea that irrespective of the controversial nature of fake news policing, governments should be responsible for containing this phenomenon, within their human rights boundaries. In this way, they will strike a fair

---

<sup>235</sup> WHO (n 89).

<sup>236</sup> Milanovic (n 59).

<sup>237</sup> Oxford Dictionary (n 70).



balance between freedom of expression and ‘the pressing social need’ for combatting fake news.<sup>238</sup> In its efforts, the government should cooperate with private companies. Finally, the fourth chapter outlined the technologies which have been developed thus far for the containment of misinformation. Automated fact-checking can be considered a natural produce digitalisation. However, until automated fact-checking has not acquired a more secure and reliable level of autonomy, there should be parallel manual fact-checking, with the former carrying out more mechanical tasks to optimise the deployment of the latter.

The tension between freedom of expression and regulation of misinformation can perhaps be best represented by the words of Martti Koskenniemi ‘*to preserve freedom, order must be created to restrict it*’.<sup>239</sup> This notion is echoed by the French President Emmanuel Macron, who asserted that ‘*the internet must be regulated (...) this is the sine qua non condition for a free, open and secure internet*’, during the drafting and subsequent implementation of the *Law No. 2018-1202 of 22 December 2018 on combating the manipulation of information*.<sup>240</sup>

In conclusion, to protect fundamental rights whilst also containing misinformation, a system of checks and balances must be developed. State actors and Service Providers should promote digital literacy aimed at capacity-building and deploy emerging technologies to curb fake news. This research shows that governments shall provide a regulatory framework for private companies to participate in this endeavour, starting from the identification of a viable definition of fake news. Therefore, whilst governments ought to intervene to contain the infodemic, it is fundamental to advance digital literacy in parallel in order to secure long-term results in the eradication of misinformation. It will enable end-users to identify the different types of fake news independently, progressively rendering the controversial heavy policing of the internet redundant. This combination will provide a solution to the grand challenge of the 21<sup>st</sup> century.

---

<sup>238</sup> ICCPR Art 19.3; ECHR Art 10.2; Leerssen (n 14) 102.

<sup>239</sup> Martti Koskenniemi, *From Apology to Utopia: The Structure of International Legal Argument* (2<sup>nd</sup> edn, CUP 2005) 71.

<sup>240</sup> Fiorentino (n 99).



## 6. Bibliography

---

### Legislation

#### International

Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III))  
(UDHR) Art 19

European Convention on Human Rights (adopted 4 November 1950, entered into force 3  
September 1953) ETS 5 Art 10,15

International Covenant on Civil and Political Rights (adopted on 16 December 1966, entered  
into force on 23 March 1976) 999 UNTS 171 Art 4,19,22,25,27

#### National Laws

British Freedom of Information Act 1967

American Freedom of Information Act 2000

Gesetz zur Regelung des Zugangs zu Informationen des Bundes (issued on 5 September 2005, last  
amended 7 August 2013)

LOI n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information

Egyptian Penal Code Law No 58 of 1937 and Its Amendments, Art 80C

Egyptian Law 180 of 2018 Regulating the Press, Media, and the Supreme Council for Media  
Regulation

Egyptian Law No. 175/2018 on Combating Information Technology Crime

2009 Reglamento General a La Ley Orgánica de Comunicación (as amended in 2019)

### Case Law

#### ECtHR

*Leander v Sweden* [1987] 9 EHRR 433

*Gaskin v UK* [1989] 12 EHRR 36

*Guerra v Italy* [1998] 26 EHRR 357

#### Human Rights Committee

*Robert Faurisson v France* (1993) 58<sup>th</sup> Session of Human Rights Committee 550

*Robert W Gauthier v Canada* (1999) 65<sup>th</sup> Session of Human Rights Committee 633



*Rakhim Mavlonov and Mr. Shansiy Sa'di v Uzbekistan* (2009) 95<sup>th</sup> Session of Human Rights Committee 1334

### American Case Law

*Whitney v People of State of California* [1927] 274 U.S. 357

### UN Documents

UNGA Res 59/1946 (14 December 1946) UN Doc A/RES/59

UN Draft Convention on Freedom of Information (adopted 13 February 1950) UN Doc E/RES/278(X)

UN ECOSOC, 'Report by the Rapporteur on Freedom of Information S.P. Lopez 1953/12' (1953)

OHCHR, 'General Comment No. 10: Article 19 (Freedom of opinion and expression)' UN Doc HRI/GEN/1/Rev.9 (1983)

UN Special Rapporteur Hussain also noted that lack of access to information undermines peoples' ability to exercise other freedoms: UN ECOSOC 'Question of The Human Rights Of All Persons Subjected To Any Form Of Detention Or Imprisonment' (14 December 1994) UN Doc E/CN.4/1995/32

UNHRC 'General Comment 34 on Article 19: Freedoms of opinion and expression' (12 September 2011) UN Doc CCPR/C/GC/34 GC 34

Joint Declaration on Freedom of Information and "Fake News", Disinformation and Propaganda (3 March 2017) UN Doc FOM.GAL/3/17

### Books

Koskeniemi M, *From Apology to Utopia: The Structure of International Legal Argument* (2<sup>nd</sup> edn, CUP 2005)

Napoli P, *Social Media and the Public Interest* (Columbia University Press, 2019)

### Articles

Brandtzaeg PB, Følstad A & Chaparro Domínguez MA, 'How Journalists and Social Media Users Perceive Online Fact-Checking and Verification Services' (2018) 12(9) *Journalism Practice* 1109  
Canham ED, 'International Freedom of Information' (1949) 14(4) *Law and Contemporary Problems* 584

Danchin PG et al, 'The Pandemic Paradox in International Law' (2020) ANU College of Law Research Paper No 20.18 [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3675519](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3675519)

Gelfert A, 'Fake News: A Definition' (2018) 38(1) *Informal Logic* 84



Leerssen P, 'Cut Out by the Middle Man: The Free Speech Implications of Social Network Blocking and Banning in the EU' (2015) 6 *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 99

McGonagle T, "'Fake news": False fears or real concerns?' (2017) 35(4) *Netherlands Quarterly of Human Rights* 203

McGonagle T, 'The development of freedom of expression and information within the UN: leaps and bounds or fits and starts?' in Tarlach McGonagle and Yvonne Donders (eds) *The United Nations and Freedom of Expression and Information* (CUP 2015)

O'Flaherty M, 'International Covenant on Civil and Political Rights: interpreting freedom of expression and information standards for the present and the future' in Tarlach McGonagle and Yvonne Donders (eds) *The United Nations and Freedom of Expression and Information* (CUP 2015)

Schwanholz J and Graham T, 'Digital Transformation: New Opportunities and Challenges for Democracy?' in Julia Schwanholz et al (eds) *Managing Democracy in the Digital Age: Internet Regulation, Social Media Use, and Online Civic Engagement* (Springer 2017)

Yemini M, 'The New Irony of Free Speech' (2018) 20 *Colum Sci & Tech L Rev* 119

### Reports and Factsheets

Brennen JS et al *Types, sources, and claims of COVID-19 misinformation* (Reuters Institute and University of Oxford, 7 April 2020) <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>

Graves L, *Understanding the Promise and Limits of Automated Fact-Checking* (Reuters Institute and University of Oxford, February 2018)

Newman N et al, *Reuters Institute Digital News Report 2020* (Reuters Institute and University of Oxford, 2020)

Pew Research Center, *Newspapers Fact Sheet* (Pew Research Center, 9 July 2019)

<https://www.journalism.org/fact-sheet/newspapers/>

US Department of State, *2019 Country Reports on Human Rights Practices: Ecuador* [state.gov/reports/2019-country-reports-on-human-rights-practices/ecuador/](https://state.gov/reports/2019-country-reports-on-human-rights-practices/ecuador/)



## Websites and Blogs

Acosta J, 'How Trump's 'fake news' rhetoric has gotten out of control' *CNN* (11 June 2019)

<https://edition.cnn.com/2019/06/11/politics/enemy-of-the-people-jim-acosta-donald-trump/index.html>

Agence France-Press, 'Facebook says it rejected 2.2m ads for breaking political campaigning rules'

*The Guardian* (18 October 2020) <https://www.theguardian.com/technology/2020/oct/18/facebook-says-it-rejected-22m-ads-seeking-to-obstruct-voting-in-us-election>

Alba D and Plambeck J, 'What You Need to Know About the Election and Disinformation' *NYT* (16

November 2020) <https://www.nytimes.com/2020/11/16/us/election-disinformation.html>

AllJazeera News Agencies, 'Egypt: Court orders release of journalist Mahmoud Hussein' *AllJazeera* (23

May 2019) <https://www.aljazeera.com/news/2019/05/egyptian-court-orders-release-mahmoud-hussein-190523100017910.html>

Allyn B 'Nearly Half Of Accounts Tweeting About Coronavirus Are Likely Bots' *NPR* (20 May 2020)

<https://www.npr.org/sections/coronavirus-live-updates/2020/05/20/859814085/researchers-nearly-half-of-accounts-tweeting-about-coronavirus-are-likely-bots>

Anderson J and Rainie L, 'The future of fake news, according to experts' *World Economic Forum* (1

November 2017) <https://www.weforum.org/agenda/2017/11/this-is-what-experts-think-will-happen-to-fake-news>

Automated Fact Checking, *Full Fact* <https://fullfact.org/automated/>

Bälz K and Mujally H, 'The New Egyptian Anti-Cybercrime Law Regulates Legal Responsibility for Web Pages and Their Content' *Lexology* (19 September 2018)

<https://www.lexology.com/library/detail.aspx?g=90440972-f53e-46dd-b225-7f7cbdea7d73>

BBC, 'ABC figures: Newspapers will no longer have to publish sales' *BBC* (21 May 2020)

<https://www.bbc.com/news/entertainment-arts-52754762>

BBC, 'Independent to cease as print edition' *BBC* (12 February 2016)

<https://www.bbc.com/news/uk-35561145>

Benkler Y et al, 'Mail-In Voter Fraud: Anatomy of a Disinformation Campaign' (*Berkman Klein Center*,

1 October 2020) <https://cyber.harvard.edu/publication/2020/Mail-in-Voter-Fraud-Disinformation-2020>

Bickert M, 'Enforcing Against Manipulated Media' *Facebook* (6 January 2020)

<https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/>

Booth R, 'Journalists to use 'immune system' software against fake news' *The Guardian* (8 August

2017) <https://www.theguardian.com/technology/2017/aug/08/fake-news-full-fact-software-immune-system-journalism-soros-omidyar>



- Booth R, 'Journalists to use "immune system" software against fake news' *The Guardian* (8 August 2017) <https://www.theguardian.com/technology/2017/aug/08/fake-news-full-fact-software-immune-system-journalism-soros-omidyar>
- Burgess M, 'Google is helping Full Fact create an automated, real-time fact-checker' *Wired* (17 November 2016). <https://www.wired.co.uk/article/automated-fact-checking-full-fact-google-funding>
- Coldewey D, 'Facebook adds labels identifying state-controlled media' *TechCrunch* (4 June 2020) <https://techcrunch.com/2020/06/04/facebook-adds-labels-identifying-state-controlled-media/>
- Coleman A, "'Hundreds dead" because of Covid-19 misinformation' *BBC* (12 August 2020) <https://www.bbc.com/news/world-53755067>
- Committee to Protect Journalists, 'Egypt arrests Al-Jazeera producer on fake news charge' *CPJ* (27 December 2016) <https://cpj.org/2016/12/egypt-arrests-al-jazeera-producer-on-fake-news-cha/>
- Confessore N, 'Cambridge Analytica and Facebook: The Scandal and the Fallout So Far' *NYT* (4 April 2018) <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>
- Culliford E and Paul K, 'With fact-checks, Twitter takes on a new kind of task' *Reuters* (31 May 2020) <https://www.reuters.com/article/us-twitter-factcheck-idUSKBN2360U0>
- Daws R, 'Twitter's latest acquisition tackles fake news using AI' *AI News* (4 June 2019) <https://artificialintelligence-news.com/2019/06/04/twitter-acquisition-fake-news-ai/>
- Dunn J, 'Americans aren't sure who to blame for the 'fake news' problem' *Business Insider* (23 December 2016) <https://www.businessinsider.com/fake-news-survey-pew-facebook-chart-2016-12>
- Dzhanova Y, 'Joe Biden and Kamala Harris changed their Twitter bios to reflect their White House victories' *Business Insider* (7 November 2020) <https://www.businessinsider.com/biden-harris-change-twitter-bios-to-reflect-white-house-victories-2020-11?IR=T>
- Eadicicco L, 'Satirical websites are testing Facebook's policy on not being the 'arbiter of truth' by running false headlines claiming Mark Zuckerberg is dead or abusive' *Business Insider* (29 May 2020) <https://www.businessinsider.com/fake-headlines-mark-zuckerberg-satire-test-facebook-policies-2020-5?IR=T>
- Facebook, 'Deepfake Detection Challenge Dataset' *AI Facebook* (25 June 2020) <https://ai.facebook.com/datasets/dfdc/>
- Facebook, 'Fact Checking on Facebook' <https://www.facebook.com/business/help/2593586717571940>
- Florentino MR, 'France passes controversial "fake news" law' *Euronews* (22 November 2018) <http://euronews.com/2018/11/22/france-passes-controversial-fake-news-law>



Fraser M, 'Debate: The legal fight against "fake news" must not veer into censorship' *The Conversation* (11 June 2018) <https://theconversation.com/debate-the-legal-fight-against-fake-news-must-not-veer-into-censorship-98049>

Frenkel S, 'Facebook Says It Deleted 865 Million Posts, Mostly Spam' *NYT* (15 May 2018) <https://www.nytimes.com/2018/05/15/technology/facebook-removal-posts-fake-accounts.html>

Full Fact, 'Automated Fact Checking' *Full Fact* <https://fullfact.org/automated>

Funke D, 'Automated fact-checking has come a long way. But it still faces significant challenges' *Poynter* (4 April 2018) <https://www.poynter.org/fact-checking/2018/automated-fact-checking-has-come-a-long-way-but-it-still-faces-significant-challenges/>

Funke D, 'Here's how close automated fact-checking is to reality' *Poynter* (28 February 2018) <https://www.poynter.org/fact-checking/2018/heres-how-close-automated-fact-checking-is-to-reality/>

Funke D, 'In Argentina, fact-checkers' latest hire is a bot' *Poynter* (11 January 2018) <https://www.poynter.org/fact-checking/2018/in-argentina-fact-checkers%C2%92-latest-hire-is-a-bot/>

Gray R, 'Lies, propaganda and fake news: A challenge for our age' *BBC* (1 March 2017) <https://www.bbc.com/future/article/20170301-lies-propaganda-and-fake-news-a-grand-challenge-of-our-age>

Greenblatt A, 'What does the future of automated fact-checking look like?' *Poynter* (4 April 2016) <https://www.poynter.org/fact-checking/2016/whats-does-the-future-of-automated-fact-checking-look-like/>

Henriksson T, 'Article 19: legal restrictions to "fake news" could be used to limit press freedom' *WAN IFRA* (29 November 2016) <https://blog.wan-ifra.org/2016/11/29/article-19-legal-restrictions-to-fake-news-could-be-used-to-limit-press-freedom>

Horowitz J, 'In Italian Schools, Reading, Writing and Recognising Fake News' *NYT* (18 October 2017) <https://www.nytimes.com/2017/10/18/world/europe/italy-fake-news.html>

Human Rights Watch, *Freedom Of Expression On The Internet* <https://www.hrw.org/legacy/wr2k/issues-04.htm#TopOfPage>

Kolodny L, 'Full Fact aims to end fake news with automated fact checking tools' *TechCrunch* (17 November 2016) <https://techcrunch.com/2016/11/17/full-fact-scores-google-grant-to-put-an-end-to-fake-news/?guccounter=1>

Lomas N, 'Fabula AI is using social spread to spot "fake news"' *TechCrunch* (7 February 2019) <https://techcrunch.com/2019/02/06/fabula-ai-is-using-social-spread-to-spot-fake-news/>



- Lomas N, 'Microsoft launches a deepfake detector tool ahead of US election' *TechCrunch* (2 September 2020) <https://techcrunch.com/2020/09/02/microsoft-launches-a-deepfake-detector-tool-ahead-of-us-election/>
- Luscombe R, "'Chill!': Greta Thunberg recycles Trump's mockery of her as he tries to stop votes' *The Guardian* (5 November 2020) <https://www.theguardian.com/environment/2020/nov/05/greta-thunberg-donald-trump-twitter-chill>
- Lyons T, 'Increasing Our Efforts to Fight False News' *Facebook* (21 June 2018) <https://about.fb.com/news/2018/06/increasing-our-efforts-to-fight-false-news/>
- Makszimov V, 'Hungary's state of emergency law is a 'blank cheque' to Orban, critics say' *Euractiv* (23 March 2020) <https://www.euractiv.com/section/justice-home-affairs/news/hungarys-state-of-emergency-law-is-a-blank-cheque-to-orban-critics-say/>
- Makszimov V, 'Hungary's state of emergency law is a 'blank cheque' to Orban, critics say' *Euractiv* (23 March 2020) <https://www.euractiv.com/section/justice-home-affairs/news/hungarys-state-of-emergency-law-is-a-blank-cheque-to-orban-critics-say/>
- McLennan D and Miles J, 'A once unimaginable scenario: No more newspapers' *Washington Post* (21 March 2018) <https://www.washingtonpost.com/news/theworldpost/wp/2018/03/21/newspapers/?noredirect=on>
- Mendel T, 'Freedom of Information as an Internationally Protected Human Right' *Article 19* <https://www.article19.org/data/files/pdfs/publications/foi-as-an-international-right.pdf>
- Middleton C, 'Facebook trained AI with 3.5 billion hashtagged Instagram photos' *Internet of Business* <https://internetofbusiness.com/facebook-training-ai-instagram/>
- Milanovic M and Schmitt MN, 'Cyber Attacks and Cyber (Mis)information Operations during a Pandemic' *EJIL: Talk!* (29 March 2020) <https://www.ejiltalk.org/cyber-attacks-and-cyber-misinformation-operations-during-a-pandemic/>
- Milanovic M, 'Viral Misinformation and the Freedom of Expression: Part I' *EJIL: Talk!* (13 April 2020) <https://www.ejiltalk.org/viral-misinformation-and-the-freedom-of-expression-part-i/>
- Milanovic M, 'Viral Misinformation and the Freedom of Expression: Part II' *EJIL: Talk!* (13 April 2020) <https://www.ejiltalk.org/viral-misinformation-and-the-freedom-of-expression-part-ii/>
- Mong A, 'Countering fake news while safeguarding free speech' *DW* (14 March 2017) <https://www.dw.com/en/countering-fake-news-while-safeguarding-free-speech/a-37924278>
- Murison M, 'Facebook using machine learning to fight fake news' *Internet of Business* (25 June 2018) <https://internetofbusiness.com/facebook-machine-learning-fake-news/amp/>



NPR Staff, 'What Legal Recourse Do Victims Of Fake News Stories Have?' *NPR* (7 December 2016)

<https://www.npr.org/2016/12/07/504723649/what-legal-recourse-do-victims-of-fake-news-stories-have?t=1598346365344>

Ortiz-Ospina E, 'The rise of social media' *Our world in data* (18 September 2019)

<https://ourworldindata.org/rise-of-social-media>

Roden L, 'Swedish kids to learn computer coding and how to spot fake news in primary school' *The Local SE* (13 March 2017)

<https://www.thelocal.se/20170313/swedish-kids-to-learn-computer-coding-and-how-to-spot-fake-news-in-primary-school>

Roose K, 'On Election Day, Facebook and Twitter Did Better by Making Their Products Worse' *NYT* (5 November 2020)

<https://www.nytimes.com/2020/11/05/technology/facebook-twitter-election.html>

Santana B, 'Jair Bolsonaro accused me of spreading 'fake news'. I know why he targeted me' *The Guardian* (22 June 2020)

<https://www.theguardian.com/commentisfree/2020/jun/22/jair-bolsonaro-fake-news-accusation-marielle-franco>

Scott M, 'Russia and China push "fake news" aimed at weakening Europe: report' *Politico* (1 April 2020)

<https://www.politico.eu/article/russia-china-disinformation-coronavirus-covid19-facebook-google/>

Shearer E, 'Social media outpaces print newspapers in the U.S. as a news source' *Pew Research Center* (10 December 2018)

<https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/>

Smith N, 'Schoolkids in Taiwan Will Now Be Taught How to Identify Fake News' *Time* (7 April 2017)

<https://time.com/4730440/taiwan-fake-news-education/>

Some K, 'Facebook Uses AI To Fight Coronavirus Misinformation And Fake News' *Analytics Insight* (13 June 2020)

<https://www.analyticsinsight.net/facebook-uses-ai-fight-coronavirus-misinformation-fake-news/>

Talbot P, 'Twitter Removes Thousands Of Accounts For Manipulating Its Platform' *NPR* (20 September 2019)

<https://www.npr.org/2019/09/20/762799187/twitter-removes-thousands-of-accounts-for-manipulating-their-platform?t=1597910217677>

Troianovski A, 'Fighting False News in Ukraine, Facebook Fact Checkers Tread a Blurry Line' *NYT* (26 June 2020)

<https://www.nytimes.com/2020/07/26/world/europe/ukraine-facebook-fake-news.html>

Twitter, 'Platform manipulation and spam policy' *Twitter* (September 2019)

<https://help.twitter.com/en/rules-and-policies/platform-manipulation>

UN Department of Global Communications, 'UN tackles "infodemic" of misinformation and cybercrime in COVID-19 crisis' *UN*

<https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%99infodemic%E2%80%99-misinformation-and-cybercrime-covid-19>



WHO, '1st WHO Infodemiology Conference' *WHO* <https://www.who.int/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference>

Wong JC and Levine S, 'Twitter labels Trump's false claims with warning for first time' *The Guardian* (27 May 2020) <https://www.theguardian.com/us-news/2020/may/26/trump-twitter-fact-check-warning-label>

Woolley S, 'We're fighting fake news AI bots by using more AI. That's a mistake' *MIT Technology Review* (8 January 2020) <https://www.technologyreview.com/2020/01/08/130983/were-fighting-fake-news-ai-bots-by-using-more-ai-thats-a-mistake/>

World Economic Forum, 'Digital Wildfires in a Hyperconnected World' *World Economic Forum* <http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/>

## Others

'Twitter Usage Statistics' *Internet Live Stats* <https://www.internetlivestats.com/twitter-statistics/> accessed on 27<sup>th</sup> November 2020

Atasanov P, 'Fake News between Artificial Intelligence and Credibility' (PhD thesis, Goce Delcev University 2018) 1020

BBC Reality Check, *BBC* [bbc.com/news/reality\\_check](http://bbc.com/news/reality_check)

Chequeado, 'Chequeabot' <https://chequeabot.chequeado.com/desgrabador/>

Collazzo M and Tyan A, 'Emergency Powers, COVID-19 and the New Challenge for Human Rights' (2020) 20(48) *IAI Commentaries* 1

IPI tracker on press freedom violations linked to COVID-19 coverage, 'COVID-19: Number of Media Freedom Violations by Region' *IPI* <https://ipi.media/covid19-media-freedom-monitoring/>

Oxford Dictionary, *Disinformation* <https://www.lexico.com/definition/disinformation>

Oxford Dictionary, *Misinformation* <https://www.lexico.com/definition/misinformation?s=t>

Pérez-Latre FJ, Sánchez-Taberner A, García-Mansilla G "'Barriers to entry" in the media industry: the launching of El Español and Netflix', University of Navarra (2-5 June 2016)

## List of Images

**Graph 1: Percentage of people using social platforms**

We Are Social USA - Global socially-led creative agency (2020). Available at: <https://wearesocial.com/>

**Graph 2: Traditional print media compared to social media for news**



*Social media users pass the 4 billion mark as global adoption soars - We Are Social (2020). Available at: <https://wearesocial.com/blog/2020/10/social-media-users-pass-the-4-billion-mark-as-global-adoption-soars>*

**Graph 3: People using social media for news as of January 2015 compared to 2020**

Reuters Institute Digital News Report (2020). Available at: <https://www.digitalnewsreport.org/>

**Table 1: Types of Fake News**

*Subject Guides: Fake News: Types of Fake News (2020). Available at:*

<https://guides.lib.uiowa.edu/c.php?g=849536&p=6077637>

**Graph 4: 'Fake News Stories Are a Problem - But Who's to Blame?'**

Richter, F. (2016) *Infographic: Fake News Stories Are a Problem - But Who's to Blame?*, Statista Infographics. Available at: <https://www.statista.com/chart/7305/public-opinion-on-fake-news/>

**Image 1: Laws implemented by countries**

Nazzah et al. (2018) *China's Lessons for Fighting Fake News*, Foreign Policy. Available at:

<https://foreignpolicy.com/2018/09/06/chinas-lessons-for-fighting-fake-news/>

Nationale, A. (2020) *Assemblée nationale ~ Les députés, le vote de la loi, le Parlement*

*français, Assemblée nationale*. Available at: <http://www.assemblee-nationale.fr/>

Борисевич, К. (2018) "Иногда форумы — грязь". *Обсуждается введение уголовной ответственности за недостоверную информацию*, TUT.BY. Available at:

<https://news.tut.by/society/595733.html>

*Kenya signs bill criminalising fake news* (2018). Available at: <https://mg.co.za/article/2018-05-16-kenya-signs-bill-criminalising-fake-news/>

*Cambodia 'fake news' laws tighten noose on press freedom - International Press Institute* (2019).

Available at: <https://ipi.media/cambodia-fake-news-laws-tighten-noose-on-press-freedom/>

*Brown signs bill to encourage media literacy in schools* (2018). Available at:

<https://www.pressdemocrat.com/article/news/california-now-has-a-law-to-bolster-media-literacy-in-schools/?sba=AAS>

*TechCrunch is now a part of Verizon Media* (2020). Available at:

<https://techcrunch.com/2019/05/09/singapore-fake-news-law/>

*The New Egyptian Anti-Cybercrime Law Regulates Legal Responsibility for Web Pages and Their*

*Content | Lexology* (2020). Available at: <https://www.lexology.com/library/detail.aspx?g=90440972-f53e-46dd-b225-7f7cbdea7d73>



*TIMEP Brief: The Law Regulating the Press, Media, and the Supreme Council for Media Regulation* (2020). Available at: <https://timep.org/reports-briefings/timep-brief-the-law-regulating-the-press-media-and-the-supreme-council-for-media-regulation/>

(2020) *Ilo.org*. Available at: <https://www.ilo.org/dyn/natlex/docs/ELECTRONIC/57560/111585/F1337119832/EGY57560.pdf>

*Taiwan Legal Update: Coronavirus and Misinformation - ASIALLIANS* (2020). Available at: <http://asiallians.com/en/taiwan-legal-update-coronavirus-misinformation/>

*In the name of 'fake news,' Asian governments tighten control on social media* (2019). Available at: <https://www.latimes.com/world/asia/la-fg-asia-fake-news-20190130-story.html>

*New 'fake news' law stifles independent reporting in Russia on COVID-19 - International Press Institute* (2020). Available at: <https://ipi.media/new-fake-news-law-stifles-independent-reporting-in-russia-on-covid-19/>

*Cambodia 'fake news' crackdown prompts fears over press freedom* (2018). Available at: [https://www.theguardian.com/world/2018/jul/06/cambodia-fake-news-crackdown-prompts-fears-over-press-freedom?CMP=share\\_btn\\_tw](https://www.theguardian.com/world/2018/jul/06/cambodia-fake-news-crackdown-prompts-fears-over-press-freedom?CMP=share_btn_tw)

Correia, M. and Viana, N. (2018) *20 projetos de lei no Congresso pretendem criminalizar fake news - Agência Pública, Agência Pública*. Available at: <https://apublica.org/2018/05/20-projetos-de-lei-no-congresso-pretendem-criminalizar-fake-news/>

Vladin 'lex Facebook' lako bi se mogao pretvoriti u masovnu cenzuru (2018). Available at: <https://faktograf.hr/2018/01/16/vladin-lex-facebook-lako-bi-se-mogao-pretvoriti-u-masovnu-cenzuru/>

STORY, B. (2020) *Govt planning to amend IT Act to crack down on apps, sites unable to curb fake news, BusinessToday.in*. Available at: <https://www.businesstoday.in/top-story/govt-planning-to-amend-it-act-to-crack-down-on-apps-sites-unable-to-curb-fake-news/story/306041.html>

*In East and Southeast Asia, misinformation is a visible and growing concern - Poynter* (2018). Available at: <https://www.poynter.org/fact-checking/2018/in-east-and-southeast-asia-misinformation-is-a-visible-and-growing-concern/>

*South Korea: Act of Spreading False Information via Internet Not Punishable | Global Legal Monitor* (2020). Available at: <https://www.loc.gov/law/foreign-news/article/south-korea-act-of-spreading-false-information-via-internet-not-punishable/>

*The Framework Act on Telecommunications (English translation)* (2020). Available at: <https://www.sec.gov/Archives/edgar/data/892450/000119312512188247/d333931dex151.htm>

Article 47 (Penal Provisions) (1) A person who has publicly made a false communication over the telecommunications facilities and equipment for the purpose of harming the public interest shall be



punished by imprisonment for not more than five years or by a fine not exceeding fifty million won.

<Amended by Act No. 5219, Dec. 30, 1996>

*“Fake news”: autoridades nacionales y regionales podrían tener nueva causal para la cesación de cargos - Senado - República de Chile (2020).* Available at: <https://www.senado.cl/fake-news-autoridades-nacionales-y-regionales-podrian-tener-nueva/senado/2019-01-29/122725.html>