

ARTICLE

Governing Fake News: The Regulation of Social Media and the Right to Freedom of Expression in the Era of Emergency

Donato Vese*

University of Turin, Department of Law, Turin, Italy. Email: donato.vese@unito.it

Abstract

Governments around the world are strictly regulating information on social media in the interests of addressing fake news. There is, however, a risk that the uncontrolled spread of information could increase the adverse effects of the COVID-19 health emergency through the influence of false and misleading news. Yet governments may well use health emergency regulation as a pretext for implementing draconian restrictions on the right to freedom of expression, as well as increasing social media censorship (ie chilling effects). This article seeks to challenge the stringent legislative and administrative measures governments have recently put in place in order to analyse their negative implications for the right to freedom of expression and to suggest different regulatory approaches in the context of public law. These controversial government policies are discussed in order to clarify why freedom of expression cannot be allowed to be jeopardised in the process of trying to manage fake news. Firstly, an analysis of the legal definition of fake news in academia is presented in order to establish the essential characteristics of the phenomenon (Section II). Secondly, the legislative and administrative measures implemented by governments at both international (Section III) and European Union (EU) levels (Section IV) are assessed, showing how they may undermine a core human right by curtailing freedom of expression. Then, starting from the premise of social media as a “watchdog” of democracy and moving on to the contention that fake news is a phenomenon of “mature” democracy, the article argues that public law already protects freedom of expression and ensures its effectiveness at the international and EU levels through some fundamental rules (Section V). There follows a discussion of the key regulatory approaches, and, as alternatives to government intervention, self-regulation and especially empowering users are proposed as strategies to effectively manage fake news by mitigating the risks of undue interference by regulators in the right to freedom of expression (Section VI). The article concludes by offering some remarks on the proposed solution and in particular by recommending the implementation of reliability ratings on social media platforms (Section VII).

*I wish to acknowledge the anonymous reviewers for valuable comments and suggestions on an earlier draft of this article. This article presents the results of the annual research I carried out thanks to the research fellowship from the University of Turin – Department of Law (supervisor Dr Valeria Ferraris). The first draft of the article was inspired by the international conference “Covid 19, pandémies et crises sanitaires, quels apports des sciences humaines et sociales?” held on 16–17 December 2020 at the École Supérieure de technologie d’Essaouira. I wish to thank the coordinator of the conference, Professor Mohamed Boukherouk, for inviting me to that important event, and I am also grateful for suggestions and comments from colleagues at the international conference: Mélissa Moriceau, Hilaire Akerekoro, Abdelkader Behtane, Miterand Lienou and Yanick Hypolitte Zambo.

I. Introduction

In one of the masterpieces of contemporary literature, *The Prague Cemetery*,¹ Simone Simonini (and his alter ego the Abbé Dalla Piccola) is a cynical forger who continuously produces and sells fake news. Simonini fakes documents not only for private individuals but also for the police, the secret services and even the State. He shapes the story to his liking, transforming it with his false documents, often based on the decontextualisation of existing documents and therefore originating from true facts. It is the principle of likelihood that makes this news plausible.

Eco's novel shows how reality can be altered, changed and even created by words. The protagonist, Simonini, demonstrates this, as does a historical case cited in the text, "The Protocols of the Elders of Zion", a fake that spread throughout Europe, being exploited by Hitler and even gaining credence among enlightened men such as Henry Ford. Eco describes the process underlying so-called fake news as a real problem for society. Its presence on the Internet, from the most harmless to the most pernicious, is growing rapidly and is increasingly difficult to recognise.

Yet Eco's novel also shows the importance of freedom of expression and opinion as the cornerstone of democratic society.

There is no doubt that disinformation and particularly fake news can pose considerable risks to society. Spreading fake news on a major social media platform could cause serious concern for users and negative externalities with ripple effects for people.

Nevertheless, freedom of expression may come at a price, and perhaps we are paying it now in this time of crisis through the spread of fake news on social media. However, it is one that we should be willing to pay without undue hesitation as the future of the democracy in which we believe is at stake.

Furthermore, is it not true that we would challenge government restrictions on the right to economic freedom, as some are already doing? The same holds true for freedom of movement and other human rights. By contrast, especially in the context of the current health emergency, we are observing how governments address fake news on social media by adopting regulatory policies that can significantly undermine freedom of expression; and this could also be a pretext for limiting freedom of expression in the future.

In this article, I seek to challenge the stringent emergency legislative and administrative measures on fake news that governments have recently put in place in order to analyse their negative implications for the right to freedom of expression and to suggest a possible solution in the context of public law.

I shall discuss these controversial government policies in order to clarify why we cannot allow freedom of expression to be jeopardised in the process of trying to manage the risks of fake news.

I start with an examination of the legal definition of fake news in academia in order to establish the essential characteristics of the phenomenon (Section II).

Secondly, I assess the legislative and administrative measures implemented by governments at both the international and European Union (EU) levels (Sections III and IV, respectively), showing how they may undermine a core human right by curtailing freedom of expression.

Then, starting from the premise of social media as a "watchdog" of democracy and moving on to the contention that fake news is a phenomenon of "mature" democracy, I will argue that public law already protects freedom of expression and ensures its effectiveness at the international and EU levels through some fundamental rules (Section V).

Lastly, I explore key regulatory approaches and, as alternatives to government intervention, I propose self-regulation and above all empowering users as strategies to manage

¹ U Eco, *The Prague Cemetery*, tr. R Dixon (London, Harvill Secker/New York, Houghton Mifflin 2011).

fake news by mitigating the risks of undue interference in the right to freedom of expression (Section VI).

In so doing, I conclude by offering some remarks on the proposed solution and in particular by recommending the implementation of reliability ratings on social media platforms (Section VII).

II. What is fake news? A definition

What is fake news? This is quite a challenging question for legal scholars, so a preliminary task is to establish a definition and identify some of the key problems raised by fake news in academic debate and at the institutional level. This section aims to do just that.

To this end, it could be argued that the definition of fake news is part of the more general notion of disinformation.² For this reason, when discussing fake news in this article, I will often refer to the broader phenomenon of disinformation.³ On this topic, it may be understood that fake news is a recent phenomenon, and both the legislations and the measures of the authorities, especially the European ones, are used to talk about disinformation including fake news in this legal context.

There is no universally agreed-upon definition of fake news. Indeed, as we shall see, scholars have suggested various meanings for the term, and a definition has recently been proposed as “the online publication of intentionally or knowingly false statements of fact”.⁴

Academics have also defined fake news as lies, namely deliberately false statements of fact distributed via news channels.⁵ Nevertheless, as has rightly been noted, current usage is not yet settled, and there are clearly different types of fake news that should not be conflated for legal purposes.⁶

For other scholars, fake news should be limited to articles that suggest, through appearance and content, the conveyance of real news, but which also knowingly include at least one material factual assertion that is empirically verifiable as false and that is not otherwise protected by the fair report privilege,⁷ even if these are often understood as fabricated news stories.⁸

Moreover, according to another definition, fake news is information that has been deliberately fabricated and disseminated with the intention of deceiving and misleading others into believing falsehoods or doubting verifiable facts.⁹

² See, eg, A Renda, “The legal framework to address ‘fake news’: possible policy actions at the EU level” (2018) European Parliament, 10–18 also available online at <[https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/619013/IPOL_IDA\(2018\)619013_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/619013/IPOL_IDA(2018)619013_EN.pdf)>.

³ The term “fake news” may be widely recognised in public debate, but some academic and above all policy sources generally advise against using it, recommending “disinformation” instead. While “misinformation” refers to material that is simply erroneous (eg due to error or ignorance), “disinformation” implies an intentional, malicious attempt to mislead.

⁴ DO Klein and JR Wueller, “Fake news: a legal perspective” (2017) 20(10) *Journal of Internet Law* 6, also available at SSRN: <<https://ssrn.com/abstract=2958790>>.

⁵ M Verstraete et al, “Identifying and countering fake news” (2017) *Arizona Legal Studies Discussion Paper* no. 17-15, 5–9.

⁶ B Baade, “Fake news and international law” (2018) 29(4) *European Journal of International Law* 1358 <<https://doi.org/10.1093/ejil/chy071>>.

⁷ C Calvert, S McNeff, A Vining and S Zarate, “Fake news and the First Amendment: reconciling a disconnect between theory and doctrine” (2018) 86 (99) *University of Cincinnati Law Review* 103 <<https://scholarship.law.uc.edu/uclr/vol86/iss1/3>>.

⁸ A Park and KH Youm, “Fake news from a legal perspective: the United States and South Korea compared” (2019) 25 *Southwestern Journal of International Law* 100–19, at 102.

⁹ T McGonagle, “Fake news: false fears or real concerns?” (2017) 35(4) *Netherlands Quarterly of Human Rights* 203–09 <<https://doi.org/10.1177/0924051917738685>>.

The authors of a well-known article on the US presidential election of 2016 define fake news as news articles that are intentionally and verifiably false and could mislead readers,¹⁰ as well as news stories that have no factual basis but are presented as news.¹¹ Incidentally, it has been appropriately noted that fake news is also the presentation of false claims purporting to be about the world in a format and with content that resembles the format and content of legitimate media organisations.¹²

Furthermore, it has been observed that fake news may also purport to describe events in the real world, typically by mimicking the conventions of traditional media reportage, yet it is known by its creators to be significantly false and is transmitted with the dual goal of being widely retransmitted and of deceiving at least some of its audience.¹³ Beyond this, it has been argued that fake news is best defined as the deliberate presentation of (typically) false or misleading claims as news, where the claims are misleading by design.¹⁴

Lastly, from analysis of previous studies that have defined and operationalised the term, other authors have stated that fake news is information constituting viral posts based on fictitious accounts made to look like news reports in contemporary discourse and particularly in media coverage.¹⁵

Outside academia, the definition of fake news has been discussed at the institutional level. In the USA, several important events have been held on the definition of fake news, such as the workshop organised on 7 March 2017 by the Information Society Project at Yale Law School and the Floyd Abrams Institute for Freedom of Expression.¹⁶ During the workshop, news organisations, information intermediaries, data scientists, computer scientists, the practising bar and sociologists explored efforts to define fake news and to discuss the feasibility and desirability of possible solutions.¹⁷

In general, most participants were reluctant to propose negative State regulations for fake news. The option of using government funding or other economic incentives to indirectly promote legitimate news and information outlets was floated, but this was rightly critiqued on similar grounds to those associated with government intervention to penalise certain kinds of speech, simply by aiming to prevent the government from determining what is true or worthy. Ultimately, nearly all participants agreed on one overarching

¹⁰ H Allcott and M Gentzkow, "Social media and fake news in the 2016 election" (2017) 31(2) *Journal of Economic Perspectives* 213. However, in defining this concept, they exclude satirical websites such as *The Onion* (<<https://www.theonion.com>>), which uses humour and exaggeration to criticise social and political issues. On this point, for a different opinion, see Klein and Wueller, *supra*, note 4, 6. See also A Bovet and HA Makse, "Influence of fake news in Twitter during the 2016 U.S. presidential election" (2019) 10 *Nature Communications* 7 <<https://doi.org/10.1038/s41467-018-07761-2>>.

¹¹ See Alcott and Gentzkow, *supra*, note 10, 5.

¹² N Levy, "The bad news about fake news" (2017) 6(8) *Social Epistemology Review and Reply Collective* 20 <<http://wp.me/p1Bfg0-3GV>>.

¹³ R Rini, "Fake news and partisan epistemology" (2017) 27(2) *Kennedy Institute of Ethics Journal* E43-E64 <10.1353/ken.2017.0025>.

¹⁴ A Gelfert, "Fake news: a definition" (2018) 38(1) *Informal Logic* 85 and 108 <<https://doi.org/10.22329/il.v38i1.5068>>, where "the phrase 'by design' is intended to reflect that what is novel about fake news – not only, but especially on electronic social media – is its systemic dimension".

¹⁵ EC Tandoc Jr, WL Zheng and R Ling, "Defining 'fake news': a typology of scholarly definitions" (2018) 6(2) *Digital Journalism* 2.

¹⁶ "Fighting Fake News Workshop Report", 3–4 <<https://law.yale.edu/isp/initiatives/floyd-abrams-institute-freedom-expression/practitioner-scholar-conferences-first-amendment-topics/fighting-fake-news-workshop>> (last accessed 9 November 2020).

¹⁷ First of all, the participants determined that the most salient danger associated with fake news is the fact that it devalues and delegitimises the voices of experts, authoritative institutions and the concept of objective data, all of which undermines society's ability to engage in rational discourse based upon shared facts. Secondly, some argued that the difficulty of defining fake news raised the attendant risk of overly broad government regulation, while others worried that opening the door to permitting government sanctions against certain kinds of public discourse would grant the government too much power to control speech in areas of public concern.

conclusion: that re-establishing trust in the basic institutions of a democratic society is critical to combatting the systematic efforts being made to devalue truth. In addition to thinking about how to fight different kinds of fake news, it is necessary to think broadly about how to bolster respect for facts.

In Europe, the European Commission promoted a public consultation from 13 November 2017 to 23 February 2018, which, among other things, afforded some criteria for defining fake news.¹⁸ In particular, organisations were asked to suggest different criteria for defining fake news, including from a legal point of view. The responses highlighted a wide range of criteria, with the consensus that fake news could be defined by looking at: (1) the intent and apparent objectives pursued by fake news; (2) the sources of such news; and (3) the actual content of news.¹⁹

Essentially, this consultation reached a definition of fake news based on the pursued objectives of the news. Thus, the concept would mainly cover online news, albeit sometimes disseminated in traditional media too, intentionally created and distributed to mislead readers and influence their thoughts and behaviour.

Moreover, fake news can polarise public opinion, opinion leaders and media by creating doubts regarding verifiable facts, eventually jeopardising the free and democratic opinion-forming process and undermining trust in democratic processes. Gaining political or other kinds of influence or funds through online advertising (eg clickbait) or causing damage to an undertaking or a person can also be major aims of fake news. The existence of a clear intention behind the fake news would establish the difference between this phenomenon and that of misinformation; that is, where wrong information is provided owing, for instance, to good-faith mistakes or to failure to respect basic journalism standards (eg verification of sources, investigation of facts, etc.).²⁰

Lastly, it should be noted that, in the consultation in question, civil society organisations and news media in particular justifiably criticised the term “fake news” as misleading and with negative connotations (ie used by those who criticise the work of the media or opposing political views). Hence, since fake news may be a symptom of a wider problem, namely the crisis of information, the use of the term “disinformation” was suggested as a more appropriate expression.

In the UK, the House of Commons Digital, Culture, Media and Sport (DCMS) Select Committee published its final report on disinformation and fake news at the end of an eighteen-month inquiry on 18 February 2019. In this Final Report, the DCMS stated that fake news is a poorly defined and misleading term that conflates a variety of false information, from genuine error through to foreign interference in democratic processes. For this reason, the DCMS Select Committee recommended that the government move away from the term “fake news” and instead seek to address disinformation and wider online manipulation by defining disinformation as the deliberate creation and sharing of false or manipulated information that is intended to deceive and mislead audiences, either for the

¹⁸ European Commission, “Public Consultation on Fake News and Online Disinformation” (26 April 2018) <<https://ec.europa.eu/digital-single-market/en/news/public-consultation-fake-news-and-online-disinformation>> (last accessed 10 November 2020). The consultation will collect information on: (1) the definition of fake information and its spread online; (2) the assessment of measures already taken by platforms, news media companies and civil society organisations to counter the spread of fake information online; and (3) the scope for future actions to strengthen the quality of information and to prevent the spread of disinformation online.

¹⁹ European Commission, “Synopsis Report of the Public Consultation on Fake News and Online Disinformation” (26 April 2018), 6 <<https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation-fake-news-and-online-disinformation>> (last accessed 10 November 2020).

²⁰ *ibid.*

purposes of causing harm or for political, personal or financial gain. Conversely, misinformation should refer to the inadvertent sharing of false information.²¹

Starting from the suggested definition of fake news, the DCMS Select Committee called for: (1) a compulsory code of ethics for social media companies overseen by an independent regulator; (2) additional powers for the regulator to launch legal action against companies breaching the code; and (3) a reform of electoral communications laws and rules on overseas involvement in elections. Finally, it also recommended creating (4) an obligation for social media companies to take down known sources of harmful content, including proven sources of disinformation.

In addition, as regards further institutional actions linked to the aforementioned definition of fake news, the UK government has been committed to maintaining a news environment, both online and offline, where accurate content and high-quality news online can prevail. While mechanisms are in place to enforce accuracy and impartiality in the broadcast and press industries, greater regulation in the online space has been considered. It has developed a range of regulatory and non-regulatory initiatives to improve transparency and accountability in the online environment where information is shared. It has also committed to ensuring that freedom of expression in the UK is protected and enhanced online, and this work will be carried out in partnership with industry, the media and civil society institutions.

The brief analysis presented so far shows that the definition of fake news is particularly complex and raises some significant issues.

Essentially, we might reasonably argue that fake news is information designed to emulate characteristics of the media in form but not in substance. Fake news sources typically do not have media editorial policies and procedures for ensuring the correctness and reliability of their information. This, in the majority opinion of academics, would likely make them unreliable and harmful to the public. Fake news adds to other information flaws, namely misinformation, meaning false or misleading information, and disinformation, meaning false information that is deliberately disseminated to deceive people.

Fake news plays a leading role in several areas, such as politics, health and the economy, which means that governments around the world are called upon to regulate the phenomenon.²²

In the next two sections, I shall consider legislative and administrative policies in order to discuss and criticise the robust measures governments have introduced to address fake news at the international and European levels (Sections III and IV, respectively). In doing so, the analysis refers specifically to States adopting particularly stringent policies to address fake news on social media during the COVID-19 pandemic, bearing in mind that policies became more stringent during and after the outbreak.

The analysis of the measures implemented by States around the world is not, therefore, based on their policy of freedom of expression, nor on their way of guaranteeing it to their citizens. On the contrary, emergency legislative and administrative measures are evaluated in order to understand whether or not they fall foul of the international and European rules that protect human rights with regard to freedom of expression.

²¹ House of Commons Digital, Culture, Media and Sport Committee, “Disinformation and ‘Fake News’: Government response to the committee’s fifth report of session 2017–19”, hc 1630 (2018), 2 <<https://publications.parliament.uk/pa/cm201719/cmselect/cmcomeds/1630/1630.pdf>>, archived at <<https://perma.cc/e92s-4ggc>> (last accessed 11 November 2020). It should be noted that, to help provide clarity and consistency, the Digital, Culture, Media and Sport Committee recommended that the government not use the term “fake news” and instead use, and define, the words “misinformation” and “disinformation”. On this point, see C Feikert-Ahalt, “Initiatives to counter fake news in selected countries – United Kingdom” 100–08 <<https://www.loc.gov/law/help/fake-news/index.php>> (last accessed 11 November 2020).

²² See, eg, MR Leiser, “Regulating fake news” (2017), available at <<https://openaccess.leidenuniv.nl/handle/1887/72154>>.

III. International regulatory responses to fake news

All over the world, some governments have issued stringent legislative and administrative measures restricting freedom of expression in order to address disinformation and especially fake news. In this regard, an important factor to consider is that the pandemic has encouraged strict government policies, which, acting under the threat of loss of life, have passed particularly invasive human rights laws to manage the risks of online disinformation.

Generally, these policies could trigger “chilling effects” that could be implemented by governments to build a climate of self-censorship that dissuades democratic actors such as journalists, lawyers and judges from speaking out.²³

It should be noted that in its latest report on “The state of the world’s human rights”, Amnesty International has emphasised the relationship between freedom of expression and fake news.²⁴ The report documented various repressions with criminal sanctions imposed by governments around the world against journalists and social media users.

In a few countries, particularly in Asia and the Middle East and North Africa, authorities prosecuted and even imprisoned human rights defenders and journalists using vaguely worded charges such as spreading misinformation, leaking state secrets and insulting authorities, or labelled them as “terrorists”. Some governments invested in digital surveillance equipment to target them.²⁵ Moreover, public authorities punished those who criticised government actions concerning COVID-19, exposed violations in the response to it or questioned the official narrative around it. Many people were detained arbitrarily and, in some cases, charged and prosecuted. In some countries, the government used the pandemic as a pretext to clamp down on unrelated criticism. In Latin America, disinformation laws that force platforms to decide whether to remove content without judicial orders have been found to be incompatible with Article 13 of the American Convention on Human Rights.²⁶

The United Nations (UN) Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression has recently declared that several States have adopted laws that grant the authorities excessive discretionary powers to compel social media platforms to remove content that they deem illegal, including what they consider to be disinformation or fake news. He has also affirmed how failure to comply is sanctioned with significant fines and content blocking.²⁷ The UN Special Rapporteur has highlighted

²³ For a definition of “chilling effects”, see L Pech, “The concept of chilling effect: its untapped potential to better protect democracy, the rule of law, and fundamental rights in the EU” opensocietyfoundations.org (4 March 2021) <<https://www.opensocietyfoundations.org/uploads/c8c58ad3-fd6e-4b2d-99fa-d8864355b638/the-concept-of-chilling-effect-20210322.pdf>> (last accessed 2 September 2021). According to Pech: “From a legal point of view, chilling effect may be defined as the negative effect any state action has on natural and/or legal persons, and which results in pre-emptively dissuading them from exercising their rights or fulfilling their professional obligations, for fear of being subject to formal state proceedings which could lead to sanctions or informal consequences such as threats, attacks or smear campaigns”. Furthermore, he said that “State action is understood in this context as any measure, practice or omission by public authorities which may deter natural and/or legal persons from exercising any of the rights provided to them under national, European and/or international law, or may discourage the potential fulfilment of one’s professional obligations (as in the case of judges, prosecutors and lawyers, for instance)”.

²⁴ Amnesty International, report 2020/21, “The state of the world’s human rights” <<https://www.amnesty.org/en/documents/pol10/3202/2021/en/>> (last accessed 13 June 2021).

²⁵ *ibid.*, 16.

²⁶ See Mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression and the Special Rapporteur for freedom of expression of the Inter-American Commission on Human Rights, OL BRA 6/2020, 3 July 2020 <<https://spcommreports.ohchr.org/TMResultsBase/DownloadPublicCommunicationFile?gld=25417>> (last accessed 16 June 2021).

²⁷ See UN Special Rapporteur, Report on disinformation, A/HRC/47/25, 13 April 2021, 12 <<https://undocs.org/A/HRC/47/25>> (last accessed 15 June 2021).

how such laws lead to the suppression of legitimate online expressions with limited or no due process or without prior court order and contrary to the requirements of Article 19(3) of the International Covenant on Civil and Political Rights (ICCPR).²⁸ In addition, a trend emerges that sees States delegating functions to online platform “speech police” that traditionally belong to the courts. The risk with such laws is that intermediaries are likely to err on the side of caution and “over-remove” content for fear of being sanctioned.²⁹

Concerning private companies, although they do not have the same human rights obligations as States, social media platforms are expected to respect human rights in their activities and operations according to the Guiding Principles on Business and Human Rights. In response to the challenges raised by fake news, the main social media platforms have adopted a range of policies and tools, generally banning what they consider to be false news and various deceptive practices that undermine authenticity and integrity on their platforms. Hence, social media giants like Twitter³⁰ and Facebook³¹ have adopted specific policies on COVID-19-related disinformation and have also established a third-party fact-checking programme,³² as well as a new community-based approach to fighting misinformation,³³ respectively.

Nonetheless, as we shall see, social media policies are still ineffective, counterproductive and in some cases detrimental to users’ rights.

With this in mind, this section of the article specifically focuses on States around the world that have taken particularly stringent emergency legislative and administrative measures to deal with fake news on social media during the COVID-19 pandemic: China, the USA, Russia, the UK, Australia, Canada, Burkina Faso, Singapore and India.³⁴ Furthermore, in grasping the extent of the legislative and administrative measures on the right to freedom of expression, the analysis will also cover the social media regulation in force.

1. China

In Asia, China has passed some of the strictest laws in the world when it comes to fake news.³⁵ In 2016, the Cybersecurity Law criminalised “any individual or organization . . . that fabricates or disseminates false information to disrupt the economic and social order” (Article 12).³⁶ In particular, China’s Cybersecurity Law requires social media platforms to

²⁸ *ibid*

²⁹ *ibid*.

³⁰ See Twitter, COVID-19 misleading information policy <<https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy>> (last accessed 16 June 2021).

³¹ See Facebook, Advertising policies related to coronavirus (COVID-19) <<https://www.facebook.com/business/help/1123969894625935>> (last accessed 16 June 2021).

³² See Facebook, Facebook’s Third-Party Fact-Checking Program <<https://www.facebook.com/journalismproject/programs/third-party-fact-checking>> (last accessed 16 June 2021).

³³ See Twitter, Introducing Birdwatch, a community-based approach to misinformation <https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation> (last accessed 16 June 2021).

³⁴ For a statistical analysis also extending to criminal measures to combat fake news during the COVID-19 pandemic, see International Press Institute, COVID-19: Number of Media Freedom Violations by Region <<https://ipi.media/covid19-media-freedom-monitoring/>> (last accessed 14 June 2021).

³⁵ See, eg, B Qin, D Strömberg and Y Wu, “Why does China allow freer social media? Protests versus surveillance and propaganda” (2017) 31(1) *Journal of Economic Perspectives* 117–40 <<https://doi.org/10.1257/jep.31.L117>>; L Guo, “China’s ‘fake news’ problem: exploring the spread of online rumors in the government-controlled news media” (2020) 8(8) *Digital Journalism* 992–1010 <<https://doi.org/10.1080/21670811.2020.1766986>>.

³⁶ See Standing Committee of the National People’s Congress, Cybersecurity Law of the People’s Republic of China, Order No. 53 of the President, 11 July 2016 <<http://www.lawinfochina.com/display.aspx?lib=law&id=22826#>> (last accessed 11 November 2020). See S Reeves, R Alcalá and E Gregory, “Fake news! China is a rule-of-law nation and respects international law” (2018) 39(4) *Harvard International Review* 42–46.

solely republish and link to news articles from registered news media. Furthermore, in 2018, the Chinese authorities started requiring microblogging sites to highlight and refute rumours on their platforms and launched a platform called Piyao that allows people to report potential fake news. The platform broadcasts real news sourced from state-owned media, party-controlled local newspapers and various government agencies. The app also uses artificial intelligence to automatically detect rumours of accounts on social media platforms like Weibo and WeChat, on which it broadcasts reports from state-owned media.

As reported in a reliable source, since the spread of the COVID-19 outbreak the Chinese government systematically started checking information on the disease available online and in the media.³⁷ Indeed, briefings and media reports describe how the government and public authorities delayed releasing information on the coronavirus outbreak to the public.³⁸ Thus, the Hubei authorities silenced and sanctioned a number of citizens for spreading rumours and disturbing the social order. Two of the people involved were medical experts who had warned the public of the spread of the coronavirus on social media.³⁹

The Cyberspace Administration of China (CAC), the government agency that controls Internet infrastructure and content within the Chinese borders, threatened websites, media platforms and accounts with sanctions for disseminating harmful or alarming news related to COVID-19. Additionally, in a CAC report, Sina Weibo, Tencent and ByteDance were flagged for an inspection of their platforms.⁴⁰

More generally, the authorities warned the public of the legal consequences of fake news. A Chinese police announcement report regarding sanctions for rumours demonstrates that many citizens received a reprimand, fines and administrative or criminal detention in January 2020.⁴¹

2. The USA

In the USA, Congress passed the Honest Ads Act in 2017, aiming to regulate online political advertising and to counter fake news. Specifically, the law requires social media platforms like Facebook and Google to keep copies of ads, make them public and keep tabs on who is paying for them and how much.⁴²

In addition, the Honest Ads Act compels companies to disclose details such as advertising spending, targeting strategies, buyers and funding. It also requires online political campaigns to adhere to stringent disclosure conditions for advertising on traditional media.⁴³

³⁷ See, eg, L Khalil “Digital authoritarianism, China and covid”, Lowy Institute <https://www.lowyinstitute.org/publications/digital-authoritarianism-china-and-covid#_edn42> (last accessed 11 November 2020).

³⁸ Forbes, “Report: China Delayed Releasing Vital Coronavirus Information, Despite Frustration from WHO” (2 June 2020) <<https://www.forbes.com/sites/isabeltogoh/2020/06/02/report-china-delayed-releasing-vital-coronavirus-information-despite-frustration-from-who/>> (last accessed 12 November 2020).

³⁹ <https://www.hubei.gov.cn/zxjy/rdhy/202001/t20200101_1862068.shtml> (last accessed 10 November 2020).

⁴⁰ <<https://mp.weixin.qq.com/s/YhjV75NwJZO4CypdepArQw>> (last accessed 10 November 2020).

⁴¹ Cf. China: Protect Human Rights While Combatting Coronavirus Outbreak <<https://mp.weixin.qq.com/s/3dYMFt1vXS-WuEJcZDytWw>> (last accessed 12 November 2020).

⁴² Cf. S.1989 – Honest Ads Act, 19 October 2017 <<https://www.congress.gov/bill/115th-congress/senate-bill/1989/text>> (last accessed 11 November 2020).

⁴³ R Kraski, “Combating fake news in social media: U.S. and German legal approaches” (2017) 91(4) *St. John’s Law Review* 923–55; see also Policy Report by G Hacıyakupoglu, J Yang Hui, VS Suguna, D Leong and M Rahman, “Countering fake news: a survey of recent global initiatives” (2018) *S Rajaratnam School of International Studies* 3–22, available at <https://www.rsis.edu.sg/wp-content/uploads/2018/03/PR180416_Counter-Fake-News.pdf>.

The US National Defense Authorization Act of 2017 approved the establishment of the Global Engagement Center⁴⁴ to lead, synchronise and coordinate the Federal Government's efforts to deal with foreign State and non-State propaganda and disinformation efforts aimed at jeopardising US national security interests.⁴⁵ By contrast, the Governor of the State of California vetoed a bill that was to create an advisory group meant to monitor the spread of misinformation on social media and find potential solutions.⁴⁶

We should also consider the Stigler Committee on Digital Platforms. This is an independent and non-partisan Committee of academics, policymakers and experts who have studied how social media platforms like Google, Twitter and Facebook impact economics and antitrust laws, data protection, the political system and the new media industry. In this regard, the Stigler Committee Report addresses the impact of social media platforms on various aspects of society and proposes policy solutions for lawmakers and regulators to consider when dealing with the power held by these companies.⁴⁷

It should be noted that on 20 February 2020, the Trump administration in its annual economic report criticised – especially in its section on antitrust enforcement for the digital economy – the Stigler Report's proposal for creating a Digital Agency, saying that it “raises a host of issues” and “that the downsides of new, far-reaching regulation need to be taken seriously”.⁴⁸ However, these criticisms have been refuted,⁴⁹ as these were based on a claim already advanced by Charles Koch Institute's Neil Chilson in the *Washington Post*,⁵⁰ an argument already refuted in an earlier *ProMarket* post.⁵¹

In 2020, during the COVID-19 pandemic, the USA tightened control over coronavirus messaging by government health officials and scientists, directing them to coordinate all statements and public appearances.⁵²

3. Russia

On 31 March 2020, the Russian authorities approved amendments to the Criminal Code and to the Code of Administrative Offences that introduced criminal penalties for the public

⁴⁴ “Global Engagement Center”, US Department of State: Diplomacy in Action <<https://www.state.gov/r/gec/>> (last accessed 11 November 2020).

⁴⁵ See S.2943 – National Defense Authorization Act for Fiscal Year 2017, Congress.Gov <<https://www.congress.gov/bill/114th-congress/senatebill/2943/text>> (last accessed 11 November 2020).

⁴⁶ See Senate of the State of California, SB-1424 Internet: social media: advisory group <https://leginfo.ca.gov/faces/billStatusClient.xhtml?bill_id=201720180SB1424> (last accessed 11 November 2020).

⁴⁷ Cf. Stigler Committee on Digital Platforms Final Report, 2019 <<https://www.publicknowledge.org/wp-content/uploads/2019/09/Stigler-Committee-on-Digital-Platforms-Final-Report.pdf>> (last accessed 26 August 2021).

⁴⁸ Cf. 2020 Economic Report of the President, 20 February 2020 <<https://trumpwhitehouse.archives.gov/wp-content/uploads/2020/02/2020-Economic-Report-of-the-President-WHCEA.pdf>> (last accessed 26 August 2021).

⁴⁹ Cf. L Zingales and FM Lancieri, “The Trump Administration attacks the Stigler Report on Digital Platforms” <<https://promarket.org/2020/02/21/the-trump-administration-attacks-the-stigler-report-on-digital-platforms/>> (last accessed 26 August 2021).

⁵⁰ Cf. N Chilson, “Creating a new federal agency to regulate Big Tech would be a disaster” <<https://www.washingtonpost.com/outlook/2019/10/30/creating-new-federal-agency-regulate-big-tech-would-be-disaster/>> (last accessed 26 August 2021).

⁵¹ Cf. L Zingales and F Scott Morton, “Why a new digital authority is necessary” <<https://promarket.org/2019/11/08/why-a-new-digital-authority-is-necessary/>> (last accessed 26 August 2021).

⁵² *New York Times*, “Pence will control all coronavirus messaging from health officials” (27 February 2020) <<https://www.nytimes.com/2020/02/27/us/politics/us-coronavirus-pence.html>> (last accessed 10 November 2020).

dissemination of knowingly fake news in the context of the COVID-19 emergency, including the social media environment.⁵³

In particular, the Government established criminal liability for the public dissemination of knowingly fake news regarding circumstances posing a threat to the life and safety of citizens under Article 207.1 of the Criminal Code of the Russian Federation (CCRF),⁵⁴ as well as for the public dissemination of knowingly false socially significant information, which entailed grave consequences under Article 207.2 of the CCRF.⁵⁵

Furthermore, the legislative changes extended criminal sanctions for violating sanitary and epidemiological regulations. Amendments to the Russian Code of Administrative Offences introduced hefty fines of up to five million roubles for journalists spreading fake news. In fact, pursuant to Article 13.15 of the Code of Administrative Offences of the Russian Federation, committing the same offence twice could lead to a fine of up to ten million roubles.

4. The UK

In the UK, the House of Commons DCMS examined the issue of disinformation and fake news from January 2017, focusing on issues such as the definition, role and legal liabilities of social media platforms.⁵⁶

The DCMS Special Committee drew up an interim report on 29 July 2018 and a final one in February 2019. A set of recommendations produced by the committee include: (1) a compulsory code of ethics for technology companies overseen by an independent regulator with powers to launch legal action; (2) changes to electoral communications laws to ensure the transparency of political communications online; and (3) the obligation for social media companies to take down known sources of harmful content, including proven sources of disinformation.⁵⁷

Moreover, the DCMS Special Committee invited elected representatives from Argentina, Belgium, Brazil, Canada, France, Ireland, Latvia and Singapore to establish committee on disinformation and fake news (the so-called International Grand Committee), which held its inaugural session in November 2018.⁵⁸

Following the session, members of the International Grand Committee signed a declaration on the “Principles of the law governing the Internet”, which affirms the

⁵³ On Russian fake news legislation, see, eg, O Pollicino, “Fundamental rights as bycatch – Russia’s anti-fake news legislation” (*VerfBlog*, 28 March 2019) <<https://verfassungsblog.de/fundamental-rights-as-bycatch-russias-anti-fake-news-legislation/>, doi:10.17176/20190517-144352-0> (last accessed 12 November 2020).

⁵⁴ See The Criminal Code of the Russian Federation, Art 207.1 “Public dissemination of knowingly false information about circumstances that pose a threat to the life and safety of citizens” <<https://rulaws.ru/uk/Razdel-IX/Glava-24/Statya-207.1/>>.

⁵⁵ *ibid*, Art 207.2 “Public dissemination of knowingly false socially significant information, which entailed grave consequences”.

⁵⁶ See the UK House of Commons Digital, Culture, Media and Sport Committee, “Fake news” inquiry, 30 January 2017 <<https://old.parliament.uk/business/committees/committees-a-z/commons-select/culture-media-and-sport-committee/news-parliament-2015/fake-news-launch-16-17/>> (last accessed 12 November 2020).

⁵⁷ See the House of Commons, Digital, Culture, Media and Sport Committee, Disinformation and “fake news”: Interim Report Fifth Report of Session 2017-19, HC 363, 29 July 2018 <<https://publications.parliament.uk/pa/cm201719/cmselect/cmcomeds/363/363.pdf>> (last accessed 13 November 2020).

⁵⁸ See the House of Commons, Digital, Culture, Media and Sport International Grand Committee Oral evidence: Disinformation and “fake news”, HC 363, 27 November 2018 <<http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/disinformation-and-fake-news/oral/92923.pdf>> (last accessed 13 November 2020).

parliamentarians' commitment to the principles of transparency, accountability and the protection of representative democracy in regard to the Internet.⁵⁹

5. Australia

In Australia, the government appointed a taskforce to address fake news threats to electoral integrity through its foreign interference laws, which passed through parliament in June 2018 and have also had some bearing on the question of disinformation. Later, the Australian Electoral Commission (AEC) commenced a social media literacy campaign and other activities to coincide with the 2019 federal election. In addition, there have also been several recent parliamentary inquiries and an inquiry by the Australian Competition and Consumer Commission (ACCC) examining issues related to fake news.

It should be noted that prior to the July 2018 federal by-elections⁶⁰ held across four states, the Turnbull government established a multi-agency body, the Electoral Integrity Assurance Taskforce,⁶¹ to address risks to the integrity of the electoral system, particularly in relation to cyber interference and online disinformation. Agencies involved include the AEC, the Department of Finance, the Department of Home Affairs and the Australian Cyber Security Centre. According to the Department of Home Affairs, the taskforce's role is to provide the AEC with technical advice and expertise in relation to cyber interference and online disinformation with regard to electoral processes.

A media report on the establishment of the taskforce suggested that its central concern is cybersecurity and disinformation, including fake news and interference with the electoral roll or AEC systems. The media report added that the potential use of disinformation and messaging and any covert operations designed to disrupt the by-elections would also be closely monitored, as the foreign interference "threat environment" had escalated even within the two years since the last federal election in 2016.⁶²

Another critical point that should be mentioned is that the National Security Legislation Amendment Act 2018 added new foreign interference offences to the Australian Commonwealth Criminal Code.⁶³ The elements of these foreign interference offences could arguably be applied to persons who weaponise fake news in certain circumstances. In particular, the offences extend to persons who, on behalf of a foreign government, engage in deceptive or covert conduct intended to influence a political or governmental process of the Commonwealth or a State or Territory, or to influence the exercise of an Australian democratic or political right or duty, whether in Australia or not.

Some positive signs are emerging from the AEC advertising campaign called "Stop and Consider". The AEC started this campaign on 15 April 2019 on social media platforms such as Facebook, Twitter and Instagram to encourage voters to carefully check the sources of

⁵⁹ See Members of the "International Grand Committee" on Disinformation and "Fake News", the Declaration on the "Principles of the law governing the Internet", 27 November 2018 <<https://old.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/news/declaration-internet-17-19/>> (last accessed 13 November 2020).

⁶⁰ See the Parliament of Australia, Australian Electoral Commission, Striving to safeguard election, 2 March 2019 <<https://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;query=Id%3A%22media%2Fpressclp%2F6529492%22>> (last accessed 18 November 2020).

⁶¹ See Australian Electoral Commission (AEC), Electoral Integrity Assurance Taskforce, (last update) 10 July 2020 <<https://www.aec.gov.au/elections/electoral-advertising/electoral-integrity.htm>> (last accessed 18 November 2020).

⁶² See the Parliament of Australia, Department of Home Affairs, ASIO, AFP call-up for by-elections, 9 June 2018 <<https://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;query=Id%3A%22media%2Fpressclp%2F6016585%22>> (last accessed 18 November 2020).

⁶³ See the Australian Government, National Security Legislation Amendment (Espionage and Foreign Interference) Act 2018 No. 67, 2018, 10 December 2018 <<https://www.legislation.gov.au/Details/C2018C00506>> (last accessed 18 November 2020).

electoral communication they would see or hear during the 2019 federal election campaign.⁶⁴ In effect, Stop and Consider was a media literacy campaign based on empowering users, positively alerting voters to the possibility of disinformation or false information intended to influence their vote and helping them to check sources of information and thus cast an informed vote.

Since December 2017, the ACCC has been conducting an inquiry into the impact of online platforms on media competition in Australia, including the implications of this impact for quality news and journalism.⁶⁵ The ACCC's preliminary report discusses a range of intersecting factors that pose a risk of increasing audience exposure to fake news.⁶⁶ These factors include commercial incentives for media companies to produce sensational "clickbait stories" optimised for search engines and news aggregators and designed to go viral. Other potential problems include those associated with news feeds on social media platforms. Such feeds show users a mix of individual news stories with no context regarding source credibility, which makes it difficult for users to discern the quality of the information. In addition, platforms select and prioritise news on the basis of users' past behaviours and preferences, so news stories that share the same perspectives may be repeatedly made available to consumers.

Recently, the Australian government has enacted the News Media and Digital Platforms Mandatory Bargaining Code, which, among other things, also promotes self-regulation by social media companies and diminishes government intervention.⁶⁷

6. Canada

In March 2018, the Canadian House of Commons Standing Committee on Access to Information, Privacy and Ethics (AIPE) began an inquiry into a breach of personal information involving Cambridge Analytica and Facebook.⁶⁸

The AIPE Committee's preliminary report contained a number of recommendations, mostly amendments to the Personal Information Protection and Electronic Documents Act.⁶⁹ The AIPE Committee's final report of December 2018 included a number of potential regulatory responses to the problem of misinformation and disinformation on social media.⁷⁰

⁶⁴ See Australian Electoral Commission (AEC), Encouraging voters to "stop and consider" this federal election, (last update) 15 April 2019 <<https://www.aec.gov.au/media/media-releases/2019/04-15.htm>> (last accessed 18 November 2020).

⁶⁵ See the Australian Competition & Consumer Commission (ACCC), Digital platforms inquiry, 4 December 2017 <<https://www.accc.gov.au/focus-areas/inquiries-ongoing/digital-platforms-inquiry>> (last accessed 18 November 2020).

⁶⁶ See the Australian Competition & Consumer Commission (ACCC), Digital platforms inquiry, Preliminary Report, 10 December 2018 <<https://www.accc.gov.au/focus-areas/inquiries-ongoing/digital-platforms-inquiry/preliminary-report>> (last accessed 18 November 2020).

⁶⁷ Cf. Treasury Laws Amendment (News Media and Digital Platforms Mandatory Bargaining Code) Act 2021, No 21/2021 <<https://www.legislation.gov.au/Details/C2021A00021>> (last accessed 27 August 2021).

⁶⁸ See Standing Committee on Access to Information, Privacy and Ethics, Breach of personal information involving Cambridge Analytica and Facebook, Reports and Government Responses. On this point, two reports can be consulted: (1) Report 16: Addressing Digital Privacy Vulnerabilities and Potential Threats to Canada's Democratic Electoral Process, 14 June 2018; and (2) Report 17, Democracy under Threat: Risks and Solutions in the Era of Disinformation and Data Monopoly, 6 December 2018 <<https://www.ourcommons.ca/Committees/en/ETHI/StudyActivity?studyActivityId=10044891>> (last accessed 23 November 2020).

⁶⁹ See Committee on Access to Information, Privacy and Ethics, Report 16: Addressing digital privacy vulnerabilities and potential threats to Canada's democratic electoral process, 14 June 2018 <<https://www.ourcommons.ca/DocumentViewer/en/42-1/ETHI/report-16/>> (last accessed 24 November 2020).

⁷⁰ See Committee on Access to Information, Privacy and Ethics, Report 17, Democracy under Threat: Risks and Solutions in the Era of Disinformation and Data Monopoly, 6 December 2018 <<https://www.ourcommons.ca/DocumentViewer/en/42-1/ETHI/report-17>> (last accessed 25 November 2020).

One recommendation was that social media platforms would be required to be more transparent with regard to the processes behind the dissemination of material online, including clear labelling of automated or algorithmically produced content.

Other suggestions were that there should be obligations on platforms to take down illegal content, including disinformation and above all fake news, with more investment by platforms and governments in digital literacy programmes and public awareness campaigns.

7. Burkina Faso

In June 2019, Burkina Faso's parliament adopted the Protection from Online Falsehoods and Manipulation Act, which seeks to punish the publication of fake news compromising security operations, false information about rights abuses or destruction of property or images and audio from a terrorist attack.

Parliament specifically amended the country's Penal Code to introduce a series of new offences that aim to fight terrorism and organised crime, fight the spread of fake news and suppress efforts to demoralise the Burkinabe armed forces.⁷¹ Offenders could face fines of up to £7000 or a maximum ten years in jail.

8. Singapore

In May 2019, the Singapore Parliament approved a law criminalising the dissemination of fake news online.⁷² The law makes it illegal to spread false statements of facts that compromise security, public peace and safety and the country's relations with other nations.

In particular, the law punishes those who post fake news with heavy fines and even jail time. In this regard, if a user shares false information, the penalty is a fine of up to \$37,000 or five years in prison. What is more, the punishment jumps to \$74,000 and a potential ten-year jail term if the falsehood was shared using an inauthentic online account or a bot.

A further important aspect is that social media platforms like Facebook and Twitter face fines of up to \$740,000 for their roles in spreading misinformation.

9. India

In India, during the nationwide lockdown imposed after the pandemic, more than fifty journalists were arrested under emergency laws for spreading fake news.

On 7 April 2020, Uttar Pradesh police lodged a First Information Report (FIR) against journalist Prashant Kanojia for allegedly making "objectionable remarks" about Prime Minister Modi and Chief Minister Yogi Adityanath on social media.

Shortly afterwards, the Uttar Pradesh police registered another FIR against The Wire, a daily news website, and its editor Siddharth Varadarajan for reporting that Yogi Adityanath had attended a public religious event after the nationwide lockdown was announced.⁷³

⁷¹ See Assemblée Nationale, Modification de la loi 31 Mat 2018 No. 25 (Code Penal) <<https://perma.cc/LQH4-72W4>> (last accessed 28 November 2020).

⁷² See Republic of Singapore, Protection from Online Falsehoods and Manipulation Act 2019, 25 June 2019 <<https://sso.agc.gov.sg/Acts-Supp/18-2019/Published/20190625?DocDate=20190625>> (last accessed 28 November 2020). See M Zastrow, "Singapore passes 'fake news' law following researcher outcry" (*Nature*, 15 May 2019) <<https://doi.org/10.1038/d41586-019-01542-7>>. See also, for a comparative law perspective, K Han, "Big Brother's regional ripple effect: Singapore's recent 'fake news' law which gives ministers the right to ban content they do not like, may encourage other regimes in south-east Asia to follow suit" (2019) 48(2) *Index on Censorship* 67–69.

⁷³ Amnesty International, *supra*, note 24, 184.

IV. Legislative and administrative measures regulating fake news in the European Union

Member States of the EU have started to regulate fake news by law, and some EU governments have issued administrative measures accordingly. However, the COVID-19 pandemic has considerably increased legislative initiatives and administrative measures imposed by EU governments. Therefore, this section of the article specifically looks at the Member States that have taken particularly stringent emergency legislative and administrative measures to manage fake news during the COVID-19 pandemic: Germany, France, Italy and Spain. In addition, the social media regulation in force is analysed in order to understand the impact of emergency measures on the right to freedom of expression.

I. Germany

On 28 October 2020, the Federal States in Germany passed the Interstate Media Treaty (*Medienstaatsvertrag* – “MStV”).⁷⁴ It addresses online disinformation – and thus also fake news – by regulating transparent algorithms, the labelling of social bots, the findability of public service content and journalistic due diligence for social media.⁷⁵

The MStV is the German implementation of the EU Audiovisual Media Services Directive 2010/13/EU,⁷⁶ as amended by Directive 2018/1808/EU.⁷⁷ It replaces the Interstate Broadcasting Treaty (*Rundfunkstaatsvertrag* – RStV⁷⁸) and is considered an important cornerstone in media policy.

The MStV plays a pivotal role in national efforts to modernise the media landscape and to make the German legislative framework fit the European social media legal environment.⁷⁹ As a result, the MStV focuses on telematics services in addition to

⁷⁴ See J Kalbhenn and M Hemmert-Halswick, “EU-weite Vorgaben zur Content-Moderation auf sozialen Netzwerken” [EU-wide guidelines on content moderation on social networks] (2021) 3 ZUM – Zeitschrift für Urheber- und Medienrecht 184–94. See also N Gielen and S Uphues, “Regulierung von Markt- und Meinungsmacht durch die Europäische Union” [Regulation of market and opinion power by the European Union] (2021) 14 Europäische Zeitschrift für Wirtschaftsrecht 627–37; J Kühling “Fake news and hate speech – Die Verantwortung von Medienintermediären zwischen neuen NetzDG, MStV und DSA” [Fake news and hate speech – the responsibility of media intermediaries between new NetzDG, MStV and DSA] (2021) 6 ZUM – Zeitschrift für Urheber- und Medienrecht 461–72.

⁷⁵ Cf. *Medienstaatsvertrag* (MStV), official German version, especially § 18 Abs. 3; § 19; § 84; § 93/94 <<https://www.rlp.de/fileadmin/rlp-stk/pdf-Dateien/Medienpolitik/Medienstaatsvertrag.pdf>> (last accessed 28 August 2021).

⁷⁶ Cf. Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) (last accessed 28 August 2021).

⁷⁷ Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018, amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities <<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018L1808&from=EN>> (last accessed 28 August 2021).

⁷⁸ Cf. Interstate Treaty on Broadcasting and Telemedia (Interstate Broadcasting Treaty) <https://www.die-medienanstalten.de/fileadmin/user_upload/Rechtsgrundlagen/Gesetze_Staatsvertraege/RStV_22_english_version_clean.pdf> (last accessed 28 August 2021).

⁷⁹ In German media law scholarship, see the recent report by B Holznapel and JC Kalbhenn, “Monitoring media pluralism in the digital era application of the media pluralism monitor in the European Union, Albania, Montenegro, the Republic of North Macedonia, Serbia & Turkey in the year 2020 – country report: Germany, Robert Schuman Centre, Issue 2021.2823, July 2021, available online at <https://cadmus.eui.eu/bitstream/handle/1814/71947/germany_results_mpm_2021_cmpf.pdf?sequence=1> (last accessed 28 August 2021).

broadcasting. Through media intermediaries, social media platforms, user interfaces and video sharing services, the MStV applies to many players in the social media market.⁸⁰

Previously, the Network Enforcement Act (*Netzwerkdurchsetzungsgesetz – NetzDG*) had been approved on 30 June 2017⁸¹ and recently amended on 28 June 2021⁸² with the specific aim of fighting hate speech and fake news on social networks by improving the enforcement of the current laws.⁸³ Basically, with the NetzDG, Germany made fighting hate speech on social networks a priority. In this legal context, the NetzDG also regulates disinformation – and thus specifically fake news – by implementing previous legislation on disinformation.⁸⁴ To achieve this goal, the German government implemented law enforcement on social networks in order to promptly remove objectively criminal content, namely incitement to hatred, abuse, defamation or content that might lead to a breach of the peace by misleading authorities into thinking a crime has been committed.

However, the obligations placed upon private companies to regulate and take down content raises concern with respect to freedom of expression. A prohibition on the dissemination of information based on vague and ambiguous criteria, such as “insult” or “defamation”, could be incompatible with Article 19 of the ICCPR.⁸⁵

The NetzDG works by requiring social media platforms to provide a mechanism for users to submit complaints about illegal content. Once they receive a complaint, platforms

⁸⁰ For Germany and disinformation, cf. The Federal Constitutional Court (Bundesverfassungsgericht) of 18 July 2018 strengthening online activities of public service media, 79–81 <https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2018/07/rs20180718_1bvr167516en.html> (last accessed 28 August 2021).

⁸¹ Act to Improve the Enforcement of Rights on Social Networks, 1 September 2017 <<http://www.gesetze-im-internet.de/netzdg/NetzDG.pdf>> English version archived at <<https://perma.cc/BAE2-KAJX>> (last accessed 29 November 2020). On the NetzDG, see M Eifert, “Evaluation des NetzDG Im Auftrag des BMJV” <https://www.bmjv.de/SharedDocs/Downloads/DE/News/PM/090920_Juristisches_Gutachten_Netz.pdf?__blob=publicationFile&v=3> (last accessed 28 August 2021). See also the report by H Tworek and P Leerssen, “An analysis of Germany’s NetzDG law” (2019) Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, available at <<https://dare.uva.nl/search?identifier=3dc07e3e-a988-4f61-bb8c-388d903504a7>>. See also the legal report by J Gesley, “Initiatives to counter fake news: Germany”, available at <https://www.loc.gov/law/help/fake-news/germany.php#_ftnref42> (last accessed 2 December 2020). In academia, see B Holznagel, “Das Compliance-System des Entwurfs des Netzwerk-durchsetzungsgesetzes – Eine kritische Bestandsaufnahme aus internationaler Sicht” [The Compliance System of the Draft Network Enforcement Act – A Critical Review from an International Perspective] (2017) 8/9 *Zeitschrift für Urheber- und Medienrecht* 615–24.

⁸² The amendment intends to enhance the information content and comparability of social media transparency reports and to increase the user-friendliness of reporting channels for complaints about illegal content. In addition, the amendment introduces an appeal procedure for measures taken by the social media company. The powers of the Federal Office of Justice are extended to encompass supervisory powers. Lastly, due to the new requirements of the EU Audiovisual Media Services Directive, the services of video-sharing platforms are included in the scope of the Network Enforcement Act. See Gesetz zur Änderung des Netzwerkdurchsetzungsgesetzes <<https://perma.cc/9W8E-GSWM>> (last accessed 11 September 2021).

⁸³ See JC Kalbhenn and M Hemmert-Halswick, “Der Referentenentwurf zum NetzDG – Vom Compliance Ansatz zu Designvorgaben” [The draft bill on the NetzDG – from compliance approach to design requirements] (2020) 8 *MMR – Zeitschrift für IT-Recht und Digitalisierung* 518–22. See also S Niggemann, “Die NetzDG-Novelle – Eine Kritik mit Blick auf die Rechte der Nutzerinnen und Nutzer” [The NetzDG amendment – a critique with a view to users’ rights] (2021) 5 *Computer und Recht* 326–31; M Cornils, “Präzisierung, Vervollständigung und Erweiterung: Die Änderungen des Netzwerkdurchsetzungsgesetzes” [Completion and expansion: the amendments to the Network Enforcement Act] (2021) 34 *Neue Juristische Wochenschrift* 2465–71.

⁸⁴ See, eg, V Claussen, “Fighting hate speech and fake news. The Network Enforcement Act (NetzDG) in Germany in the context of European legislation” (2018) 3 *Media Laws* 113–15, available online at <<https://www.medialaws.eu/wp-content/uploads/2019/05/6-Claussen.pdf>> (last accessed 28 August 2021).

⁸⁵ See Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 1 June 2017, OL DEU 1/2017, 4 <<https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf>> (last accessed 16 June 2021).

must investigate whether the content is illegal and must remove it within seven days only if the content is “manifestly unlawful”.⁸⁶ Failing that, public authorities may impose high fines for failing to comply with existing legal obligations. In particular, platforms that do not remove clearly illegal content may be fined up to €50 million.⁸⁷

Nevertheless, the provisions imposing high fines for non-compliance with the obligations set out in the bill raise concerns, as the obligations as mentioned above may represent undue interference with the right to freedom of expression and privacy. The heavy fines raise proportionality concerns, and they may prompt social networks to remove content that may actually be lawful.⁸⁸

Moreover, the NetzDG has been cited by other countries seeking to introduce unduly restrictive intermediary laws or social media regulations that would enable the removal of fake news without a judicial or even a quasi-judicial order. Hence, several criticisms have rightly been levelled against the NetzDG, and some political parties have unsuccessfully presented proposals to modify the law since it is considered unconstitutional, particularly regarding freedom of expression.⁸⁹

2. France

In France, a law against the manipulation of information was approved by the National Assembly on 22 December 2018 with the aim of providing further protecting democratic principles from the spread of fake news.⁹⁰ In particular, this law targeted the widespread and rapid dissemination of fake news by means of digital tools, especially through the channels of dissemination offered by social media platforms influenced by foreign States.⁹¹

Thus, the new legislation concentrates especially on election campaigns, both just before and during elections, in order to focus available regulation tools on the real risk,

⁸⁶ It should be noted that the NetzDG did not introduce new obligations but tightened the time frame within which criminal content must be deleted.

⁸⁷ Regarding other States, on 30 March 2020, the Hungarian government issued a law making the spread of false or distorted information punishable with up to five years imprisonment and, on 4 May 2020, it issued a decree introducing derogations from the principle of the right to information; this is the Decree of 4 May 2020 <<https://magyarkozlony.hu/dokumentumok/008772a9660e8ff51e7dd1f3d39ec056853ab26c/megtekintes>>. The decree limits the application of data subjects' rights safeguarded under Arts 15–22 of the GDPR in relation to the processing of data conducted by both public and private entities for the purpose fighting the COVID-19 crisis. On this point, see A Gergely and V Gulyas, “Orban uses crisis powers for detentions under fake news law” (Bloomberg.com, May 2020) <<http://search.ebscohost.com/login.aspx?direct=true&db=bsu&AN=143195291&site=ehost-live>> (last accessed 14 December 2020); see also *The Guardian*, “Hungary passes law that will let Orbán rule by decree” (30 March 2020) <<https://www.theguardian.com/world/2020/mar/30/hungary-jail-for-coronavirus-misinformation-viktor-orban>> (last accessed 14 December 2020). Moreover, Bulgaria's government used its state of emergency decree to try to amend the criminal code and introduce prison sentences for spreading what it deems fake news about the outbreak with up to three years in prison or a fine of up to €5,000. Furthermore, another bill submitted to Parliament by a ruling coalition party on 19 April 2020 would, if approved, give the Bulgarian authorities more powers to suspend websites for spreading Internet misinformation beyond the immediate health emergency.

⁸⁸ *ibid.*, 4.

⁸⁹ See, eg, G Nolte, “Hate-Speech, Fake-News, das ‘Netzwerkdurchsetzungsgesetz und Vielfaltsicherung durch Suchmaschinen’ [Hate speech, fake news, the “Network Enforcement Act” and assuring diversity through search engines] (2017) 61 *Zeitschrift für Urheber- und Medienrecht ZUM* 552–54. For the proposals of the parties, see, eg, the draft act submitted by the Green Party, BT-Drs. 19/5950 <<http://dipbt.bundestag.de/dip21/btd/19/059/1905950.pdf>>, archived at <<http://perma.cc/FBW8-FJDP>>.

⁹⁰ See “Law against the Manipulation of Information”, which also introduced three new articles (L. 112, L. 163-1 and L. 163-2) to the French Electoral Code <<https://www.legifrance.gouv.fr/loda/id/JORFTEXT000037847559/2020-12-15/>> (last accessed 3 December 2020).

⁹¹ R Craufurd Smith, “Fake news, French law and democratic legitimacy: lessons for the United Kingdom?” (2019) 11(1) *Journal of Media Law* 52–81.

namely attempts to influence election results, as occurred during the US presidential elections and the Brexit campaign.

Legally speaking, the law provides the administrative measures requiring a transparency obligation for platforms, which must report any sponsored content by publishing the name of the author and the amount paid. Additionally, platforms exceeding a certain number of hits per day must have a legal representative in France and publish their algorithms. There is also a legal injunction that allows the circulation of fake news to be quickly halted.

Last but not least, during the three months preceding an election, an interim judge will qualify fake news as defined in the 1881 law on the freedom of the press according to three criteria: (1) the news must be manifestly fake; (2) it must have been disseminated deliberately on a massive scale; and (3) it must be liable to cause a breach of the peace or compromise the outcome of an election.

The French law against the manipulation of information establishes a duty of cooperation on the part of social media platforms in order to encourage them to introduce measures to prevent fake news and make these measures public. In this regard, the French Broadcasting Authority, the Superior Audiovisual Council, is assigned the role of preventing, suspending and stopping the broadcast of television services controlled by, or influenced by, other States and that are harmful to the fundamental interests of the country.

In the midst of the health emergency, in order to fight fake news, on 30 April 2020, the French government created a specific section on COVID-19 on its website.⁹² However, this initiative raised issues among journalists who consider that the government should not judge information.⁹³ Consequently, the French government took down the fake news COVID-19 page after accusations that it had overstepped its constitutional role and infringed upon press freedoms.⁹⁴

In greater detail, the national assembly approved a controversial law referring to COVID-19 information that gives platforms a one-hour deadline to remove related content after being instructed to do so by the authorities.⁹⁵ Critics reasonably claimed that the law limits freedom of expression and is difficult to apply.⁹⁶ Indeed, on 18 June 2020, the Constitutional Council judged that social media obligations to remove illegal content within twenty-four hours were not compatible with freedom of expression.⁹⁷

3. Italy

In Italy, following the constitutional referendum campaign in December 2016, a huge debate arose, and political actors began calling for new regulations to address the proliferation of fake news online. Thus, a number of proposals were made to prevent it; these could lead to the imposition of strict liability on social media platforms.

On 7 February 2017, a Member of Parliament, Senator Adele Gambaro, introduced a particularly controversial bill (the so-called *DDL Gambaro*) proposing fines and criminal

⁹² French government “Information Coronavirus” <<https://www.gouvernement.fr/info-coronavirus>> (last accessed 4 December 2020).

⁹³ *Libération*, “Désinfox coronavirus: l’Etat n’est pas l’arbitre de l’information” (3 May 2020) <https://www.liberation.fr/debats/2020/05/03/desinfox-coronavirus-l-etat-n-est-pas-l-arbitre-de-l-information_1787221> (last accessed 4 December 2020).

⁹⁴ On the academic debate see, more generally, Craufurd Smith, *supra*, note 91.

⁹⁵ The National Assembly, Bill to fight against hateful content on the internet, 13 May 2020 <<https://perma.cc/C7FD-J62S>> (last accessed 5 December 2020).

⁹⁶ FC Bremner, “French fake news law ‘will censor free speech’” (2018) *The Times* (London, England) 27.

⁹⁷ Decision n. 2020-801 DC of June 18, 2020, Law to combat hate content on the internet <<https://perma.cc/72VE-SMDJ>> (last accessed 5 December 2020).

penalties for anyone who publishes or spreads “false, exaggerated, or biased” news reports online.⁹⁸ However, after severe criticism from public opinion, the bill did not move forward in Parliament and was not adopted.

The Italian Minister of the Interior aimed to combat fake news by promoting a system of reporting manifestly unfounded and biased news or openly defamatory content. In this regard, on 18 January 2018, the Minister of the Interior introduced the “Red Button Operational Protocol” to fight the dissemination of fake news online at the time of the political elections in 2018.

The Ministry of the Interior introduced this specific online procedure in order to limit the actions of those who, with the sole intent of conditioning public opinion and tendentially orientating their thoughts and choices, design and spread news relating to topics or subjects of public interest with no foundation. Specifically, the protocol provided a “red button” signalling service where users can report a possible network of contents linked to the phenomenon of fake news. The unit of the Italian state police that investigates cyber-crime was tasked with reviewing reports and act accordingly.⁹⁹

The Constitutional Affairs Committee of the Senate of the Italian Republic is currently discussing bill No. 1900 aimed at establishing a parliamentary committee of inquiry in order to examine the problem of disinformation and, more precisely, the dissemination of fake news on a massive scale.¹⁰⁰ The bill under discussion does not establish any binding measures to deal with the spread of fake news. On the contrary, its purpose is to empower a committee with a variety of tasks:

- (1) Investigating the massive dissemination of illegal, false, non-verified or intentionally misleading information and content via traditional and online media.
- (2) Ascertaining whether such activities are backed by subjects, groups or organisations that receive financial support, including from foreign sources, with the specific aim of manipulating information and influencing public opinion, also in the context of electoral or referendum campaigns.
- (3) Assessing the impact of disinformation on health and in the context of the COVID-19 pandemic.
- (4) Evaluating whether disinformation activities pursue the goal of inciting hatred, discrimination and violence.
- (5) Exploring whether any connection exists between disinformation and commercial activities, most notably pursued by websites and digital platforms.
- (6) Verifying the status quo from a legal standpoint, as well as the existence and adequacy of procedures implemented by media platforms and social media service providers for the removal of false pieces of information and illegal content.
- (7) Assessing the existence of social, educational and literacy measures and best practices or of initiatives aimed at raising the awareness of individuals about the importance of fact checking and reliable sources of information.
- (8) Determining whether legal or administrative measures aimed at countering and prevent disinformation, as well as crimes committed via the media, are necessary,

⁹⁸ See Senate of the Republic, Bill No. 2688 of 7 February 2017 <<http://www.senato.it/service/PDF/PDFServer/BGT/01006504.pdf>> (last accessed 6 December 2020).

⁹⁹ See D Kaye, “Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression” OL ITA 1/2018, 20 March 2018 <<https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-ITA-1-2018.pdf>> (last accessed 7 December 2020).

¹⁰⁰ See Senate Act No. 1900 – Establishment of a parliamentary commission of inquiry into the massive dissemination of false information <<http://www.senato.it/leg/18/BGT/Schede/Ddliter/53197.htm>> (last accessed 7 December 2020).

also with regard to the negative consequences of disinformation on the development of minors and their learning abilities.¹⁰¹

The Italian government further engaged in transparency and debunking actions. On 4 April 2020, it set up a monitoring taskforce to combat the spread of fake news related to COVID-19 on the web and on social networks. The main goals of the government monitoring taskforce are to contain the risk that the spread of online disinformation could weaken pandemic containment measures and to promote initiatives in order to increase citizens' control over the reliability of social network information.

The Italian Ministry of Health also took action to tackle fake news on COVID-19 on the institutional homepage, where it promoted the information campaign "Beware of Hoaxes" with a collection of the most recurrent fake news on social media.¹⁰² In addition, the Communication Authority (AGCOM) launched an observatory on disinformation online by publishing reports, sending information to ministries and involving various stakeholders in the design of policies against disinformation.¹⁰³

The actions taken by the Italian government have been rightly criticised, as these pose many problems concerning the right to freedom of expression, media freedom and media pluralism,¹⁰⁴ as well as for having adopted a notion of fake news that is in disagreement with that of the European Commission, which deemed it inadequate and misleading.¹⁰⁵

4. Spain

In Spain, too, the government started monitoring online information in order to contain the spread of fake news regarding COVID-19 by establishing a specific protocol against disinformation campaigns.¹⁰⁶

On 30 October 2020, the Spanish government issued a ministerial order approving the protocol called "Procedure for Intervention against Disinformation" by Spain's National Security Council in order to prevent, detect and respond to disinformation campaigns in addition to establishing coordination mechanisms.¹⁰⁷

The document makes provision for the possibility of carrying out communication campaigns to counter fake news stories without censoring them. It is up to the government

¹⁰¹ See E Apa and M Bianchini, "Parliament considers establishing an ad-hoc parliamentary committee of inquiry on the massive dissemination of fake news" (2020) 10 *Iris* 1-3 <<https://merlin.obs.coe.int/article/9004>>.

¹⁰² The Ministry of Health "Beware of Hoaxes" <<http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioContenutiNuovoCoronavirus.jsp?lingua=italiano&id=5387&area=nuovoCoronavirus&menu=vuoto>> (last accessed 7 December 2020).

¹⁰³ AGCOM, "Covid-19 for users" <<https://www.agcom.it/covid-19-per-gli-utenti>> (last accessed 7 December 2020).

¹⁰⁴ See, eg, B Ponti, "The asymmetries of the monitoring unit to combat fake news on COVID-19" (*La Costituzione.info*, 7 April 2020) <<https://www.lacostituzione.info/index.php/2020/04/07/le-asimmetrie-dellunita-di-monitoraggio-per-il-contrasto-alle-fake-news-sul-covid-19/#more-6961>> (last accessed 7 December 2020).

¹⁰⁵ European Commission, "Final report of the High Level Expert Group on Fake News and Online Disinformation" (12 March 2018) <<https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>> (last accessed 12 November 2020).

¹⁰⁶ For academic debate, see, eg, J Muñoz-Machado Cañas, "Noticias falsas. confianza y configuración de la opinión pública en los tiempos de internet" [Fake news. confidence and configuration of public opinion in the times of the Internet] (2020) 86-87 *El Cronista del estado social y democrático de derecho* 122-39.

¹⁰⁷ Order PCM/1030/2020, of 30 October, by which the Action procedure against disinformation was approved by the National Security Council <<https://www.boe.es/boe/dias/2020/11/05/pdfs/BOE-A-2020-13663.pdf>> (last accessed 13 December 2020).

to decide what exactly constitutes misinformation, with no representatives from the media or journalist associations involved in the process.¹⁰⁸

The order passed by the coalition government – led by the Socialist Party (PSOE), with junior partner Unidas Podemos – is based on the concept that the use of fake news to destabilise a country or interference in public opinion by a third country constitute forms of attack.

However, opposition parties accused the government of creating a “Ministry of Truth” that would allegedly make decisions regarding content and provide media outlets with guidelines to follow. Although the Madrid Press Association acknowledged that the State needs to fight disinformation, it warned of the risk that the plan would lead the government to act more like a censor than as a guarantor of truth.

In reality, the Spanish plan leaves everything in the hands of the government and calls for the media to be consulted only if needed, but it is the media that should control the government rather than the government controlling the media.

V. Safeguarding the right of expression in international and European Union law

The administrative measures governments have recently adopted to regulate fake news may undermine fundamental rights, especially the right to freedom of expression, and the COVID-19 pandemic is emphasising this trend.¹⁰⁹ Indeed, we have seen that by governing disinformation and especially by fighting the spread of fake news¹¹⁰ on social media platforms,¹¹¹ many governments have enacted legislative and administrative measures restricting freedom of expression.¹¹²

In this section, disputing the government policies in question, I shall argue that some important principles of public law guarantee and protect freedom of expression as a fundamental individual right at the international and European levels, never subjecting it to conditions contrary to the enjoyment of such right.¹¹³

Indeed, although it might seem clear in the legal literature, discussing the guarantees of the right to freedom of expression in relation to social media platforms and especially in relation to the recent phenomenon of fake news requires a review of the main international and European rules that expressly establish the protection of this right. Such guarantees will be the focus of this section. In a democratic, free and open society, the right to express one’s opinion may not be restricted in the name of fighting fake news.

Hence, as several scholars have clearly stated, in international public law, the expression of ideas and opinions is considered a basic right of all persons as it is part of the full

¹⁰⁸ *El País*, “Spain to monitor online fake news and give a ‘political response’ to disinformation campaigns” (9 November 2020) <<https://english.elpais.com/politics/2020-11-09/spain-to-monitor-online-fake-news-and-give-a-political-response-to-disinformation-campaigns.html>> (last accessed 14 December 2020).

¹⁰⁹ See R Radu, “Fighting the ‘infodemic’: legal responses to COVID-19 disinformation” (2020) 6(3) *Social Media + Society* 1–4 <<https://doi.org/10.1177%2F2056305120948190>>. For a non-legal discussion, see S Laato, AN Islam, MN Islam and E Whelan, “What drives unverified information sharing and cyberchondria during the COVID-19 pandemic?” (2020) 29 *European Journal of Information Systems* 288–305 <<https://doi.org/10.1080/0960085X.2020.1770632>>.

¹¹⁰ Baade, *supra*, note 6, 1358.

¹¹¹ See, eg, the recent C Marsden, T Meyer and I Brown, “Platform values and democratic elections: how can the law regulate digital disinformation?” (2020) 36 *The Computer Law & Security Report* 1–18 <<https://doi.org/10.1016/j.clsr.2019.105373>>.

¹¹² See, eg, E Howie, “Protecting the human right to freedom of expression in international law” (2018) 20(1) *International Journal of Speech Language Pathology* 12–15 <<https://doi.org/10.1080/17549507.2018.1392612>>.

¹¹³ See, eg, E Barendt, *Freedom of Speech* (2nd edn, Oxford, Oxford University Press 2005); see also K Greenawalt, *Fighting Words: Individuals, Communities, and Liberties of Speech* (Princeton, NJ, Princeton University Press 1995).

development of an individual, but this right also represents a milestone for free and democratic society, providing the vehicle for the exchange and development of opinions.¹¹⁴ Many important rules of international law safeguard people's freedom of expression, so that we might at least refer to the ones we indicate below to argue for the limits governments encounter when regulating the phenomenon of fake news on social media.

Firstly, the Universal Declaration of Human Rights (UDHR), proclaimed by the UN General Assembly in Paris on 10 December 1948, ensures an effective protection of the freedom of expression pursuant to Article 19 of the UDHR in the following terms: "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 frontiers".¹¹⁵

In particular, the last part of Article 19 of the UDHR provides a legal basis for the dissemination "through any media" of information by establishing that it can be conveyed "without interference".

According to this interpretation of Article 19 of the UDHR, a free, uncensored and unhindered media is essential in any society to ensure freedom of opinion and expression as the basis for the full enjoyment of a wide range of other human rights (eg rights to freedom of assembly and association, the exercise of the right to vote). In other words, it constitutes one of the cornerstones of a democratic society.¹¹⁶

Secondly, the ICCPR, adopted by the General Assembly on 16 December 1966,¹¹⁷ guarantees freedom of expression and opinion in paragraphs 1 and 2 of Article 19, which read: "1. Everyone shall have the right to hold opinions without interference. 2. Everyone shall have the right to freedom of expression; this right shall include 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".¹¹⁸

In spite of this, paragraph 3 of Article 19 of the ICCPR imposes certain limitations on the exercise of the right to freedom of expression laid down in paragraph 2 of the same norm involving special duties and responsibilities.¹¹⁹ Consequently, governments may invoke Article 19(3) of the ICCPR relating (1) to the respect of the rights or reputations of others and (2) to ensuring the protection of national security and public order or of public health or morals.

Yet if we consider the specific terms of Article 19(1) of the ICCPR as well as the relationship between opinion and thought laid down by Article 18 of the ICCPR, we can affirm that any derogation from paragraph 1 would be incompatible with the object and purpose

¹¹⁴ See Howie, *supra*, note 112, 13. See also Human Rights Committee, communication No. 1173/2003, *Benhadj v. Algeria*, Views adopted on 20 July 2007 <http://www.worldcourts.com/hrc/eng/decisions/2007.07.20_Benhadj_v_Algeria.htm>; No. 628/1995, *Park v. Republic of Korea*, views adopted on 5 July 1996 <http://www.worldcourts.com/hrc/eng/decisions/1996.07.05_Park_v_Republic_of_Korea.htm> (last accessed 8 January 2021).

¹¹⁵ See United Nations General Assembly, Universal Declaration of Human Rights (UDHR), 10 December 1948, Paris (General Assembly resolution 217 A) <<https://www.un.org/en/universal-declaration-human-rights/>> (last accessed 8 January 2021).

¹¹⁶ See Human Rights Committee, communication No. 1128/2002, *Marques v. Angola*, Views adopted on 29 March 2005 <<http://hrlibrary.umn.edu/undocs/1128-2002.html>> (last accessed 9 January 2021).

¹¹⁷ The ICCPR has been widely ratified throughout the world, with 168 States Parties. Notably, the following States have not signed the ICCPR: Myanmar (Burma), Malaysia, Oman, Qatar, Saudi Arabia, Singapore, South Sudan and the United Arab Emirates.

¹¹⁸ See, eg, M O'Flaherty, "Freedom of expression: Article 19 of the International Covenant on Civil and Political Rights and the Human Rights Committee's General Comment No 34" (2012) 12(4) *Human Rights Law Review* 627–54 <<https://doi.org/10.1093/hrlr/ngs030>>.

¹¹⁹ It should be noted that Art 19 is also limited by another article, Art 20, which prohibits any propaganda of war or any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.

of the Covenant, even in the event of public health risks such as the COVID-19 pandemic.¹²⁰ This argument finds support in the UN Human Rights Committee claim that freedom of opinion is an element from which “it can never become necessary to derogate . . . during a state of emergency”.¹²¹

As the free communication of information and ideas about public and political issues is indispensable for democracy, the ICCPR embraces the right for the media to receive information on the basis of which it can fulfil its function.¹²² This entails a free media able to comment on public issues without censorship or restraint and to inform public opinion,¹²³ and the public consequently has a corresponding right to receive media output.¹²⁴

We can consistently note from this perspective that Article 19(2) of the ICCPR explicitly includes in this right the “freedom to seek, receive and impart information and ideas . . . through any other media of . . . choice”. Thus, governments should ensure that legislative and administrative frameworks for the regulation of social media platforms are consistent with the provisions of Article 19 of the ICCPR.

We may further observe, in accordance with the statements of the UN Human Rights Committee,¹²⁵ that any restrictions on the operation of websites, blogs or any other Internet-based, electronic or other such information dissemination system, including systems to support such communication (eg Internet service providers or search engines), are only permissible if they comply with paragraphs 1 and 2 of Article 19 of the ICCPR. In addition, permitted restrictions should generally be content-specific and, on the other hand, generic bans on the operation of certain sites and systems must be considered incompatible with Article 19 of the ICCPR. Likewise, it is also inconsistent with the same rule to forbid a site or information dissemination system to publish material solely because it may be critical of the government or the political social system it espouses.¹²⁶

In international law, other rules enshrine the right to freedom of expression, namely the International Covenant on Economic, Social and Cultural Rights (ICESCR),¹²⁷ which guarantees the right to freedom of expression under Article 15(3),¹²⁸ and the International

¹²⁰ See the UN Human Rights Committee (HRC), *CCPR General Comment No. 24: Issues Relating to Reservations Made upon Ratification or Accession to the Covenant or the Optional Protocols thereto, or in Relation to Declarations under Article 41 of the Covenant*, 4 November 1994, CCPR/C/21/Rev.1/Add.6 <<https://www.refworld.org/docid/453883fc11.html>> (last accessed 10 January 2021).

¹²¹ UN Human Rights Committee 102nd session Geneva, 11–29 July 2011 General comment No. 34, Art 19: Freedoms of opinion and expression, at para 5 <<https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf>> (last accessed 8 January 2021). General Comment No. 34 is a document adopted by the UN Human Rights Committee in July 2011 that gives States more specific guidance on the proper interpretation of Art 19 of the ICCPR.

¹²² J Farkas and J Schou, *Post-Truth, Fake News and Democracy: Mapping the Politics of Falsehood* (New York/London, Routledge 2020).

¹²³ UN Human Rights Committee (HRC), *CCPR General Comment No. 25: Article 25 (Participation in Public Affairs and the Right to Vote), The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service*, 12 July 1996, CCPR/C/21/Rev.1/Add.7 <<https://www.refworld.org/docid/453883fc22.html>> (last accessed 11 January 2021).

¹²⁴ See Human Rights Committee, communication No. 1334/2004, *Mavlonov and Sa'di v. Uzbekistan* <http://www.worldcourts.com/hrc/eng/decisions/2009.03.19_Mavlonov_v_Uzbekistan.htm> (last accessed 11 January 2021).

¹²⁵ See *supra*, note 121, at para 43 <<https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf>> (last accessed 11 January 2021).

¹²⁶ *ibid.*

¹²⁷ International Covenant on Economic, Social and Cultural Rights adopted and opened for signature, ratification and accession by General Assembly Resolution 2200A (XXI) of 16 December 1966, entered into force on 3 January 1976, in accordance with Art 27 <<https://www.ohchr.org/en/professionalinterest/pages/cescr.aspx>> (last accessed 11 January 2021). The ICESCR has been signed and ratified by 163 States Parties.

¹²⁸ Art 15(3) of the ICESCR: “The States Parties to the present Covenant undertake to respect the freedom indispensable for scientific research and creative activity”.

Convention on the Elimination of All Forms of Racial Discrimination (ICERD),¹²⁹ which expresses the right to freedom of expression under Article 5(d)(vii) and (viii), such as the rights to freedom of thought, conscience and religion and the freedom of opinion and expression, respectively.

As regards the EU's legal system, freedom of expression, media freedom and pluralism are enshrined in the EU Charter of Fundamental Rights (CFR), as well as in the European Convention on Human Rights (ECHR). Furthermore, we might claim more generally that no country can join the EU without guaranteeing freedom of expression as a basic human right according to Article 49 of the Treaty on European Union (TEU).¹³⁰

The ECHR¹³¹ recognises the right to freedom of expression under Article 10(1), which states: "Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises".¹³²

In particular, we can clearly note that Article 10(1) of the ECHR describes several components of the right to freedom of expression, including the freedom to express one's opinion and the freedom to communicate and receive information. As a matter of principle, we therefore claim that the protection given by Article 10 of the ECHR should extend to any expression, regardless of its content, disseminated by any individual, group or type of media. Therefore, based on what we have seen, we can argue that the right protected by Article 10 of the ECHR entirely covers freedom of expression on social media as it includes the right to hold opinions and to receive and impart ideas and information without interference regardless of frontiers, which are removed thanks above all to social media platforms.

The CFR¹³³ guarantees and promotes the most important freedoms and rights enjoyed by EU citizens in one legally binding document. The CFR under Article 11(1) states that "everyone has the right to freedom of expression", and at the same time it ensures that this right includes "freedom to hold opinions and to receive and impart information and ideas without interference by public authorities and regardless of frontiers".

Moreover, the CFR explicitly requires governments to respect "the freedom and pluralism of the media" in accordance with Article 11(2), thus avoiding any legislative and administrative measures that could undermine or even merely jeopardise human rights, and especially the freedom of expression.

We should bear in mind that, based on the interpretation of the ECHR and CFR Charters, in addition to the other fundamental rules of the EU legal system, the European Court of Human Rights (ECtHR) has repeatedly stated that freedom of expression constitutes "one

¹²⁹ International Convention on the Elimination of All Forms of Racial Discrimination, adopted and opened for signature and ratification by General Assembly Resolution 2106 (XX) of 21 December 1965, entered into force on 4 January 1969, in accordance with Art 19 <<https://www.ohchr.org/en/professionalinterest/pages/cerd.aspx>> (last accessed 11 January 2021). The ICERD has been ratified by 177 States.

¹³⁰ See Art 49 of the TEU <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:12012M049>> (last accessed 12 January 2021).

¹³¹ The European Convention on Human Rights was opened for signature in Rome on 4 November 1950 and came into force in 1953 <<https://www.echr.coe.int/Pages/home.aspx?p=basictexts&c>> (last accessed 12 January 2021).

¹³² Yet we should consider some of the "duties and responsibilities" set forth in Art 10(2) of the ECHR. See the seminal work by D Voorhoof, "The European Convention on Human Rights: The Right to Freedom of Expression and Information restricted by Duties and Responsibilities in a Democratic Society (2015) 7(2) Human Rights 1–40.

¹³³ The CFR was declared in 2000 and came into force in December 2009. For the official version of the Charter, see <https://eur-lex.europa.eu/eli/treaty/char_2012/oj> (last accessed 13 January 2021).

of the essential foundations of a democratic society and one of the basic conditions for its progress and for each individual's self-fulfilment".¹³⁴

In its decision-making process, the ECtHR has considered national constitutional practices that afford a high level of protection to freedom of expression and has frequently ascribed the protection of freedom of expression to the ICCPR as well as other international documents.

The Strasbourg Court has consistently underlined some primary tasks that the media must accomplish without having to suffer interference from governments in order to fulfil people's enjoyment of freedom of expression. So, in summary, media should: (1) share information and ideas concerning matters of public interest; (2) perform the leading role of public watchdog; and (3) meet the need for impartial, independent and balanced news,¹³⁵ information and comment.

The ECtHR case law principles on the right to freedom of expression are summarised in *Bédát v. Switzerland* of 29 March 2016.¹³⁶ It can basically be argued that these principles set out in the aforementioned judgment can protect fake news by including it in the exercise of the right of freedom of expression under certain conditions.

Indeed, the ECtHR recognises that freedom of expression represents one of the basic foundations of a democratic society and one of the essential conditions for its progress and each individual's self-fulfilment. Indeed, in *Bédát v. Switzerland*, the Strasbourg Court established that, under Article 10(2) of the ECHR, this right is applicable not only to "information" or "ideas" that are favourably received or regarded as inoffensive or as matters of indifference, but also to those that – as could be the case for fake news – offend, shock or disturb. Such are the demands of pluralism, tolerance and broadmindedness, without which there is no "democratic society". As set forth in Article 10, this freedom is subject to exceptions, which must, however, be construed strictly, and the need for any restrictions must be established convincingly.¹³⁷

Furthermore, we can also observe, concerning the level of protection, that there is little scope under Article 10(2) of the ECHR for restrictions on freedom of expression in two fields, namely political speech and matters of public interest.¹³⁸ Accordingly, a high level of protection of freedom of expression, with the authorities thus having a particularly narrow margin of appreciation, will normally be accorded where the remarks concern a matter of public interest, as in the case of comments on the functioning of the judiciary, even in the context of proceedings that are still pending.¹³⁹ A degree of hostility¹⁴⁰ and the

¹³⁴ ECtHR, *Lingens v. Austria*, 8 July 1986; *Şener v. Turkey*, 18 July 2000; *Thoma v. Luxembourg*, 29 March 2001; *Marônek v. Slovakia*, 19 April 2001; *Dichand and Others v. Austria*, 26 February 2002. See JF Flauss, "The European Court of Human Rights and the freedom of expression" (2009) 84(3) *Indiana Law Journal* 809. See also G Ilić, "Conception, standards and limitations of the right to freedom of expression with a special review on the practice of the European Court of Human Rights" (2018) 1 *Godišnjak Fakulteta Bezbednosti* 29–40.

¹³⁵ On this point, see M Eliantonio, F Galli and M Schaper, "A balanced data protection in the EU: conflicts and possible solutions" (2016) 23(3) *Maastricht Journal of European and Comparative Law* 391–403 <<https://doi.org/10.1177/1023263X1602300301>>.

¹³⁶ *Bédát v. Switzerland* App no 56925/08 (ECtHR GC 29 March 2016) para 48 <<http://hudoc.echr.coe.int/eng?i=001-161898>>; see also *Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland* App no 931/13 (ECtHR GC 27 June 2017) <<http://hudoc.echr.coe.int/eng?i=002-11555>>.

¹³⁷ *Bédát v. Switzerland*, supra, note 136, para 48.

¹³⁸ See *Morice v. France* App no 29369/10 (ECtHR GC 23 April 2015) para 125; see also *Sürek v. Turkey* App no 26682/95 (ECtHR GC 8 July 1999) para 61; *Lindon, Otchakovsky-Laurens and July v. France* App no 21279/02 and 36448/02 (ECtHR GC 22 October 2007) para 46; *Axel Springer AG v. Germany* App no 39954/08 (ECtHR 7 February 2012) para 90.

¹³⁹ See, *mutatis mutandis*, *Roland Dumas v. France* App no 34875/07 (ECtHR 15 July 2010) para 43; *Gouveia Gomes Fernandes and Freitas e Costa v. Portugal* App no 1529/08 (ECtHR 29 March 2011) para 47.

¹⁴⁰ See *E.K. v. Turkey*, App no 28496/95 (ECtHR 7 February 2002) paras 79 and 80.

potential seriousness of certain remarks¹⁴¹ do not obviate the right to a high level of protection, given the existence of a matter of public interest.¹⁴²

This section has analysed some important rules of international and European law and has argued that freedom of expression is protected and guaranteed as a fundamental human right, never subjecting it to conditions contrary to its enjoyment.

The review of these fundamental rules is useful to argue that government policies addressing online disinformation and especially the phenomenon of fake news are lawful only if they comply with the right to freedom of expression as they are established at the international and European levels.

The next section of this paper will discuss the main regulatory approaches – namely empowering users, self-regulation and government intervention – that States around the world might consider when regulating fake news in observance of the right to freedom of expression.

VI. Empowering users, self-regulation and government intervention

Having emphasised the importance of protecting freedom of expression in international and European law, this section surveys regulatory approaches that may be implemented by States in order to govern instances of disinformation and especially fake news.

In doing so, I thus recommend self-regulation and above all empowering users rather than government intervention as regulatory strategies for addressing fake news by minimising undue interference in the right to freedom of expression.

Basically, at least three different regulatory approaches to managing information on social media platforms could be proposed. The first is based on empowering users,¹⁴³ leveraging the ability of individuals to evaluate and detect fake news.¹⁴⁴ The second and third approaches concern the accountability of the social media platforms that can be implemented by either self-regulation or government intervention.

The first approach is essentially based on fact checking by individual websites.¹⁴⁵ Specifically, a number of major organisations in the USA, such as *PolitiFact*, *FactCheck.org*, *The Washington Post* and *Snopes*, fact check rumours, health stories and political claims, especially those that often appear on social media.¹⁴⁶ Fact checking is also carried out by reliable sources of information such as newspapers, where news is almost

¹⁴¹ See *Thoma v. Luxembourg* App no 38432/97 (ECtHR 29 March 2001) para 57.

¹⁴² See *Paturel v. France* App no 54968/00 (ECtHR 22 December 2005) para 42.

¹⁴³ See MM Madra-Sawicka, JH Nord, J Paliszkiwicz and T-R Lee, “Digital media: empowerment and equality” (2020) 11(4) *Information* 225. The authors argue that empowerment is a process by which powerless people become conscious of their situation, organise collectively to improve it and access opportunities, as an outcome of which they take control over their own lives, gain skills and solve problems. See also N Kabeer, “Resources, agency, achievements: reflections on the measurement of women’s empowerment” (1999) 30 *Development and Change* 435–64. For this author, empowerment means expanding people’s ability to make strategic life choices, particularly in the context in which this ability had been denied to them. From another point of view, according to N Wallerstein and E Bernstein, “Empowerment education: Freire’s ideas adapted to health education” (1988) 15 *Health Education and Behavior* 379–94, empowerment is a process that supports the participation of people, organisations and communities in gaining control over their lives in their community and society.

¹⁴⁴ Recently, eg, see EK Vraga et al, “Empowering users to respond to misinformation about Covid-19” (2020) 8(2) *Media and Communication (Lisboa)* 475–79. These authors claim that “if much of the misinformation circulating on social media is shared unwittingly, news and scientific literacy that helps people distinguish between good and bad information on Covid-19 could reduce the amount of misinformation shared”.

¹⁴⁵ See, eg, AL Wintersieck, “Debating the truth: the impact of fact-checking during electoral debates” (2017) 45(2) *American Politics Research* 304–31.

¹⁴⁶ For a list of fact-checking websites, cf. “Fake News & Misinformation: How to Spot and Verify” available at <<https://guides.stlcc.edu/fakenews/factchecking>>. For an extend analysis on the methodologies of three major fact-checking organisations in the USA, cf. <https://ballotpedia.org/The_methodologies_of_fact-checking>.

always subject to editorial scrutiny.¹⁴⁷ Moreover, another approach based on empowering users seeks to increase the ability of individuals to assess the quality of sources of information by educating them, although it is unclear whether such efforts can actually improve the ability to assess credibility and, if so, whether this will have long-term effects.¹⁴⁸

Yet in many cases this approach inexorably collides with social reality. Recently, the behavioural sciences have demonstrated that people are predictably irrational.¹⁴⁹ This may mean that people are irrational consumers of news as they prefer information that confirms their pre-existing attitudes (selective exposure), view information that is consistent with their pre-existing beliefs as more persuasive than dissonant information (confirmation bias¹⁵⁰) and are inclined to accept information that satisfies them (desirability bias).¹⁵¹ By preselecting information that interests them, social media users tend to reinforce their own worldviews and opinions. These aspects entail social media users tending to aggregate into ideologically homogeneous groups (communities), on which they focus their attention and from which they obtain news. Furthermore, it has been authoritatively argued that people are inclined to remain within these specific communities, thus contributing to the emergence of polarisation.¹⁵² Distinct and separate communities that do not interact with each other arise spontaneously, creating opposing groups each of which focuses only on a specific narrative – which therefore tends to strengthen and polarise within the group – ignoring the alternatives (echo chambers).¹⁵³

I am persuaded, despite the problems just mentioned, that the approach of empowering users – if properly pursued and implemented by significantly enhancing quality information and actively correcting disinformation – might represent an effective and democratic strategy to address fake news on social media platforms by avoiding undue interference with the right to freedom of expression.¹⁵⁴ The empowerment approach, in essence, can

¹⁴⁷ To take just one example, the French newspaper *Le Monde* has identified and corrected nineteen misleading statements made by Marine Le Pen, the right-wing candidate who reached the runoff of the 2017 French presidential election, during her televised debate against Emmanuel Macron; the article is available at <http://www.lemonde.fr/les-decodeurs/article/2017/05/03/des-intox-du-debat-entre-emmanuel-macron-et-marine-le-pen-verifiees_5121846_4355770.html> (last accessed 26 January 2021).

¹⁴⁸ D Lazer et al, “The science of fake news: addressing fake news requires a multidisciplinary effort” (2018) 359(6380) *Science* 1094–96.

¹⁴⁹ D Ariely, *Predictably Irrational: The Hidden Forces that Shape our Decisions* (New York, Harper Collins 2009).

¹⁵⁰ See RS Nickerson, “Confirmation bias: a ubiquitous phenomenon in many guises” (1998) 2(2) *Review of General Psychology* 175–220.

¹⁵¹ Lazer et al, *supra*, note 148, 1095.

¹⁵² CR Sunstein, *Republic.com 2.0* (Princeton, NJ, Princeton University Press 2009) pp 57–58. According to Sunstein’s point of view, the polarisation of a group is a phenomenon whereby individuals, after deliberation with likeminded individuals, are likely to adopt a more extreme position than the one they originally held. From another standpoint, Sunstein also considers fragmentation, arguing that the Internet’s ability to reinforce narrow interests encourages self-isolation, which, in turn, leads to group polarisation. For a discussion of fragmentation and the Internet, see CR Sunstein, “Deliberative trouble? Why groups go to extremes” (2000) 110 *Yale Law Journal* 71.

¹⁵³ See, eg, RK Garrett, “Echo chambers online? Politically motivated selective exposure among Internet news users” (2009) 14 *Journal of Computer-Mediated Communication* 265–85. See also S Flaxman, S Goel and JM Rao, “Filter echo chambers, and online news consumption” (2016) 80(S1) *Public Opinion Quarterly* 298–320.

¹⁵⁴ See Z Li, “Psychological empowerment on social media: who are the empowered users?” (2016) 42(1) *Public Relations Review* 50–52 and the review of the literature described therein. In academia, according to the leading scholarship, empowerment is a multi-level, open-ended construct that includes the individual level (see A Schneider, G Von Krogh and P Jäger, “What’s coming next? Epistemic curiosity and lurking behavior in online communities” (2013) 29(1) *Computers in Human Behavior* 293–303), the organisational level (see NA Peterson and MA Zimmerman, “Beyond the individual: toward a nomological network of organizational empowerment” (2004) 34(1–2) *American Journal of Community Psychology* 129–45) and the community level (see MA Zimmerman,

create a sound foundation for dealing with fake news by providing social media users with the tools to detect and share high-quality information.

Empowering users could potentially allow social media platforms to play a leading role in reducing the spread and impact of fake news through algorithms and bots.¹⁵⁵ Major platforms like Google, WhatsApp, Twitter and Facebook may use complex statistical models that predict and maximise engagement with content in order to improve the quality of information.¹⁵⁶ The platforms can therefore signal to users the quality of the information and their relative sources by incorporating these signals in the algorithmic rankings of their content. They can similarly decrease the personalisation of political information relating to other types of content, thus reducing the phenomenon of “echo chambers.”

Likewise, the platforms can effectively minimise the impact of the automated dissemination of content using bots (ie users who automatically share news from a set of sources with or without reading them). Some of the major platforms, especially Facebook and Twitter, have recently pursued the aim of managing fake news by modifying their algorithms to improve the quality of their content and counteracting bots that disseminate disinformation.

The second and third approaches refer to self-regulation and government intervention, respectively. We will analyse them here by comparing the disadvantages and drawbacks of the latter with the advantages and benefits of self-regulation, bearing in mind that the policies of the States we have discussed were essentially based on government intervention (see *supra*, Sections III and IV).

Self-regulation generally increases online accountability while providing more flexibility and safeguards than government intervention. Indeed, self-regulation may help preserve the independence of social platforms by safeguarding them from government interference. Fundamentally, self-regulatory mechanisms aim to foster public trust in the media. In fact, through self-regulation, social media platforms understand how they work best and therefore have an incentive to provide effective rules over more short-sighted and therefore more invasive government measures.

As social networks have caused considerable concern, States around the world have increasingly called for government intervention in order to protect their citizens from (fake) news that is considered harmful. Legislative and administrative measures have shown, however, that government intervention pursuing legitimate goals can easily cause

“Empowerment theory: psychological, organizational and community levels of analysis” in J Rappaport and E Seidman (eds), *Handbook of Community Psychology* (New York, Plenum Press 2000) pp 43–63.

¹⁵⁵ See European Commission, Report of the independent High Level Group on fake news and online disinformation, “A multi-dimensional approach to disinformation”. According to this Report, the platforms should implement the development of built-in tools/plugin and applications for browsers and smartphones to empower users to better control access to digital information. In particular, platforms should consider ways to encourage users’ control over the selection of the content to be displayed as results of a search and/or in news feeds. Such a system should give the user the opportunity to have content displayed according to quality signals. Moreover, content recommendation systems that expose different sources and different viewpoints around trending topics should be made available to users in online platforms. It should give users a certain degree of control.

¹⁵⁶ However, in German case law, incentivising the visibility of government sources has been found to be illegal under competition law. Specifically, Google had favoured information from the German Ministry of Health by placing it “at the top” of its search engine results. In this regard, the Munich Regional Court (Landgericht München I) has provisionally prohibited a cooperation between the Federal Government and the Internet company Google on a health portal. The judges essentially granted two applications *ad interim* injunctions against the Federal Republic, represented by the Federal Ministry of Health, and the US company Google, as the Regional Court announced. The judgments are not legally binding. The Federal Government and Google will first examine the decision. For more details, see the German press release at <<https://rsw.beck.de/aktuell/daily/meldung/detail/kooperation-zwischen-gesundheitsministerium-und-google-untersagt>> (last accessed 10 September 2021).

negative side effects, including becoming a tool for suppressing opposition and critics (see *supra*, Sections III and IV).

By contrast, as little government intervention as possible is required if the media are to continue to fulfil their role as watchdogs of democracy. Thus, the self-regulatory approach can help prevent stringent legislative and administrative measures against social media platforms, which would undermine rather than protect fundamental rights such as the freedom of expression. Moreover, although self-regulation might provide an alternative to courts for resolving media content complaints in this context, members of the public can still choose to take matters to court, as this remains an irrepressible core human right.

The social network environment makes legal supervision difficult and therefore opens up new prospects for social media self-regulation. First of all, because we are experiencing a time of rapid and constant change in media technology, self-regulation offers more flexibility than the government regulation option. Secondly, self-regulation is less costly for States and society in general. However, to avoid the risk of self-regulation benefitting (only) the interests of social media companies, provisions on transparency and efficiency must be reinforced.

A coherent strategy based on self-regulation and empowering users with aim of addressing the spread of disinformation, and especially fake news, across social networks has recently been implemented in EU law. To this end, the European Commission passed the so-called Code of Practice on Disinformation (CoP).¹⁵⁷ The CoP was warmly welcomed by representatives of leading social media platforms (as well as advertisers and the advertising industry) who volunteered to establish self-regulatory standards to fight online disinformation. Indeed, media giants such as Facebook, Google, Twitter, Mozilla, Microsoft and TikTok officially signed the CoP and accepted the rules on periodic European Commission monitoring.¹⁵⁸

In 2019, the European Commission implemented monitoring policies to define the progress of the measures adopted by social media platforms to comply with the commitments envisaged by the CoP. In particular, the European Commission asked platforms to report on a monthly basis on the actions undertaken to improve the scrutiny of ad placements, to ensure transparency of political and issue-based advertising and to tackle fake news and the malicious use of bots. In addition, further monitoring actions have been undertaken to counter the spread of fake news during the COVID-19 pandemic. The platforms that signed the CoP have stepped up their efforts against disinformation and fake news on the coronavirus by providing specific reports on the actions taken to implement transparency.¹⁵⁹ As a result, a first general study to assess the implementation of the CoP has been released.¹⁶⁰

Concerning the fake news monitoring policies put in place by social media platforms during the health crisis, signatories effectively joined the *ad hoc* “Fighting COVID-19 Disinformation Monitoring Programme”.¹⁶¹ Indeed, the COVID-19 disinformation

¹⁵⁷ The EU Code of Practice of Disinformation <<https://ec.europa.eu/digital-single-market/en/code-practice-disinformation>>.

¹⁵⁸ See, eg, M Dando and J Kennedy, “Combating fake news: The EU Code of Practice on Disinformation” (2019) 30(2) Entertainment Law Review 44–42.

¹⁵⁹ The European Commission published “First results of the EU Code of Practice against disinformation” (29 January 2019) available at <<https://ec.europa.eu/digital-single-market/en/news/first-results-eu-code-practice-against-disinformation>> (last accessed 7 March 2021). For more details, see RÓ Fathaigh, “European Union. European Commission: Reports on the Code of Practice on Disinformation” (2019) 3(3) Iris 1–5.

¹⁶⁰ See I Plasilova et al, “Study for the assessment of the implementation of the Code of Practice on Disinformation Final Report” Publications Office, 2020, available at <<https://op.europa.eu/en/publication-detail/-/publication/37112cb8-e80e-11ea-ad25-01aa75ed71a1/language-en>>.

¹⁶¹ First baseline reports – Fighting COVID-19 Disinformation Monitoring Programme <<https://digital-strategy.ec.europa.eu/en/library/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme>> (last accessed 19 June 2021).

monitoring programme has provided an in-depth overview of the actions taken by platforms to fight false and misleading information regarding coronavirus and vaccines. It has proven to be a useful transparency measure to ensure platforms' public accountability and has put the CoP through a stress test.

Specifically, the signatories to the CoP have been requested to provide information regarding: (1) the initiatives to promote authoritative content at the EU and Member State levels; (2) tools to improve users' awareness; (3) information on manipulative behaviour on their services; and (4) data on flows of advertising linked to COVID-19 disinformation on their services and on third-party websites.

Fundamentally, the baseline reports from Facebook, Google, Microsoft-LinkedIn, TikTok, Twitter and Mozilla summarise the actions taken by these platforms to reduce the spread of false and misleading information on their services, covering a period from the beginning of the health emergency until 31 July 2020. More importantly, these reports offer a comprehensive overview of the relevant actions. Overall, baseline reports demonstrate that the signatories to the CoP have intensified their efforts compared with the first year of implementation of the Code's commitments.

In general, it can be said that the platforms have enhanced the visibility of authoritative sources by conferring significance to COVID-19 information from the World Health Organization (WHO) and national health authorities and by providing new tools and services to facilitate access to relevant and reliable information concerning the health emergency.

Furthermore, the reports reveal that the platforms have addressed a considerable quantity of content encompassing false or misleading information, especially by removing or degrading content liable to cause physical harm or weaken public health policies. From this perspective, platforms have increased their efforts to detect cases of social media manipulation, operations intended to have a negative influence or coordinated inauthentic behaviour. In doing so, platforms did not detect coordinated disinformation operations with a specific focus on COVID-19 on their services, even if they detected a high number of items including false information concerning the coronavirus.

In addition, it may be noted that these reports emphasise strong measures to reduce the flow of advertising on third-party web pages supplying disinformation about the coronavirus, while providing free COVID-19-related advertising space for government and public health authorities.

In detail, the reports encompass quantitative data exemplifying the impact of social media platform policies. In particular, we can affirm that Google has placed importance to articles published by EU fact-checking organisations, which generated more than 155 million impressions over the first half of 2020.¹⁶² With regard to Mozilla, this platform has enhanced the use of the browser space (Firefox snippets) and features (Pocket) to promote important public health information from the WHO, leading to more than 35 million impressions and 25,000 clicks on the snippets in Germany and France alone, while the curated coronavirus hub in Pocket generated more than 800,000 page views from more than 500,000 users around the globe. It has also provided expertise and opened up datasets on Firefox usage in February and March to help researchers investigating social distancing measures.¹⁶³

We can also see how Microsoft-LinkedIn shares with interested members a "European Daily Rundown", namely a summary of the day's news written and curated by experienced journalists and distributed to members in all twenty-seven EU Member States. The

¹⁶² Cf. Google COVID-19 report – August 2020 <<https://digital-strategy.ec.europa.eu/en/library/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme>> (last accessed 19 June 2021).

¹⁶³ Cf. Mozilla COVID-19 report – August 2020 <<https://digital-strategy.ec.europa.eu/en/library/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme>> (last accessed 19 June 2021).

“European Daily Rundown” has a reach of approximately 9.7 million users in the EU.¹⁶⁴ Regarding Facebook, the platform has referred over two billion people globally to resources from the WHO and other public health authorities through its “COVID-19 Information Center”, with over 600 million people clicking through to learn more.¹⁶⁵ Twitter’s COVID-19 information pages were visited by 160 million users. Such pages bring together the latest tweets from a number of authoritative and trustworthy government, media and civil society sources in local languages.¹⁶⁶ Lastly, we can also note that TikTok’s informational page on COVID-19 has been visited over 52 million times across their five major European markets (the UK, Germany, France, Italy and Spain).¹⁶⁷

Legally speaking, the CoP aims to achieve the objectives defined by the European Commission in its Communication COM/2018/236, “Tackling Online Disinformation: A European Approach”, presented in April 2018,¹⁶⁸ by setting a wide range of commitments, from transparency in political advertising to the demonetisation of purveyors of disinformation. Communication COM/2018/236 in fact outlined the key overarching principles and objectives that should guide the actions of Member States to raise public awareness about disinformation and fake news. I contend that, in doing so, the EU has correctly interpreted self-regulation as a more effective approach than government intervention for addressing fake news as it is more flexible and above all less detrimental to users’ rights than the severe administrative measures that governments might adopt.

However, what I wish to emphasise even more strongly is that the European Commission has rightly recommended implementing user empowering. In this regard, Section I point (x) of the CoP seeks directly to promote the empowerment of users through tools enabling a customised and interactive online experience so as to enable users to fully grasp the meaning of content and to easily access different news sources reflecting alternative viewpoints, as well as providing appropriate and effective means for them to report disinformation and fake news.

To do so, Section I point (xi) of the CoP fosters “fact-checking activities” on social media platforms with reasonable measures to enable privacy-compliant access to data and to cooperate by providing relevant data on how their services function, including data for independent investigation by academic researchers and general information on algorithms.¹⁶⁹

¹⁶⁴ Cf. Microsoft–LinkedIn COVID-19 report – August 2020 <<https://digital-strategy.ec.europa.eu/en/library/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme>> (last accessed 19 June 2021).

¹⁶⁵ Cf. Facebook COVID-19 report – August 2020 <<https://digital-strategy.ec.europa.eu/en/library/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme>> (last accessed 19 June 2021).

¹⁶⁶ Cf. Twitter COVID-19 report – August 2020 <<https://digital-strategy.ec.europa.eu/en/library/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme>> (last accessed 19 June 2021).

¹⁶⁷ Cf. TikTok COVID-19 report – August 2020 <<https://digital-strategy.ec.europa.eu/en/library/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme>> (last accessed 19 June 2021).

¹⁶⁸ The Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, “Tackling Online disinformation: A European Approach” COM/2018/236 final <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236>>.

¹⁶⁹ Arguably, from this perspective, it should be noted that German higher courts are beginning to issue the first rulings on fact checking by arguing that social media platforms, such as Facebook, must take into account the guarantee of freedom of expression when applying fact checking on users. Basically, this case law requires a higher threshold for deleting fake news. In this regard, see Der 6. Zivilsenat des Oberlandesgerichts Karlsruhe [The 6th Civil Senate of the Higher Regional Court of Karlsruhe], which is responsible, *inter alia*, for disputes concerning unfair competition. It issued an urgent decision on 27 May 2020 on the requirements for the presentation of a fact check on Facebook. In particular, the 6th Civil Senate granted the emergency application for an injunction against the defendant’s specific entry in the plaintiff’s post, which was based on an infringement of competition law, and amended the judgment of the Mannheim Regional Court, which had reached the opposite conclusion, accordingly. The decisive factor was that the concrete design of the test entry was, in the opinion of the Senate, misleading for the average Facebook user. Specifically, the linking of the entries on Facebook could be misunderstood to mean that the check and the objections referred to the plaintiff’s reporting, instead of – as was

A first example that seems to go in this direction is represented by the Facebook Oversight Board. In 2018, the Founder and Chief Executive Officer of Facebook, Mark Zuckerberg, announced the creation of this board with the aim of ensuring an additional and independent control policy on content removal or account suspensions for alleged violations of Facebook community rules.¹⁷⁰ Indeed, the scope of the Oversight Board – a body of independent experts who review Facebook’s most challenging content decisions and focus on important and disputed cases – was to serve as an appellate review system for user content and to make content-moderation policy recommendations to Facebook.¹⁷¹

It has been argued that the Oversight Board exemplifies an important innovation and represented an appreciated attempt to disperse the enormous power over online discourse held by Facebook. Specifically, it could help to make that power more transparent and legitimate by encouraging dialogue around how and why Facebook’s power is exercised in the first place. However, this represents a more modest purpose than becoming an independent source of universally accepted free speech norms, but it is still ambitious for an institution that is breaking new ground.¹⁷²

More importantly, without going into the question of whether it can authoritatively resolve clashing ideas of freedom of expression, the fact of establishing an autonomous board that judges potentially competent rights in the right balance certainly represents a more substantial advantage over Facebook’s internal moderation practices that have proved incomplete and largely ineffective. The Facebook Oversight Board might represent the first “platform-scaled moment” of transnational Internet adjudication of online speech. Thus, it means a step towards empowering users by involving them in private platform governance and providing them with a minimum of procedural due process.¹⁷³

It should be borne in mind, however, that this goal cannot be effectively achieved by Facebook or other social media platforms in the absence of targeted government policies that facilitate the implementation of user empowerment strategies.

Hence, and more generally, I claim that empowering users could be a crucial task that the Member States of the EU and other States around the world should take on in the future. Empowering users could become a key challenge, especially in order to avoid or at least limit stringent government interventions and ineffective or counterproductive administrative measures that would undermine fundamental human rights.

As it is based on a careful combination of self-regulation and empowering users, we can reasonably sustain that the recent regulatory approach applied in the European Commission’s CoP might play a role in national policies to face the challenge of fake news. Nevertheless, it is also my opinion that careful adjustments will need to be made to regulatory strategies depending on the social, political and legal frameworks in which they are

actually the case for the most part in the opinion of the Senate – to the “open letter”, which the plaintiff had only reported on. It should be noted that, according to the ruling of the 6th Civil Senate of the Higher Regional Court of Karlsruhe, the legality of fact checks on Facebook in general has not been decided in such proceedings. A press release on the fact-checking jurisdiction by the court is available online in German at <<https://oberlandesgericht-karlsruhe.justiz-bw.de/pb/,Lde/6315824/?LISTPAGE=1149727>> (last accessed on 10 September 2021). See also the German Federal Constitutional Court’s Order of 22 May 2019, 1 BvQ 42/19 on account deletion <https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2019/05/qk20190522_1bvq004219en.html> (last accessed on 30 August 2021). The 2019 Supreme Court decision, although mainly on a hate speech case, can also be considered similar for fact-checking cases, as it is on the effects of third parties on the right to freedom of expression for monopolists of online platforms.

¹⁷⁰ For an in-depth analysis, see K Klonick, “The Facebook Oversight Board: creating an independent institution to adjudicate online free expression” (2020) 129(8) *Yale Law Journal* 2448–73.

¹⁷¹ *ibid.*, 2464.

¹⁷² See E Douek, “Facebook’s ‘Oversight Board’: move fast with stable infrastructure and humility” (2019) 21(1) *North Carolina Journal of Law and Technology* 76, available at <<https://scholarship.law.unc.edu/ncjolt/vol21/iss1/2>>.

¹⁷³ Klonick, *supra*, note 170, at 2499.

to be implemented, so that any measures will be proportionate and adequate to the risks to be managed, the problems to be solved and, ultimately, the human rights to be protected.

Theoretically, as I underline here, if we try to look beyond fake news as a negative phenomenon and start considering it as a shared concern of “mature” democracies, and at the same time if we bear in mind the key role that social media play as a watchdog in a democratic society,¹⁷⁴ then we might wonder whether we will still be willing to accept government policies that violate our freedom of expression.

Arguably, a first and significant finding in this direction emerges in the European Commission’s policy, which seems to take these aspects into due consideration in the legal context of the CoP. As a matter of fact, in line with Article 10 of the ECHR and the principle of freedom of expression, Section I point (vii) of the CoP establishes that social media platforms should not be compelled by governments, nor should they adopt voluntary policies, to delete or prevent access to otherwise lawful content or messages solely on the basis that they are thought to be false.¹⁷⁵

Lastly, *de jure condendo*, some consideration should be made on the Digital Services Act (DSA), with particular regard to the empowerment of users.¹⁷⁶ Indeed, the proposal for regulation on a Single Market for Digital Services – namely the DSA – represents one of the key measures within the European strategy for digital.¹⁷⁷ In line with what was announced by the European Commission in the Communication “Shaping Europe’s Digital Future”,¹⁷⁸ the initiative was presented with a view to an overall review of the European-based regulatory body. In particular, the DSA aims, on the one hand, to increase and harmonise the responsibilities of social media platforms and information service providers by strengthening control over the content policies of platforms in the EU and, on the other hand, to introduce rules to ensure the fairness and contestability of digital markets.

More specifically, the European Commission has taken into consideration the recent emergence of three fundamental problems: firstly, the increased exposure of citizens to online risks, with particular regard to damage from illegal activities and violations of fundamental rights; secondly, the coordination and effectiveness of supervision of platforms, which are considered ineffective due to the limited administrative cooperation framework established by the E-Commerce Directive to address cross-border issues; and thirdly, the fragmented legal landscape due to the first initiatives to regulate digital services at a national level – this has resulted in new barriers in the internal market that

¹⁷⁴ The approach that describes the media as a watchdog is beginning to be seen as insufficient. To this end, see the recent case law of the German constitutional court analysing the role of public service media as a counterweight to disinformation.

¹⁷⁵ Recently, see I Katsirea, “‘Fake news’: reconsidering the value of untruthful expression in the face of regulatory uncertainty” (2019) 10(2) *Journal of Media Law* 159–88, which criticised “the restriction of ‘fake news’ unless if there was a pressing social need”.

¹⁷⁶ Recently, to fight disinformation, there is also the new proposed Artificial Intelligence (AI) Act by the Commission with labelling rules for chatbots and deep fakes interacting with DSA (eg Art 52 of the AI Act). On this AI Act, see J Kalbhenn, “Designvorgaben für Chatbots, Deep fakes und Emotionserkennungssysteme: Der Vorschlag der Europäischen Kommission zu einer KI-VO als Erweiterung der medienrechtlichen Plattformregulierung [Design specifications for chatbots, deep fakes and emotion recognition systems: the European Commission’s proposal for an AI regulation as an extension of media law platform regulation] (2021) 8–9 *ZUM – Zeitschrift für Urheber- und Medienrecht* 663–74. See also M Vaele and BF Zuiderveen, “Demystifying the draft EU Artificial Intelligence Act” (2021) 4 *Computer Law Review International* 97–112.

¹⁷⁷ Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, COM/2020/825 final, Brussels, 15 December 2020, 2020/0361(COD) <<https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN>> (last accessed 20 June 2021).

¹⁷⁸ EC Communication, “Shaping Europe’s Digital Future” (19 February 2020) <https://ec.europa.eu/info/publications/communication-shaping-europes-digital-future_it> (last accessed 20 June 2020).

have produced competitive advantages for already-existing very large platforms and digital services.

It should be noted that the DSA seeks to ensure the proper functioning of the single market in relation to the provision of online intermediation services across borders. It sets out a number of specific objectives, such as: (1) maintaining a secure online environment; (2) improving the conditions for innovative cross-border digital services; (3) establishing effective supervision of digital services and collaboration between authorities; and especially (4) empowering users and protecting their fundamental rights online.

Regarding empowering users, the DSA upholds the general principle that “what is illegal offline must also be illegal online”.¹⁷⁹ With this in mind, the new regulation introduces: (1) new harmonised procedures for the faster removal of illegal contents, products and/or services; (2) more effective protection of online users’ rights and internal complaint-management systems – including mechanisms for user reporting and new obligations regarding vendor traceability; and (3) a general framework of enforcement of the legislation thanks to the coordination between the national authorities, and particularly through the designation of the new figure of the “digital services coordinator”.

With respect to self-regulation, the DSA is based on the rationale that regulating social media companies through the CoP was not sufficient to address online harms. Indeed, in my opinion, the risk assessment and mitigation set forth in Articles 26 and 27 of the DSA can be viewed as an effective self-regulatory tool. Moreover, from this perspective, it can reasonably be said that the code of conduct has not been successful, which has led to the DSA having stringent specifications.

It must be said that the DSA has its strengths and weaknesses. The main objective of subjecting social platforms to specific obligations, hitherto substantially ungoverned by national and European regulatory frameworks, can be advantageous, but only if it serves to promote freedom of expression and not just the security of the platforms themselves.

To this end, from the perspective of empowering users, transparency policies should be encouraged in order to provide users with the tools to fully understand the potential of the digital environment on social media platforms and therefore to actively participate in the public debate.

Finally, the motto “what is illegal offline is illegal offline” cannot be taken literally. In my opinion, the particular context must be considered – namely the online and digital sphere, in which the rights and duties of both users and platforms are exercised.

VII. Conclusion

This article seeks to challenge the stringent legislative and administrative measures governments have recently put in place, analysing their negative implications for the right to freedom of expression and suggesting different regulatory approaches in the context of public law.

It began by exploring the legal definition of fake news in academia in order to establish the essential characteristics of the phenomenon (Section II).

It then went on to assess the legislative and administrative measures implemented by governments at both the international and EU levels (Sections III and IV, respectively), showing how they risk undermining a core human right by curtailing freedom of expression, but adding that many governments worldwide are regulating the spread of information on social media under the pretext of addressing fake news.

¹⁷⁹ In this regard, it should be mentioned that the DSA promotes empowering users when it states that recommender systems should offer a no-profiling option (eg Art 28 DSA).

I have emphasised how governments are doing this above all to prevent the risk of an uncontrolled dissemination of information that could well increase the adverse effects of the COVID-19 health emergency by exploiting false and misleading news.

This paper also claims that there is an equally non-negligible risk that governments might use the health emergency as a ruse for implementing draconian restrictions on the right to freedom of expression, in addition to increasing social media censorship (eg chilling effects).

Specifically, starting from the premise of social media as a “watchdog” of democracy and moving on to the contention that fake news is a phenomenon of “mature” democracy, I have argued that public law already protects freedom of expression and ensures its effectiveness at the international and EU levels through some fundamental rules (Section V).¹⁸⁰

Lastly, I have explored key regulatory approaches and, as an alternative to government intervention, have proposed empowering users and self-regulation as strategies to manage fake news by mitigating the risks of undue interference in the right to freedom of expression (Section VI).

To conclude, in this section, I offer some remarks on the proposed solution by recommending the implementation of legal tools such as “reliability ratings” of social media platforms in order to enhance the management of information and particularly to minimise the risks related to fake news, especially in a time of emergency.

The regulatory approaches proposed to manage fake news phenomena on social media, self-regulation and especially empowering users (see *supra*, Section VI) might be implemented by public policies focused on long-term behavioural incentives rather than stringent short-term legislative and administrative measures.

As a matter of fact, behavioural science, and particularly psychology,¹⁸¹ can play a leading role in public policy.¹⁸² Recent important surveys on the behavioural approach support the long-term effectiveness of active psychological inoculation as a means of building resistance to fake news and disinformation in general.¹⁸³

¹⁸⁰ Recently, see RK Helm and H Nasu, “Regulatory responses to ‘fake news’ and freedom of expression: normative and empirical evaluation” (2021) 21(2) Human Rights Law Review 302–28 <<https://doi.org/10.1093/hrlr/ngaa060>>. I do not doubt that there are well-founded reasons why the authors argue to implement criminal sanctions in order to address fake news. However, I am not persuaded by the authors’ argument when they affirm that “criminal sanctions as an effective regulatory response due to their deterrent effect, based on the conventional wisdom of criminal law, against the creation and distribution of such news in the first place” (cf. p 323, and more generally p 303, where it has been said that “this article identifies, albeit counter-intuitively, criminal sanction as an effective regulatory response”). Based on what I have argued in this article, the goal should not be to “worship freedom of expression” (p 326), but to protect it as a fundamental human right. Basically, the use of criminal sanction might undermine people’s rights more than is necessary to protect other interests, namely national security, public order and public health, and therefore should be avoided whenever other legal instruments that are less invasive than human rights can be used. Thus, in my opinion, the implementation of criminal sanctions ought to represent an “*extrema ratio*” for government policies precisely in order not to jeopardise fundamental human rights.

¹⁸¹ See R Greifeneder, ME Jaffé, EJ Newman and N Schwarz, *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation* (London, Routledge 2021), especially Part II, pp 73–90. See also Li, *supra*, note 154, at 51, Section 2.3, “Social media empowerment”. See also JR Anderson, *Cognitive Science Series. The Architecture of Cognition* (Hillsdale, NJ, Erlbaum 1983).

¹⁸² See CR Sunstein, *Behavioral Science and Public Policy* (Cambridge, Cambridge University Press 2020). See also the seminal book RH Thaler and CR Sunstein, *Nudge: Improving Decisions about Health, Wealth and Happiness* (London, Penguin 2009).

¹⁸³ See R Maertens, J Roozenbeek, M Basol and S van der Linden, “Long-term effectiveness of inoculation against misinformation: three longitudinal experiments” (2021) 27(1) Journal of Experimental Psychology 1–16 <<https://doi.org/10.1037/xap0000315>>. See also S van der Linden and J Roozenbeek, “Psychological inoculation against fake news” in R Greifeneder, M Jaffé, EJ Newman and N Schwarz (eds), *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation* (London, Routledge 2020) <<https://doi.org/10.4324/9780429295379-11>>. In addition, see J Cook, S Lewandowsky and UKH Ecker, “Neutralizing misinformation through inoculation:

Overall, these findings might strengthen the argument in favour of introducing legal tools for measuring the trustworthiness of information on social media platforms. To do so, in my opinion, governments could promote policies empowering users through a reliability rating mechanism.

In this regard, governments may set up the management of reliability ratings through independent third-party bodies, namely regulatory agencies or authorities, in order to carry out the various steps of the procedure transparently, checking its correctness and reporting the results to the public.

Specifically, reliability ratings can measure the frequency and percentage of fake news spread on a social media platform over a certain time frame. Thus, based on the frequency and percentage of fake news it disseminates, the platform could be considered statistically reliable or otherwise by those using it.¹⁸⁴

It might be reasonable to assume that reliability ratings may act as an incentive to enhance the efficient management of fake news among both users and owners of social media. If users can consciously decide whether or not to use a certain platform based on reliability ratings, the owner would probably be induced to adopt more effective measures to deal with fake news in order not to lose users.

At the same time, by maximising its reliability rating, one platform may incentivise another to do so too in order not to lose its reputation and therefore its users. Indeed, the threat of losing users would incentivise a platform to ensure the reliability of information in a better and more effective way. In other words, reliability ratings can stimulate competition among social media platforms to effectively manage information and address fake news.

It should be noted that reliability ratings also reveal the importance of self-regulation. To avoid or at least minimise the risk of losing their reputation and users, the owners of social media platforms will certainly implement effective legal systems to regulate information and mitigate fake news phenomena.

Hence, it may be reasonable to argue, from a legal point of view, that empowering users can play a crucial and proactive role in effectively managing fake news on social media in the era of emergency for at least three compelling reasons: (1) to build and maintain the trust of other users; (2) to ensure and preserve the function of social media as a watchdog in democratic society; and (3) to promote and protect freedom of expression by limiting undue interference by regulators.

On this last point, it must be said that reliability ratings in governing fake news – especially in emergencies – should not interfere with users' freedom of expression. Users, in my view, remain absolutely “free” to consider a social media platform trustworthy or not based on their own personal and unquestionable opinion. In fact, the

exposing misleading argumentation techniques reduces their influence” (2017) 12 PLoS ONE e0175799 <<https://doi.org/10.1371/journal.pone.0175799>>.

¹⁸⁴ Recently, with respect to reliability ratings in social media, see A Kim, P Moravec and AR Dennis, “Combating fake news on social media with source ratings: the effects of user and expert reputation ratings” (2019) 36(3) *Journal of Management Information Systems* 931–68, also available at SSRN <<https://ssrn.com/abstract=3090355>> or <<https://doi.org/10.2139/ssrn.3090355>>. The authors define and explain some reputation ratings, namely: (1) the expert rating, where expert fact checkers rate articles and these ratings are aggregated to provide an overall source rating related to new articles as they are published; (2) the user article rating, where users rate articles and these ratings are aggregated to provide the source rating on new articles; and (3) the user source rating, where users directly rate the sources without considering any specific articles from the source. On credibility evaluation online, see MJ Metzger, AJ Flanagin and RB Medders, “Social and heuristic approaches to credibility evaluation online” (2010) 60 *Journal of Communication* 413–39 <<https://doi.org/10.1111/j.1460-2466.2010.01488.x>>. On the same point, see also BK Kaye and TJ Johnson, “Strengthening the core: examining interactivity, credibility, and reliance as measures of social media use” (2016) 11(3) *Electronic News* 145–65 <<https://doi.org/10.1177/1931243116672262>>.

contents of a platform that has implemented reliability ratings should not be “labelled” as reliable *a priori*, precisely because they are subject to the right of freedom of expression.

It should also be said that in order not to create oligopolies or monopolies, reliability ratings should not per se create advantageous positions when users search for news on social media platforms using Internet search engines (eg because news from “reliable” platforms is placed by default “at the top” in the search results on the Internet and thus becomes more visible in general).

Specifically, an effective and, above all, “democratic” reliability rating model should promote user participation in the *ex post* evaluation of a platform’s reliability. In doing so, the reliability rating can take into account the diversity of opinions typical of freedom of expression and an open and plural society, promoting the comparison between different thoughts on the same news.

Lastly, at the same time, involving users in the debate on the effective reliability of a social media platform could trigger mechanisms of democratic control by the public in order to govern fake news in the “era of emergency”.