TRUTH AND DECEPTION ACROSS THE ATLANTIC: A ROADMAP OF DISINFORMATION IN THE US AND EUROPE

*Oreste Pollicino**
*Elettra Bietti***

Abstract

This paper tries to offer a bird's eye view of the complex dynamics and legal constraints that shape the digital information ecosystem, and how lawyers and policy-makers should think about possible solutions to the issues at hand. The Authors believe that some action against disinformation is needed, and tend to favour actions that regulate platforms rather than direct regulation by the state, eg. ensuring that platforms have effective mechanisms for eradicating fake account and coordinate disinformation efforts, ensuring greater transparency and traceability of disinformation and the financial incentives related to it, ensuring appropriate remedies for individuals affected. It seems that governments and institutions around the world, including some European countries, are so eager to regulate fake news that they might overstep their legitimacy bounds in doing so. The Authors warn against that, advocating a nuanced approach which takes into account the specific political and technical circumstances in which platforms and states operate to devise adequate measures for regulating online speech in the digital economy.

TABLE OF CONTENTS

1. Introduction

When, in October 2017, Donald Trump claimed to have coined the term "fake news" many believed him.[1] The term in fact exists in the United States at least since 1890,[2] but the advent of the internet and digital culture seems to have exponentially increased its use and salience in at least two ways, being used both as a diagnostic for an increasingly complex and harmful information ecosystem, and as a kind of political shield which in turn contaminates public discourse.[3] With the renewal of interest in

* Full Professor of Constitutional Law and Media Law, Bocconi University, drafted sections 1-2.
** SJD Candidate at Harvard Law School, drafted sections 3-4. The paper is the product of a joint collaboration.

[1] C. Cillizza, *Donald Trump just claimed he invented "fake news"*, CNN (October 26, 2017) https://www.cnn.com/2017/10/08/politics/trump-huckabee-fake/index.html (last visited Jun 14, 2018).

[2] *The Real Story of "Fake News"*, Merriam Webster Blog, https://www.merriam-webster.com/words-at-play/the-real-story-of-fake-news (last visited Jun 14, 2018). An example is as follows: "*Secretary Brunnell Declares Fake News About His People is Being Telegraphed Over the Country*," headline of the Cincinnati Commercial Tribune (Cincinnati, OH), 7 Jun. 1890. More recently, the coining of this term has been attributed to the journalist Craig Silverman, see C. Silveaman, *This Analysis shows How Viral Fake Election News Stories Outperformed Real News on Facebook*, Buzzfeed News (November 16, 2016), https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook.

[3] *See* graph from Google Ngram, use of the term "fake news" in books from 1800 to 2008. One notices a peak from about 2000 onwards.

"fake news" and disinformation in a digital context comes increasing concern around the role of online intermediaries and the effects of information sharing and spreading in the networked information age.

In this paper, we present a roadmap to diagnosing and addressing the information ecosystem's present ills, through a comparison of the legal and regulatory landscape in the United States and Europe. We articulate how various layers of legal and regulatory complexity constrain the universe of possible solutions to disinformation in those two regions, and why a universal solution might be difficult to devise at present. Our core aim is to help lawyers and policy-makers refine the sets of questions they must ask before proposing regional regulatory solutions.

We proceed in two parts. In Part I we pose a definitional problem: what are "fake news" and "disinformation", what is the main issue that needs addressing and why should free speech scholars care? In a recent Public Data Lab report, "fake news" is defined as content that is false and widely shared:

"*If a blog claims that Pope Francis endorses Donald Trump, it's just a lie. If the story is picked up by dozens of other blogs, retransmitted by hundreds of websites, cross-posted over thousands of social media accounts and read by hundreds of thousands, then it becomes fake news.*"[4]

While acknowledging the expression's ambiguity and controversial nature, we identify three factors that in combination provide conceptual clarity on the identification of disinformation: factual accuracy, the actor's intent, and the resulting harm. Based on these factors, we develop an analytical framework for identifying where regulatory and legal intervention are necessary.

The second question, which we tackle in Parts II and following, we explore various questions that must be answered in order to determine how "fake news" and "disinformation" should be regulated. We start in Part II with current constitutional and transnational free speech doctrines in the United States and Europe, presenting various critiques of those doctrines. In Part III, we then turn to an analysis of intermediary obligations and safe harbors and their complex relationship and tension with free

[4] *Public Data Lab, A Field Guide to "Fake News" and Other Information Disorders,* at 62, http://fakenews.publicdatalab.org/ (last visited Jun 16, 2018).

speech guarantees. Finally, in Part IV we assess current regulatory possibilities and reforms in the 'fake news' and 'disinformation' space. Our conclusion is that for any reform to make sense, lawyers and policy-makers must carry out an in-depth review of the regional legal and regulatory landscape, as well as of the technical possibilities and constraints that the networked information ecosystem presents.

2. Mapping Disinformation: Definitions and Problems

2.1. Fake News and Information Operations in Context

Misleading and sensational news are not an isolated phenomenon, they are characteristic of media strategies used to capture attention in an ecosystem characterized by attention scarcity. To understand the phenomenon, one must understand how content is generated, shared and further re-circulated.

According to the European Commission's High Level Group on Fake News and Online Disinformation (HLEG), problems of disinformation are driven on the one hand by actors and on the other hand by manipulative uses of communication infrastructures, uses "*that have been harnessed to produce, circulate and amplify disinformation on a larger scale than previously, often in new ways that are still poorly mapped and understood*".[5] In Data & Society's report on *Media Manipulation and Disinformation Online*,[6] Alice Marwick and Rebecca Lewis offer an in-depth overview of media manipulation in context, with a focus on right wing misinformation efforts. Their story is one of a very complex interaction and collusion between hyper partisan right wing actors and trolls on the one hand, and the mainstream media on the other hand, highlighting the media's tendency to gravitate toward sensationalism, the need for constant novelty, and the aim of achieving profits instead of professional ethical standards and civic responsibility.[7]

---

[5] European Commission, *A multi-dimentional approach to disinformation, 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, at 5.

[6] A. Marwick & R. Lewis, *Media Manipulation and Disinformation Online*, 106 (2017).

[7] *Id. supra*, at 47.

Platforms such as Facebook have acknowledged their role in the phenomenon and made significant efforts. Of course, as we discuss further below, appraising the success of their efforts depends on how we as a society and they themselves define the scope of their responsibility. In a white paper setting out their efforts, they lay out a taxonomy of information operations as a set of activities aimed at spreading inaccurate information and at shaping beliefs, emphasizing the role of different actors in the ecosystem and outlining how they propose to tackle the problem.[8] They define the umbrella category of "*Information (or Influence) Operations*" as the actions taken by governments or organized non-state actors to distort domestic or foreign political sentiment, most frequently to achieve strategic and/or geopolitical outcomes. "*False News*" are defined as news articles that purport to be factual, but which contain *intentional* misstatements of fact. "*False Amplifiers*" are ideologically-motivated coordinated activities by inauthentic accounts that are carried out with the intent to manipulate public opinion (e.g., by discouraging some speakers and encouraging or amplifying other speakers). The networks of accounts involved can be large networks of fake accounts used by dedicated professionals to share high volumes of information, or smaller networks of carefully curated online personas.[9] The goals of the creators and promoters of false amplifiers include the promotion or denigration of a specific cause or issue, the fostering of distrust in political institutions, or the general spread of confusion.[10] Financial gain is rarely their ultimate goal. "*Disinformation*" is a broader category which applies whenever

[8] Facebook, *Facebook and Information Operations* (April 27, 2017), https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf. But see an update here: https://newsroom.fb.com/news/2017/09/information-operations-update/.

[9] Alice Marwick & Rebecca Lewis, *Media Manipulation and Disinformation Online, Data & Society* (May 15th, 2017), https://datasociety.net/output/media-manipulation-and-disinfo-online/, at 8.

[10] Facebook, *Facebook and Information Operations* (April 27, 2017), https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf. But see an update here: https://newsroom.fb.com/news/2017/09/information-operations-update/.

inaccurate or manipulated content is spread intentionally.[11] Facebook see their responsibility as that of tackling devious speech whether it is somewhat true such as cherry-picked statistics, or outright falsehoods such as those that led to the *Pizzagate* scandal.[12]

The contributions and role of platforms, political and other actors to the spread of disinformation is controversial. Yochai Benkler, Robert Faris and Hal Roberts for instance have recently emphasized the important role of the political context, in particular the specific contributions of the right-wing media ecosystem to problems of disinformation in the United States, and have argued that the role of technology platforms, bots and foreign spies has tended to be overemphasized.[13]

2.2. Three Parameters to Understand Fake News and Disinformation

How to define "fake news" and "disinformation"? Is there a test that can guide lawyers, academics, policy-makers and platforms to consistently determine whether or not certain content deserves to remain visible and/or deserves constitutional protection?

According to the HLEG, "disinformation" is a more adequate term than "fake news" for at least two reasons: (1) the problem is not limited to news specifically but covers the spread of false or misleading information more generally including through fake accounts, videos and other fabricated media, through advertising and other organized information operations; (2) the term "fake news" has been adopted by politicians to contest information that is against their interests.[14] HLEG defines

---

[11] Alice Marwick & Rebecca Lewis, *Media Manipulation and Disinformation Online*, *Data & Society* (May 15th, 2017), https://datasociety.net/output/media-manipulation-and-disinfo-online/, at 5.

[12] See "Facing Facts" (May 23, 2018), available at: https://newsroom.fb.com/news/2018/05/facing-facts-facebooks-fight-against-misinformation/.

[13] Y. Benkler, R. Faris, H. Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (2018).

[14] European Commission, *A multi-dimentional approach to disinformation, 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, at 10.

disinformation as "*false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit. The risk of harm includes threats to democratic political processes and values, which can specifically target a variety of sectors, such as health, science, education, finance and more.*"[15] HLEG distinguishes the notion of disinformation from that of "misinformation", i.e. "*misleading or inaccurate information shared by people who do not recognize it as such*,"[16] and excludes from the notion of disinformation all questions related to illegal forms of speech such as defamation, hate speech, incitement to violence, etc., and also issues related to the spread of parody and satire. While we agree that the notion of "disinformation" is a more accurate term than "fake news," we will be using both terms in what follows.

A possible way of conceptualizing some important distinctions between different types of manipulative and problematic information, is to take an analytical approach to fake news and disinformation, focusing on three parameters: the content's factual truth, the intent and strategic goals associated with the content's generation and initial sharing, the harm caused by the content's release into the public sphere. While each of these factors is difficult to ascertain in practice and heavy reliance on any one of them can be somewhat unreliable, a broad taxonomy can be developed through these three factors. The presence of some harm coupled with some level of factual inaccuracy presents regulatory issues, which can often be satisfactorily addressed through existing laws (eg. defamation laws). Instead, devising *ad hoc* legal and regulatory remedies is urgently needed where factual inaccuracy and harm are coupled with the existence, on the part of one or more actors, of a diffuse intentionality to manipulate, fabricate and propagate false or deceitful information. This includes the existence of significant levels of fault on the part of employers and intermediaries.

In an attempt to address the difficulties attached to understanding the various shades of intentionality that are at play, Claire Wardle of First Draft News identifies seven types of

[15] *Id. supra.*
[16] *Id. supra.*

misleading speech in a media context and places them on a spectrum loosely based on the intent to deceive:[17]

a. ***Satire or parody****, ie. news where there is no intention to cause harm but a potential to fool (eg. The Onion);*

b. ***False connection****, ie. news whose headlines, visuals or captions do not support the content, which is factually accurate. Often these include news whose headlines and imagery were crafted to attract large audiences;*

c. ***Misleading news****, ie. these are news that include misleading use of words or information to frame an issue or an individual and slightly change the meaning of the message;*

d. ***False content****, ie. where genuine content is shared with false contextual information and that changes the understanding and interpretation of the information provided;*

e. ***Imposter content****; ie. where genuine sources are impersonated with false, made-up sources. Here there is no longer an issue of presenting true information in ways that are more or less suited to the author's goals, but rather of constructing news that are deliberately false and giving them an appearance of truth;*

f. ***Manipulated content****, ie. where genuine information or imagery is manipulated to deceive;*[18]

g. ***Fabricated content****: content that is 100% false, deliberately designed to deceive and do harm.*

To guide our further discussion, we thus tentatively conclude that:

- innocent and inaccurate news that cause *de minimis* harm do not raise urgent issues from a legal and regulatory standpoint;
- (negligently) inaccurate news that cause harm require some legal and regulatory redress, where such redress is not already provided by existing laws (eg. defamation or journalistic codes of ethics), new obligations for media outlets and intermediaries must be envisaged to ensure that the spread of harmful information is reduced to a manageable level; and
- the intentional manipulation and fabrication of news and information, whether or not facilitated by intermediaries

[17] C. Wardle, *Fake news. It's complicated*, First Draft News (February 16, 2017), https://firstdraftnews.org/fake-news-complicated/ (Retrieved April 22, 2017).

[18] *See* eg.: https://blogwatch.tv/2017/10/fake-news-types/manipulated-content/

without the relevant intent, requires new legal and regulatory prohibitions in the digital media ecosystem.

Our approach can be summarized as follows:

Types of Information	Factual Inaccuracy	Intent to Deceive	Harm
Innocent	No	No	No
Inaccurate (non-Harmful)	Yes	No	No
Harmful Inaccuracies	Yes	No	Possible
Manipulated	Maybe	Yes	Yes
Fabricated	Yes	Yes	Yes

3. Free Speech and "Fake News"

In *On Liberty,*[19] John Stuart Mill makes the case for a generous understanding of freedom of speech arguing that truth as an ideal can only be achieved if both true and false statements and opinions are allowed to remain uncensored. In the United States, false statements in newspapers have been held to have no constitutional or moral value in themselves.[20] Instead, in the US and in Europe false statements are often given legal protection so as to protect other values such as diversity of opinion, a wider scope of debate, freedom of the press, etc.[21] In other words, it seems that false news have instrumental rather than intrinsic value, and they are valuable and worthy of constitutional protection only insofar as they enable the promotion of other values such as plurality or democracy. Yet it seems that in the current context, overgenerous constitutional protection for speech may be one of the factors leading to a loss of epistemic trust in democracy. Thus, some of the guarantees that are in place to allow for speech to flourish might need to be re-examined. It is arguable that most forms of false news and disinformation, if they are constitutionally protected, fall within this second broad category

[19] J. S. Mill, *On Liberty* (1859).
[20] *See* below *New York Times v. Sullivan*, 376 U.S. 254 (1964).
[21] *See* below *Gertz v. Robert Welch, Inc.*, 418 U.S. 323 (1974).

of speech. Our view is that such protected speech not only should be given less value, but that the existing normative claims against its protection should be given more weight in deciding whether to deny it protection.

Our aim in this section is to formulate this argument while providing an overview and comparison of United States and European free speech standards. In Parts III and IV, we then move to exploring the kinds of remedies and safeguards that ought to be put in place to reduce and eradicate disinformation efforts that have no constitutional or human rights value. The latter depends in large part on the role and responsibility of intermediaries and media organizations in preventing the spread of misleading and false information.

3.1. Constitutional standards of protection for false news in the United States under the First Amendment

The First Amendment, adopted on December 15 1791 as part of the US Bill of Rights, is worded as an absolute, with no carve outs specified:

"*Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.*"

The history of the First Amendment is a long and relatively silent one until the 1920ies when US courts began to interpret the provision as conferring onto US citizens a wide scope of protection for free expression. While First Amendment jurisprudence has developed as a complex patch of sometimes inconsistent doctrines, what an external observer sees as the high watermark of American free speech is its very wide scope of constitutional protection.

The most famous rationale for such wide scope of protection is the notion of the "*marketplace of ideas*", a notion imported into First Amendment jurisprudence in the 1920ies, as illustrated by Justice Holmes' famous dissent in *Abrams v United States*:[22]

"*[T]he ultimate good desired is better reached by free trade in ideas – that* ***the best test of truth is the power of the thought to get***

[22] *Abrams v. United States*, 250 U.S. 616 (1919).

itself accepted in the competition of the market*, and that truth is the only ground upon which their wishes safely can be carried out… I think that we should be eternally vigilant against attempts to check the expression of opinions that we loathe and believe to be fraught with death, unless they so imminently threatened immediate interference with the lawful and pressing purposes of the law that an immediate check is required to save the country.*"[23] (emphasis added)

In *Reno v. ACLU*,[24] the Supreme Court applied the 'marketplace of ideas' analogy to the internet. Applying Holmes' metaphor, the court held that certain limitations on obscene online speech under the Communications Decency Act had the effect of curtailing speech that was not obscene, and therefore that the relevant portions of the Act were unconstitutional. This case highlighted for the first time the extent to which attempts to zone online speech can have spillover and chilling effects. It has also left a void in the regulation of online speech, which appears to be widening with time.

From a more theoretical point of view, as it has already been stated,[25] the reference to the "*marketplace of ideas*," should be handled with care. A metaphor implies knowledge transfer across domains (from the Greek *meta pherein*, to "carry over"): the free market of ideas metaphor carries over from the source domain of economic activity to the target domain of speech a systematic set of entailments that supersedes the limitations of the older free speech model. The economic context of Holmes' use of the metaphor should not be overlooked. In 1919, *laisser faire* market capitalism was triumphant. The concept of a free market provided a meaningful model for truth: like for any efficient allocation of goods and services which leads to market equilibrium, truth would arise as the end result of a free exchange of true and false ideas. Similarly, when the US Supreme Court borrowed the metaphor in 1997 calling the internet the "*new market place of ideas*," it was a relatively unregulated and open environment not yet characterized by the prevalence of monopolies and oligopolies. In this context, the metaphor of a free marketplace of ideas may have made perfect sense. By contrast, today, the same metaphor

[23] *Id. supra,* at 630.

[24] *Reno v. ACLU*, 521 U.S. 844 (1997).

[25] See O. Pollicino, Editorial, *Fake News, Internet and Metaphors (to be handled carefully)*, in this Journal, 2017 9(1), 1-5

appears inadequate. Digital markets, in particular, are today far from "perfectly" competitive, and instead appear to be concentrated and skewed in favor of a handful of businesses. Similarly, if fake news are arguably a significant and pervasive form of failure in the marketplace of ideas, this makes it reasonable to advocate for a degree of regulatory intervention in this area.

As regards freedom of the press, the First Amendment generally affords a wide scope of speech protection to the press and other speakers, shielding them from liability for the spread of false news and disinformation. One exception to this large protective shield is the *New York Times v. Sullivan*[26] line of cases which established that those who spread false information about public officials do not benefit from constitutional protection if they act maliciously. The rule is narrower than it seems in that it only applies to false defamatory speech against public officials, and only if actual malice is shown. Actual malice under the First Amendment means knowledge or reckless disregard of falsity, on the part of the publisher of false information. In that case Justice Brennansaid, relying on John Stuart Mill, that "*even a false statement may be deemed to make a valuable contribution to a public debate.*"[27] In *Gertz v. Robert Welch, Inc.*,[28] Justice Powell similarly stated that "*[t]he First Amendment requires that we protect some falsehood in order to protect speech that matters.*"[29] In attempting to strike a fair balance between the victims of false defamatory speech and freedom of the press, the Supreme Court in both cases acknowledged that a degree of inaccurate or misleading news is instrumentally necessary to achieve a healthy information ecosystem, and that therefore the protection of false news should be widely construed.

Outside the media context, in *United States v. Alvarez,*[30] the Supreme Court held that a law criminalizing false statements about having a military decoration or medal, the Stolen Valor Act, violated the First Amendment: thus that false statements deserve some constitutional protection. Mr. Alvarez had been convicted

[26] *New York Times v. Sullivan*, 376 U.S. 254 (1964).
[27] *Id. supra,* at 279 footnote 19.
[28] *Gertz v. Robert Welch, Inc.,* 418 U.S. 323 (1974).
[29] *Id. supra,* at 341.
[30] *United States v. Alvarez*, 567 U.S. 709 (2012).

under the Act for falsely stating in court that he was a retired marine who had been awarded a medal of honor. A majority of the Court agreed that the Stolen Valor Act was unconstitutional, albeit for different reasons.

Overall, in our view the moment has come for First Amendment law to be construed, if not more narrowly, at least in accordance with substantive notions of equality and pluralism, and compatibly with the necessary protection against private and commercial speech by powerful online actors. The state action doctrine negates the applicability of First Amendment protection between two privates, whether they are individuals, platforms or media organizations. Such doctrine proves somewhat problematic when it comes to speech harms that occur in a highly privatized digital public sphere. If private, public or a hybrid mix of actors are causing the spread of disinformation then it is wrong to keep them shielded from liability based on claims that First Amendment law requires an unregulated marketplace of ideas.[31]

3.2. European human rights standards of protection for false news

In the last sixty years or so, speech in Europe has evolved very differently from the United States. While in the United States, courts were willing to water down any restrictions on the ability of the media to publish inaccurate information, in Europe courts have followed a more cautious approach which focuses on the value of human dignity and pluralism. As a result, online speech and fake news regulation might predictably encounter more resistance in the United States than in Europe.

A "free marketplace of ideas" model of protection applies with difficulty to the European context, which becomes clear when one compares the wording of the First Amendment of the US Constitution with Article 10 of the European Convention on Human Rights. Article 10 of the ECHR has nevertheless played an important role, as EU Member States (and the EU itself, in the

[31] *See* M. D. Conover, J. Ratkiewicz, M. Francisco, B. Gonc¸alves, A. Flammini, F. Menczer, *Political Polarization on Twitter*, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (2011) https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2847/3275. Also *see* C. R Sunstein, #*Republic: Divided Democracy in the Age of Social Media* (2017).

future) join the Council of Europe and are bound by the Convention. The existence of such a binding 'constitutional' parameter is reflected in a common standard of protection within national legal systems. Contrary to the wording of the First Amendment, the wording of Article 10 of the European Convention on Human Rights (ECHR) places emphasis on the limits to free speech and is very clear in rejecting the view of free speech as an absolute.[32] The structure of this provision is twofold: Article 10(1) attaches to freedom of expression the value of a human right, while Article 10(2) admits for interferences with free speech that are necessary, in a democratic society, to meet certain social pressing needs. In other words, contracting states can then legitimately impose restrictions on freedom of expression, provided that the criteria set forth under art. 10(2) are respected.[33] The Article reads:

"*(1) 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.*

(2) The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary."

The difference is not only one of scope, but also of focus. Whilst the First Amendment addresses mainly the active

[32] See, for a comment on art. 10 ECHR, J. F. Flauss, *The European Court of Human Rights and the freedom of expression*, 84 Indiana Law Journal 809 (2009).

[33] As noted by Voorhoof, only a limited number of cases the Court came to the conclusion that the condition 'prescribed by law' was not fulfilled. According to Voorhoof, this condition requires foreseeability, precision, publicity or accessibility of the restriction. See D. Voorhoof, *Freedom of Expression under the European Human Rights System. From Sunday Times (n° 1) v. U.K. (1979) to Hachette Filipacchi Associés ("Ici Paris") v. France (2009),* 1–2 *Inter-American and European Human Rights Journal/Revista Interamericana y Europa de Derechos Humanos, Intersentia,* 3–49 (2009).

dimension of speech, i.e. the speaker's right to impart information freely, Article 10 of the European Convention and Article 11 of the EU Charter of Fundamental Rights emphasize the passive dimension of speech, i.e. the audience's right to be pluralistically informed and receive information. In this respect, it could be argued that false news and misleading or deceitful information do not fall within the scope of European free speech protection. European courts for instance would not adopt the US Supreme Court's approach in *Gertz* according to which "*[u]nder the First Amendment there is no such thing as a false idea. However pernicious an opinion may seem, we depend for its correction… on the competition of other ideas*."[34]

As noted by commentators, Article 10 ECHR protects 'several freedoms of speech', not just one.[35] A large number of decisions have been taken on complaints based on Art. 10. In this regard, the Strasbourg Court has introduced a distinction between political and commercial speech. Whereas in the first case contracting states enjoy a wider margin of appreciation, restrictions are less tolerable, in the Court's view, when rights to political speech are at stake.

In the leading case, *Sunday Times*,[36] the Court for the first time found that Art. 10 had been violated. The review concerned a judicial order preventing the publication by a newspaper of an article concerning a drug. The Court held that the restriction was not '*necessary in a democratic society*'. In this way, it becomes clear how review under art. 10 works as an additional layer of control over the protection of freedom of expression in Europe.

In an attempt to provide an overview of the same issues addressed by US Courts in the context of a free speech scrutiny, some remarks can be made.

With respect to hate speech,[37] it should be noted that the Court in Strasbourg often refers to art. 17 of the Convention, which encapsulates the 'abuse clause', preventing the exercise of

[34] *Gertz*, 418 U.S. 323 (1974), at 339-340.

[35] See J.F. Flauss (2009), cit. at. 33, 810.

[36] *The Sunday Times v. United Kingdom* App no 6538/74 (ECHR, 26 April 1979).

[37] Further to the contribution of Gillespie, Chapter 20, see also F. Tulkens, *Freedom of expression and hate speech in the case-law of the European Court of Human Rights*, in J. Casadevall, E. Myjer, M. O'Boyle (eds.), *Freedom of Expression. Essays in Honour of Nicolas Bratza*, Oisterwijk: Wolf Legal Publishers (2012).

the fundamental rights protected under the ECHR in a way that is likely to undermine the enjoyment of other freedoms established therein.[38] In other terms, the Court has pointed out that disputing the existence of 'clearly established historical facts'[39] amounts to an abuse of freedom of expression that Contracting States may legitimately restrict – upon certain conditions – when it is necessary to preserve other fundamental values underlying the Convention.[40]

This has allowed some Contracting States, including Austria, France and Germany, to enact laws against hate speech.[41]

The case of *Jersild*[42] is of particular importance, amongst others. The applicant was a journalist sentenced in Denmark for having conducted an interview with some members of a young racist organisation where offensive and insulting expressions were used. Although Mr. Jersild had clearly dissociated himself from these statements and rebutted part of them, he was convicted for aiding and abetting the youths interviewed. The ECtHR said that, under the circumstances of the case, such interference with the enjoyment of freedom of expression was not *necessary in a democratic society* and that, in particular, the means employed were disproportionate to the aim of protecting the reputation or rights of others.

A last case is worth mentioning *Handyside,*[43] a case where the Court rejected the complaint of an editor who had been convicted for having published a schoolbook containing sexually explicit contents. The Court found that the restrictions to freedom of expression imposed in the case, including the seizure of copies of the book, met the criteria set forth under art. 10(2) of the Convention. *Obiter dictum*, the Court stressed that:

---

[38] For a more specific analysis of the abuse clause, see H. Cannie and D. Voorhoof, *The abuse clause and freedom of expression in the European Human Rights Convention*: An added value for democracy and human rights protection? 29(1) *Netherlands Quarterly of Human Rights* 54 (2011).

[39] *Lehideux and Isorni v. France* App no 24662/94 (ECHR, 23 September 1998).

[40] *Garaudy v. France* App no 65831/01 (ECHR, 7 July 2003). See also *Peta Deustschland v. Germany* App no 43481/09 (ECHR, 8 November 2012).

[41] See in this regard R. Kahn, *Why do Europeans ban hate speech? A debate between Karl Lowenstein and Robert Post*, 41 Hofstra Law Review 545 (2013).

[42] *Jersild v. Denmark* App no 15890/89 (ECHR, 23 September 1994).

[43] *Handyside v United Kingdom* App no 5493/72 (ECHR, 7 December 1976).

"Freedom of expression constitutes one of the essential foundations of such a society, one of the basic conditions for its progress and for the development of every man. Subject to paragraph 2 of Article 10 (art. 10-2), it is applicable not only to 'information' or 'ideas' that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population. Such are the demands of that pluralism, tolerance and broadmindedness without which there is no 'democratic society'."

3.3. Online speech

The complex question of how to balance harm with freedom of the press in defining the scope of protection for false news acquires further complexity in the context of the internet. However, if we analyse the most recent Court decisions taken, it appears that the limitations to speech in Article 10(2) have been widely construed in relation to the Internet.

The Court appears to have paid more attention to the cases where the Internet was likely to threatened the protection of fundamental rights – restrictions in those cases were then found justified – than to those cases in which the Internet appeared as a new opportunity for the enjoyment of speech.

Even though the Court has repeatedly held that the 'safe harbors' under Art. 10(2) must be construed strictly, the coming of the Internet has determined an increase of the consideration given to restrictions to free speech. In particular, in the European Court's view, the specific medium represented by the Internet has given rise to an amplification of the threats to fundamental rights compared to the past. This point emerged, for the first time, in *Editorial Board of Pravoye Delo and Shtekel v. Ukraina*,[44] concerning freedom of press:

"The risk of harm posed by content and communications on the Internet to the exercise and enjoyment of human rights and freedoms, particularly the right to respect for private life, is certainly higher than that posed by the press. Therefore, the policies governing reproduction of material from the printed media and the Internet may differ. The latter undeniably have to be adjusted according to technology's specific features

[44] *Editorial Board of Pravoye Delo and Shtekel v. Ukraina* App no 33014/05 (ECHR, 5 May 2011).

in order to secure the protection and promotion of the rights and freedoms concerned."

The assumption behind the reasoning of the ECtHR is that the internet raises new problems for the protection of fundamental rights and that the measures applying to traditional media do not effectively work in the digital environment. A new balance between freedom of expression and other human rights must be sought and, according to the court, such balance had to be resolved in favor of more restrictions on freedom of expression.

Compare the ECtHR's approach to that of the US Supreme Court which, in the aforementioned decision *Reno v. ACLU,* expressed a completely opposed view:[45]

"*The record demonstrates that the growth of the Internet has been and continues to be phenomenal. As a matter of constitutional tradition, in the absence of evidence to the contrary, we presume that governmental regulation of the content of speech is more likely to interfere with the free exchange of ideas than to encourage it. The interest in encouraging freedom of expression in a democratic society outweighs any theoretical but unproven benefit of censorship.*"

In *K.U. v. Finland,*[46] the ECtHR highlighted the non-absolute nature of the protection of certain fundamental rights on the internet. The case concerned the dissemination of the personal data of a minor by an anonymous individual who had posted an online advertisement where he claimed to be in search of a sexual relationship. When the applicant filed a complaint with the local court, there were no legal grounds in domestic law to force an ISP to disclose personal data in a criminal case. Then, the domestic legislation failed to strike a balance between the right to data protection and other interests. Although the complaint was not based on Art.10, the ECtHR expressed significant remarks as to the enjoyment of freedom of speech on the Internet:

"*Although freedom of expression and confidentiality of communications are primary considerations and users of telecommunications and Internet services must have a guarantee that their own privacy and freedom of expression will be respected, such guarantee cannot be absolute and must yield on occasion to other legitimate imperatives, such as the prevention of disorder or crime or the protection of the rights and freedoms of others … [I]t is nonetheless the*

[45] 521 US 844, 885.

[46] *K.U. v. Finland* App no 2872/02 (ECHR, 2 December 2008).

task of the legislator to provide the framework for reconciling the various claims which compete for protection in this context."

The Court adopted a stricter approach where the limitations imposed on freedom of expression were disproportionate to their aim. For instance, in *Ahmet Yildirim v. Turkey*,[47] Strasbourg judges concluded that Turkey had violated Art. 10 of the Convention by imposing a disproportionate restriction on internet access. In the context of criminal proceedings against the owner of a website where expressions insulting Ataturk's memory had been posted, an administrative authority ordered the blocking of Google Sites as a whole to prevent access to the site in question, without ascertaining whether a less far-reaching measure could have been taken. The applicant, who owned a website where his academic works were published and which was affected by the extension of the blocking order, alleged a violation of his right to freedom of expression. The Court noted that the blocking of a website falls within the legitimate restrictions that Contracting States may adopt in accordance with Art. 10(2) of the Convention, but only upon the condition that such a restriction meets the requirements referenced therein. In the case at stake, both a strict legal framework defining the scope of the ban and a judicial review were lacking.

The approach of the ECtHR has proved to be very cautious. On the one hand, it concluded that a violation of Art. 10 had occurred when the restrictions to freedom of expression did not fulfill the conditions set forth under Art. 10(2). On the other hand, however, the Court conceded that free speech is not an absolute, nor a right to which a greater protection is attached compared to other fundamental rights: then, given the risks brought by the Internet, the right to freedom of expression can more likely be limited than in the non-digital context.

The same view is behind the decision in the *Pirate Bay*[48] case, where – on the contrary – the ECtHR rejected the application based on Art. 10. The applicants were the owners of a well-known online platform where users were provided with links to

[47] *Ahmet Yildirim v. Turkey* App no 3111/10 (ECHR, 18 December 2012).

[48] *Fredrik Neij and Peter Sunde Kolmisoppi (The Pirate Bay) v. Sweden* App n 40397/12 (ECHR, 13 March 2013).

download illegally copyrighted materials through the use of peer-to-peer systems. They had been sentenced under a Swedish law which criminalised copyright infringements but complained that their right to freedom of expression had been violated in this way. The Court declared the complaint inadmissible, as the restriction imposed on free speech met the conditions under Art. 10(2) and, in particular, was proportionate to the legitimate aim pursued.[49]

Thus, the view taken by the Court in Strasbourg is that new technologies, and the internet in particular, did not generally expand the scope of freedom of expression; rather, it created more opportunities for this right to enter into conflict with other interests protected under national constitutions.

This suggestion is confirmed by looking at how the ECtHR has reacted to freedom of the press in relation to the internet. Freedom of the press is regarded as an essential part of the freedom of speech and a pillar of democracy. In the case of *Stoll*,[50] the assumption behind the reasoning of the ECtHR is that by virtue of the new technologies the duties of journalists have become more onerous:

"*[T]he safeguard afforded by Article 10 to journalists in relation to reporting on issues of general interest is subject to the proviso that they are acting in good faith and on an accurate factual basis and provide 'reliable and precise' information in accordance with the ethics of journalism ... These considerations play a particularly important role nowadays, given the influence wielded by the media in contemporary society: not only do they inform, they can also suggest by the way in which they present the information how it is to be assessed. In a world in which the individual is confronted with vast quantities of information circulated via traditional and electronic media and involving an ever-growing number of players, monitoring compliance with journalistic ethics takes on added importance.*"

These observations are confirmed by similar cases in the offline environment. For instance in the case of *Yildirim v. Turkey*, where a single publication was found to be defamatory, then there were legal grounds to prevent its circulation, but not other publications. The problem of proportionality is crucially connected to the nature of the technical means that are employed

[49] See also *Ashby Donald and other v. France* App n 36769/08 (ECHR, 10 January 2013).

[50] *Stoll v. Switzerland* App no 69698/01 (ECHR, 10 December 2007).

in distributing the content: together, these factors are critical to the protection of freedom of expression on the internet.

Further, two cases demonstrate the Court's approach to intermediary liability: *Delfi v Estonia*[51] and *MTE v Hungary*.[52] In *Delfi* it was held to be lawful for Estonian courts to order a news portal to pay damages for defamatory comments posted on the site, even though these had been removed without delay after the news portal had been notified. The decision has been criticized as inequitable for failing to state an actual knowledge standard for platforms.[53] In *MTE v Hungary* the Court partly reviewed the *Delfi* ruling.[54] In this case the ECtHR held that punishing a non-profit self-regulatory body of Internet content providers (MTE) and an Internet news portal (Index) for offensive comments posted on their sites violated Article 10. The case was distinguished from *Delfi* because the comments here did not involve hate speech and also because the Hungarian courts in *MTE* had not carried out a thorough balancing exercise between the applicants' right to freedom of expression and the real estate websites' right to respect for their commercial reputation. The ECtHR has not clarified a clear position of principle on the extent to which platforms should be made responsible for policing online speech. Arguably, the ECtHR wants to avoid a conflict with European Union courts, who are also competent on such matters.

4. The Role of Online Intermediaries

Recent events such as the Cambridge Analytica scandal or the allegations of Russian interference into the 2016 US elections campaign have exposed how important online intermediaries and platforms are to the spread of disinformation. Jack Balkin has emphasized the important role that infrastructure plays in the protection of free expression.[55] In his view the protection of speech

[51] *Delfi AS v. Estonia* (2015) ECtHR 64669/09.
[52] *MTE v. Hungary* (2016) ECtHR 22947/13
[53] *See* R. Caddell, *The last post? Third party internet liability and the Grand Chamber of the European Court of Human Rights: Delfi AS v Estonia revisited*, 21 Commun Law 49–52 (2016).
[54] *Id. supra.*
[55] J. M. Balkin, *Freedom of the Press: Old-School/New-School Speech Regulation*, 127 Harv. L. Rev. 2296 (2014).

requires certain infrastructural affordances, and the hallmark of speech in the digital age, which is calls new-school speech regulation, is a revolution in such affordances. More recently, he has described speech in a digital setting as having a triangular structure which takes into account the importance of such intermediaries.[56] Rather than exclusively focusing on the traditional bilateral relationship between speakers and the state, regulators must adopt a broader view of speech as flowing between three groups of players: on the one hand the state, municipalities, transnational and supranational entities; on the other hand internet infrastructure, including social media companies, search engines, ISPs, payment systems and others; and finally speakers including mass media organizations, civil society, etc. According to Balkin, applying the laws and constitutional standards that have been developed for speech in the offline context to the internet would be fruitless.[57] In order to protect speech values in an internet context we must thus be ready to adopt technical, regulatory and legal solutions that go beyond the traditional forms of free speech regulation and that entail interventions at the platform level. While constitutional doctrines will keep guiding us on the proper scope of speech protection, we must look elsewhere for immediate solutions to the most harmful and widespread forms of false news and disinformation.

4.1. Immunity of Internet Intermediaries under US Law

In the United States, internet intermediaries have a large amount of freedom when it comes to speech violations. Section 230(c)(1) of the Communications Decency Act ("CDA") [58] confers full immunity on "*interactive computer service(s)*", for any content shared using their services, directing that "*no provider or user of an interactive computer service shall be treated as the publisher or speaker of*

[56] J. M. Balkin, *Free Speech is a Triangle*, 118 Columbia L. Rev., no. 7 (2018).
[57] *Id. supra*, at 20-22 and 29. Also *see* J. M. Balkin, *Free Speech and Press in the Digital Age: the Future of Free Expression in a Digital Age*, 36 Pepperdine Law Review 427–1161 (2009); J. M. Balkin, *Digital Speech And Democratic Culture: A Theory Of Freedom Of Expression For The Information Society*, 79 New York University Law Review (2004).
[58] 47 U.S.C. §230.

any information provided by another information content provider."[59] In other words, "interactive computer services" must not be treated as publishers responsible for the content being published, but as passive intermediaries that channel communications and have no responsibility for the content being communicated. Section 230(c) immunity is particularly strong in that it applies even if the intermediary knows of the defamatory content on their service and knowingly fails to remove it.[60] In the CDA, the notion of "*interactive computer service*" is defined as:

"*any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions.*"[61]

The Act is unclear on the distinction between such a service and the notion of "*information content provider*" defined instead as:

"*any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service.*"[62]

The relevant factor that differentiates "*interactive computer service*" from "*information content provider*" is the notion of "*creation or development of information*." What is questionable is that we could easily understand a platform to engage in acts that closely resemble creating or developing information, for instance by arranging the way in which the information is displayed or presented to the end-user. In a long list of cases, US courts have nonetheless protected intermediaries from liability by interpreting the notion of "*interactive computer service*" very broadly.[63] Lawsuits seeking to hold intermediaries liable for editorial functions such as

[59] Note that at §230(f)(3) information content provider is defined as follows: "*any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service.*"

[60] M. Lavi, *Taking Out of Context*, 31 Harv. J. Law Technol. 145, 164 (2017).

[61] *Id. supra* section (f)(2).

[62] Communications Decency Act § 230(f)(3).

[63] *See* Zeran v. AOL, 129 F.3d 327 (4th Cir. 1997); Nemet Chevrolet, Ltd. v. Consumeraffairs.com, Inc., 591 F.3d 250, 254 n.4 (4th Cir. 2009); Giordano v. Romeo 2011 WL 6782933 (Fla. App. Ct. Dec. 28, 2011); Global Royalties v. Xcentric 544 F supp 2d 929 DIST(2008); Caraccioli v. Facebook, Inc., 2016 WL 859863 (N.D. Cal. March 7, 2016).

selection of content or decisions on how and when to display it, have generally been blocked by this section.[64]

The section 230(c) CDA immunity is thus understood to apply without exceptions to content sharing platforms such as Facebook, YouTube, Twitter and others, alongside internet service providers, web hosting services and other internet intermediaries that enable people to express themselves online. Yet the growing awareness that these platforms clearly engage in curatorial activities akin to those of publishers is generating debates over the appropriateness of the immunity. The line between being a publisher responsible for published content and acting as a 'passive' platform intermediary is no longer fully tenable.[65] The line between pre- and post-publication has become blurred because content can exist online yet not be spread widely or be visible or findable at all.[66] Jonathan Zittrain has suggested distinguishing between large and smaller intermediaries, using a size or level of activity threshold for immunity.[67] Jack Balkin on the other hand has suggested a procedural threshold. [68] His view is that if a platform complies with the Manilla principles,[69] which require a level of procedural fairness and transparency on the platforms' part, then it should benefit from immunity. If it does

[64] *See* Zeran v. AOL; Gentry v. eBay, Inc., 99 Cal. App. 4th 816 (Cal. App. Ct. June 26, 2002); Levitt v. Yelp Inc, 2011 U.S. Dist. LEXIS 124082 (N.D Cal. Oct. 26, 2011); Levitt v. Yelp! Inc., 2014 WL 4290615 (9th Cir. Sept. 2, 2014).

[65] M. Lavi, *Taking Out of Context*, forthcoming: Harvard Journal of Law and Technology (2018), J. Zittrain, *CDA 230 Then and Now: Does Intermediary Immunity Keep the Rest of Us Healthy?* The Recorder (Nov. 10, 2017), available at: https://www.law.com/therecorder/sites/therecorder/2017/11/10/cda-230-then-and-now-does-intermediary-immunity-keep-the-rest-of-us-healthy/.

[66] J. Zittrain, *CDA 230 Then and Now: Does Intermediary Immunity Keep the Rest of Us Healthy?* The Recorder (November 10, 2017), https://www.law.com/therecorder/sites/therecorder/2017/11/10/cda-230-then-and-now-does-intermediary-immunity-keep-the-rest-of-us-healthy/?slreturn=20180922190041.

[67] *Id. supra.*

[68] J. M Balkin, *Free Speech is a Triangle*, Columbia L. Rev., cit. at. 56, 44 (2018).

[69] In J. Balkin's own words, "*[t]he Manilla Principles require, among other things (1) clear and public notice of the content regulation policies companies actually employ; (2) an explanation and an effective right to be heard before content is removed; and (3) when this is impractical, an obligation to provide to post-facto explanation and review of a decision to remove content as soon as practically possible.*" *Id. supra*, at 41.

not comply, a platform cannot benefit from full immunity and must instead comply with constitutional limitations on intermediary liability, which are at present uncertain. This view, which is reasonable, falls short of the European safe harbor standards that we discuss below.

Revising the current immunities under section 230(c) CDA would be an important move, which legislators must avoid making without caution. They must focus on the various policy options that internet infrastructure allows. For instance, requiring a correction, an apology or a tweak in the algorithm that governs the spread of the problematic content in question, if practicable, may be more effective than monetary compensation by the company that contributed to such spread. As explained by Lawrence Lessig in response to the regulatory void left by the Supreme Court's ruling in *Reno*, one cannot leave harmful and deceitful online speech unregulated. The best approach requires taking into account the specificities of internet architecture and the possibilities it offers.[70]

Other kinds of limits on speech are in fact being placed indirectly through the copyright infringement notice and takedown procedure in place under the Digital Millennium Copyright Act.[71] The *Google Transparency Report* for instance reveals that a number of requests submitted through their copyright removal system have little or no connection to copyright, eg. a driving school requesting the removal of a competitor's homepage from search based on a very weak copyright claim, or requests for removal of non-infringing material publicising past copyright removal requests.[72] Further, in the US Google's policy is to remove libellous material upon presentation of a court order demonstrating its defamatory nature. Using the Lumen database,[73] Professor Eugene Volokh has brought new light on the practice of individuals and companies to

[70] L. Lessig, P. Resnick, *Zoning Speech on the Internet: A Legal and Technical Model,* in The Berkman Center for Internet & Society Research Publication No. 1999-06 12/1999.

[71] 17 U.S.C. § 512.

[72] *Google Transparency Report*, http://www.google.com/transparencyreport/removals/copyright/faq/#abusive_copyright_requests (last visited August 30, 2018).

[73] *See* https://www.lumendatabase.org/pages/about.

abuse of such removal system.[74] He has shown that companies and individuals are increasingly creating or obtaining fraudulent court orders to censor online content.[75] This and similar practices illustrate the complexity regulators face in designing effective solutions to the policing of harmful online content, in spite of the cooperative efforts of online platforms and other intermediaries concerned about the quality of the end product they offer, and increasingly also worried about their reputation.

4.2. EU Safe Harbours and Content Removal Obligations

As discussed above, one reason for the ECtHR's reluctance to express firm views regarding online content may have been the parallel and evolving nature of the EU law of intermediary liability.[76] The EU has a rich panoply of existing and prospective legislation for the digital single market. The e-Commerce Directive[77] confers partial immunity from liability for speech violations on intermediaries that passively transmit, cache or host online content, the so-called "safe harbours."[78] The Directive's safe harbours apply to "*information society service providers*" that act as:

- *Mere conduits*, providing a service "*that consists of the transmission in a communication network of information provided by a recipient of the service, or the provision of access to a communication network,*"[79]

[74] A. Holland, *Lumen Research In the News - Texas AG sues CA company over falsified court orders business model,* Lumen Blog (2017), https://www.lumendatabase.org/blog_entries/801 (last visited Aug 30, 2018).

[75] Adam Holland, *More apparently fraudulent court orders lumen blog* (2018), https://www.lumendatabase.org/blog_entries/802 (last visited Aug 30, 2018). Also see: E. Volokh, *Libel takedown injunctions and fake notarizations*, Washington Post, March 30, 2017, https://www.washingtonpost.com/news/volokh-conspiracy/wp/2017/03/30/libel-takedown-injunctions-and-fake-notarizations/ (last visited Aug 30, 2018).

[76] Note that the ECHR and the EU are separate transnational governance frameworks that, to a large extent, operate independently of one another. The EU's accession to the ECHR is not yet finalized, see: http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI%282017%29607298

[77] Directive 2000/31/EC OJ L 178, 17/07/2000.

[78] *See* Articles 12-14 of the Directive.

[79] Article 12 of the E-Commerce Directive

- *Caching service providers*, enabling "*transmission in a communication network of information provided by a recipient of the service,*"[80]
- *Hosting service providers*, allowing for the "*storage of information provided by a recipient of the service.*"[81]

According to the Directive's recital (18) these include entities operating a marketplace such as eBay or Amazon[82] and media-sharing platforms such as YouTube and Facebook.[83]

Contrary to the US regime under section 230 CDA, in the EU intermediaries will benefit from immunity only as long as they remain passive and unaware facilitators. They become liable to remove the illegal content once they are aware of its presence on their service, for instance if they are notified by a user. In *Google France v Louis Vuitton*,[84] the EU Court of Justice stated that Article 14 of the Directive, which provides immunity to hosting providers, would only apply if the storage of content was "*of a mere technical, automatic and passive nature*"[85] and if the entity did not play "*an active role of such a kind as to give it knowledge of, or control over, the data stored.*"[86] In *L'Oréal v eBay*, neutrality was said to be unlikely where an intermediary optimizes or promotes users' stored content.[87] The question of whether the use of content sorting algorithms can fall within the Directive's definition of neutrality was interestingly explored by Advocate General Poiares Maduro in *Google France*. The Advocate General considered Google not to have a direct interest in the content being presented on its search platform, and concluded that its use of algorithms to sort search results was therefore neutral, save for the display of Google AdWords advertisements in which Google had a direct interest:

---

[80] Article 13 of the E-Commerce Directive
[81] Article 14 of the E-Commerce Directive
[82] *See* eg. C-324/09 L'Oréal SA v eBay International AG EU:C:2011 [2012].
[83] *See* eg. Case No. 11/2014 *Gestevision Telecinco SA v YouTube LLC* [2014] 2 CMLR 13 (Court of Appeal of Madrid, 14 January 2014).
[84] C-236/08, C-237/08 and C-238/08 *Google France Sarl v Louis Vuitton Malletier SA* [2010] ECR I-2417.
[85] *Id. supra*, at paras 113-114.
[86] *Id. supra*, at para 120.
[87] *See* C-324/09 *L'Oréal SA v eBay International AG* EU:C:2011 [2012], at paras 140-146.

"Google's search engine… is neutral as regards the information it carries. Its natural results are a product of automatic algorithms that apply objective criteria in order to generate sites likely to be of interest to the internet user. The presentation of those sites and the order in which they are ranked depends on their relevance to the keywords entered, and not on Google's interest in or relationship with any particular site."[88]

The Advocate General's view on the neutrality of Google's content sorting algorithms is no longer as plausible today, especially in light of some of the significant antitrust investigations that the European Commission has carried out into Google's search activities, finding Google liable for displaying results in a way that advantages its own services.[89] This and other recent developments render Google and other content sharing and sorting platforms' activities less likely to squarely fall within the e-Commerce Directive's immunities.

On the question of whether it is appropriate for platforms to use automated content filtering mechanisms to filter out harmful content, the Court of Justice of the EU has expressed itself more than once against automated filtering and in favour of broad free speech guarantees. In its 2012 judgment in S*abam* v *Netlog*, [90] the Court of Justice of the EU held that Belgian content filtering requirements imposed on an intermediary violated Article 15 of the e-Commerce Directive, which exempts hosting service providers from a general obligation to monitor the content transmitted through their platform. [91] The Court also held that, requiring an intermediary to install the contested filtering system did not strike a fair balance between the right to intellectual property (Article 17 of the EU Charter), on the one hand, and the freedom to conduct business (Article 16 of the EU Charter), the right to protection of personal data (Article 8 of the EU Charter) and the freedom to receive or impart information (Article 11 of the EU Charter), on the other. EU Member States are thus arguably not allowed to require the placing of content filtering mechanisms

[88] *Google France case*, Opinion of Mr Advocate General Poiares Maduro delivered on 22 September 2009, para 144.

[89] *See* this EU Commission Press Release dated 29 June 2017: http://europa.eu/rapid/press-release_IP-17-1784_en.htm.

[90] Case C‑360/10 *Sabam v Netlog*, 16 February 2012 ECLI:EU:C:2012:85

[91] The Court confirmed its *Scarlet Extended* ruling Case C‑70/10 Scarlet Extended [2011] ECR I-11959.

of their own initiatives. Our view is that in order to strike a fair balance between freedom of expression and other values on the internet, a proportionate and fact-specific approach is needed. Considering *Sabam v Netlog* as the EU's final word on the question of internet filtering would likely be a mistake.

One of the e-Commerce Directive's safe harbours' rationales was to "*secure the free flow of information*" and to maintain a "*free and open public domain*." [92] Yet another important objective of the Directive was to facilitate the functioning of the EU single market. Promoting the values of freedom of expression online does not go hand in hand with interpreting these safe harbours extensively. Indeed it seems that in some circumstances imposing obligations on intermediaries can lead to better functioning online markets, as in *L'Oréal v eBay*.[93]

The fact that internet intermediaries can be immune from content liability under EU law in fact does not mean that they have no obligations with regard to the content that they make available. A variety of intermediary obligations to monitor and police content are arising under EU law. The General Data Protection Regulation (GDPR),[94] alongside other instruments, strongly regulates intermediaries' obligations to police content. For instance, the GDPR imposes a number of monitoring and transparency obligations on intermediaries such as the responsibility to enforce EU citizens' right of erasure or "*right to be forgotten*" on the internet.[95]

In relation to harmful news, the conflict between speech and data protection typically arises where a newspaper publishes information that either harms or affects the reputation of an individual, and such individual requests the information's removal, first from the newspaper and subsequently from the online platform that hosts the content. In *Google Spain*,[96] the Court of Justice of the European Union recognized that search engines (Google) have an obligation to remove information searchable on

[92] J. Riordan, *The liability of internet intermediaries* (First edition. ed. 2016).
[93] C-324/09 *L'Oréal SA v eBay International AG* EU:C:2011 [2012].
[94] Regulation (EU) 2016/679 (General Data Protection Regulation) OJ L 119, 04.05.2016.
[95] *See* Article 17, the right to be forgotten, also see other obligations of intermediaries eg. under Articles 24, 25 and 33 in Chapter 4 of the Regulation.
[96] Case C-131/12, *Google Spain SL v AEPD and Mario Costeja González*, May 2014.

the web through an individual's name (Mr Gonzalez) upon the individual's request and provided it would be proportionate to do so. After establishing that the activities of a search engine constitute "processing" and that a search engine should be considered a "controller" under Directive 95/46, the antecedent of the current GDPR, the court went on to state that individual rights to privacy and data protection must be interpreted as requiring that upon request:

"*the operator of a search engine is obliged to remove from the list of results displayed following a search made on the basis of a person's name links to web pages, published by third parties and containing information relating to that person, also in a case where that name or information is not erased beforehand or simultaneously from those web pages, and even, as the case may be, when its publication in itself on those pages is lawful.*" [97]

Jonathan Zittrain has argued from a US perspective that the Court of Justice in this case went both too far and not far enough.[98] It went too far because it required a search engines to censor lawful speech that should not be delist-able under the First Amendment. It imposed a speech standard that was too stringent given the borderless nature of the internet and the different standards applicable in other countries. At the same time, the Court did not go far enough because it required Google to de-list the information, but did not require its complete removal from the internet. Information that remains on the internet can be further accessed and processed by third parties, arguably in violation of the individual's privacy rights. For Zittrain, the *Google Spain* ruling is disproportionate because it introduces dangerous internet censorship while insufficiently enforcing privacy rights. As we will see below, some of the legislative efforts that are being made to regulate fake news are objectionable on similar grounds.

No matter the jurisdiction and because of the borderless nature of the internet, the immunities and obligations of online intermediaries in relation to the content that they make available are extremely complex and any decisions as to how to strike a fair balance between immunity and other considerations must be very fact specific. This means that legislation that seeks to impose on

[97] *Id. supra*, at para 88.

[98] J. Zittrain, *Don't Force Google to 'Forget'*, New York Times (May 14, 2014), https://www.nytimes.com/2014/05/15/opinion/dont-force-google-to-forget

intermediaries an overarching and rigid obligation to police and remove harmful news content will hardly comply with constitutional and human rights free speech guarantees. On the other hand, immunities fall short of protecting the values underlying free expression when they are rigidly interpreted.

5. Tackling "Fake News": Intermediaries, Law and Technology

We explored the boundary between harmful speech that must be protected for its own sake, and speech that must be tolerated for the sake of other values, we have investigated the role of intermediaries and how they are currently regulated when it comes to online speech and news in particular. In light of jurisdictional differences in constitutional guarantees, platform immunities and monitoring obligations, regulating "fake news" is not only difficult, but may seem a flawed enterprise to begin with. In this section we try to debunk this cyberlibertarian misconception, outlining a series of possible approaches to the eradication of harmful forms of disinformation and false news.

5.1. From New Rights for Individuals to New Obligations for Platforms and News Organizations

Regulating news is difficult: the right of the press and of speakers to communicate their ideas freely requires a level of protection of harmful speech that, in jurisdictions such as the United States, makes disinformation hard to attack. One way of regulating harmful speech, which we alluded to above, is through the recognition and enforcement of competing individual or collective rights such as the rights of audiences and of the individuals affected by the harmful speech in question.

An example is the European "*right to be forgotten*" discussed above which resulted from the coming into force of data protection legislation and from the *Google Spain* judgment. As Jonathan Zittrain's view on *Google Spain* above illustrates, there is much disagreement on whether the "*right to be forgotten*" correctly balances third parties' rights to speak and access information online with individual privacy. Content can be removed from the internet even if it is completely lawful, provided there is no legitimate interest on the part of others in accessing it. This brings to light an aspect of speech that is reflected in commercial speech

discussions in the United States: in *Central Hudson*,[99] for example, the Supreme Court made clear that it was protecting the right of the public to receive information, not the autonomy of advertisers as speakers.[100] The right to speak and the right to receive or access information are two sides of a coin, and in some circumstances the right to speak freely may be trumped by competing considerations, provided the information that is being communicated is not information that others have a right to access. If we apply this consideration to the question of how to regulate disinformation: it seems that no one has a right to access disinformation per se. Therefore, any rights that compete with the speaker's right to intentionally utter false and harmful information will trump the speaker's rights unless excessive policing of that speech has a high likelihood of chilling other legitimate speech and turning into censorship. Thus, if in a hypothetical case, an audience or individual's right to prevent disinformation can be envisaged, then there would be no strong reasons for a court not to enforce that right over the speaker's right. In order to justify that their right should be protected, and not the audience's, a speaker would need to adduce strong empirical evidence showing that such an approach can chill other speech.

The broad transparency obligations and the data subject rights that the General Data Protection Regulation provides are also a fruitful ground for thinking about innovative solutions to the fake news problem. Indeed, being provided with a better understanding of how false news spreads, either through the operation of a "right of access"[101] or possibly even a "*right to algorithmic explanation*" under the GDPR is a first step toward being able to stop those news at source and also toward defending oneself from harmful information sources and bad actors. The European Data Protection Supervisor, in its *Opinion 3/2018 on*

[99] *Central Hudson Gas & Electric Corp. v Public Service Commission of NY*, 447 U.S. 557, 563 (1980).
[100] *See* J. M. Balkin, *Information Fiduciaries and the First Amendment*, 49 UC Davis Law Rev. 1183, 1213 (2016) and R. Post, A. Shanor, *Adam Smith's First Amendment*, 128 Harv. L. Rev. F. 165, 172 (2015).
[101] Articles 12 to 15 of the GDPR.

online manipulation and personal data,[102] emphasizes the need to combine formal rights with other mechanisms that make the effective exercise of those rights by individuals possible.[103]

The right to be forgotten, the right to an algorithmic explanation and the rights associated to the protection against disinformation can be a part of the answer, but they all appear to have an underlying collective dimension that pure rights language does not easily capture. In an attempt to clarify the relational nature of rights and obligations in a data intensive digital ecosystem, in the United States Jack Balkin introduced a new framework for thinking about online intermediaries and data processors as "*information fiduciaries*,"[104] also as a way of circumventing some of the First Amendment barriers to regulating these entities. He understands information fiduciaries as trustees that have a special relationship of trust and confidence with their customers under private law similar to that of financial advisors, doctors or lawyers, and that therefore have a fiduciary obligation to treat customer information with due care and in a way that is aligned with customers' interests:

"*Because of their special power over others and their special relationships to others, information fiduciaries have special duties to act in ways that do not harm the interests of the people whose information they collect, analyze, use, sell, and distribute. These duties place them in a different position from other businesses and people who obtain and use digital information. And because of their different position, the First Amendment permits somewhat greater regulation of information fiduciaries than it does for other people and entities.*"[105]

Whether or not fiduciary obligations as conceived by Jack Balkin are practicable, configuring new rights, imagining new fiduciary obligations, and ensuring that these rights and obligations are effectively enforceable and can be exercised could lead to a series of prospective improvements for the regulation of false news and disinformation: first, such solutions could

[102] EDPS Opinion 3/2018 on online manipulation and personal data (March 19, 2018), available at: https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf (last visited September 13, 2018).
[103] *Id. supra*, at 20-22.
[104] J. M. Balkin, *Information Fiduciaries and the First Amendment*, 49 UC Davis Law Rev. 1183 (2016).
[105] *Id. supra*, at 1186.

contribute to the weakening of business models that rely on opaque data harvesting practices and on the support of unreliable data brokers to spread their news and increase their profitability; second, platforms would have to put more resources into explaining how they select, display and suggest content to a user, leaving online users more empowered with regard to how their beliefs are shaped; third, keeping track of information and of the value flows related to such information (including advertising revenues) would become an easier task for individuals and regulators. There seems to be an urgent need to enable the exercise of new rights, and in parallel to look beyond traditional rights-based models for new fiduciary and professional standards of care for platforms and news outlets respectively.

5.2. Is Self-Regulation a Solution to Disinformation?

Platforms are using artificial intelligence to improve and streamline news operations.[106] Journalists and media organizations are increasingly tapping into AI's potential to sort content, produce news articles, police comments, and now even to recognize fake news.[107] Yet as some recent documented experiences show, AI is far from being a perfect tool and its use and impact must be assessed with care, looking at wide societal impact and beyond narrow sets of operational technicalities.[108]

In Europe, the General Data Protection Regulation imposes restrictions on the use of machine learning and automated algorithms to make decisions that affect individuals without any human intervention.[109] This raises some questions on the legality

[106] *See* eg. J. B. Merrill and Ariana Tobin, *Facebook's Screening for Political Ads Nabs News Sites Instead of Politicians*, ProPublica (June 15, 2018), https://www.propublica.org/article/facebook-new-screening-system-flags-the-wrong-ads-as-political (last visited September 13, 2018).

[107] *See* eg. C. Underwood, *Automated Journalism - AI Applications at New York Times, Reuters, and Other Media Giants*, TechEmergence (January 17, 2018), https://www.techemergence.com/automated-journalism-applications/ (last visited Jun 29, 2018).

[108] K. Crawford and R. Calo, *There is a blind spot in AI research*, Nature (October 13, 2016), https://www.nature.com/news/there-is-a-blind-spot-in-ai-research-1.20805 (last visited Jun 29, 2018).

[109] *See* Article 22 of the GDPR, also *see* F. Kaltheuner and E. Bietti, *Data Is Power: Towards Additional Guidance on Profiling and Automated Decision-Making in the GDPR*, 2 Journal of Information Rights, Policy and Practice (2018).

of automated mechanisms put in place by platforms such as Facebook to remove harmful speech, such as terrorist content or fake news. But the exact scope of the GDPR restrictions on automated decision-making and on the use of algorithms for the moment remains far from clear.[110]

In relation to misinformation, in April 2017 Facebook announced four actions that they were taking to tackle misinformation:

1. Collaborating with others to find industry solutions to this societal problem;
2. Disrupting economic incentives, to undermine operations that are financially motivated;
3. Building new products to curb the spread of false news and improve information diversity; and
4. Helping people make more informed decisions when they encounter false news.[111]

Following the Cambridge Analytica scandal in early April 2018 during which it was discovered that Facebook had been tacitly accepting the harvesting of user data by Cambridge Analytica, a political consulting firm with ties to the Trump campaign, [112] Mark Zuckerberg was invited to testify before the Senate and before the House of Representatives.[113] At these and subsequent hearings, Zuckerberg and other prominent Facebook employees have been grilled about Facebook's role in the spread of political disinformation and propaganda, with mixed results. [114] In May 2018, Facebook made other announcements regarding its

[110] *Id. supra.*

[111] https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf

[112] *See* N. Badshah, *Facebook to contact 87 million users affected by data breach*, The Guardian (April 8, 2018), https://www.theguardian.com/technology/2018/apr/08/facebook-to-contact-the-87-million-users-affected-by-data-breach.

[113] *See* coverage of the hearings, eg. Z. Wichter, *2 Days, 10 Hours, 600 Questions: What Happened When Mark Zuckerberg Went to Washington*, New York Times (April 12, 2018), https://www.nytimes.com/2018/04/12/technology/mark-zuckerberg-testimony

[114] *See* European Parliament hearings of May 22nd, see Senate Hearing of September 5th, well covered by Evelyn Douek, Senate Hearing on Social Media and Foreign Influence Operations: Progress, But There's A Long Way to Go, Lawfare (September 6, 2018), https://www.lawfareblog.com/senate-hearing-social-media-and-foreign-influence-operations-progress-theres-long-way-go.

efforts to tackle misinformation.[115] It importantly unveiled three new initiatives:[116] first, they released the "Facing Facts" Short Film;[117] second, they released an updated news literacy campaign which explains how to spot false news and provides information on what actions Facebook is taking; and third, they announced renewed efforts to work with an independent academic commission[118] to better understand the role of social media in misinformation and democracy:

"*the commission will lead a request for proposals to measure the volume and effects of misinformation on Facebook. They will then manage a peer review process to select which scholars will receive funding for their research, and access to privacy-protected data sets from Facebook. This will help keep us accountable and track our progress over time.*"[119]

In parallel, Facebook has been strengthening the transparency of political online advertising on their platform, including new transparency obligations for advertisers of political content, new forms of labelling of advertisements and the disclosure to the public of who paid for the advertisement's display.[120] In October 2018, they announced in their newsroom that three new independent studies confirmed that Facebook's efforts to tackle misinformation had been successful.[121] While it

---

115 Published on May 23, 2018, available at: https://newsroom.fb.com/news/2018/05/facing-facts-facebooks-fight-against-misinformation.

116 Also *see* N. Thompson, *Exclusive: Facebook Opens Up About False News*, Wired, 2018, https://www.wired.com/story/exclusive-facebook-opens-up-about-false-news/ (last visited Jul 11, 2018).

117 *Id. supra.*

118 As announced in April 2018: https://newsroom.fb.com/news/2018/04/new-elections-initiative

119 "Facing Facts" (May 23, 2018), available at: https://newsroom.fb.com/news/2018/05/facing-facts-facebooks-fight-against-misinformation/.

120 *See* J. B. Merrill, A. Tobin, M. Varner, *What Facebook's New Political Ad System Misses,* ProPublica (May 24, 2018), https://www.propublica.org/article/what-facebooks-new-political-ad-system-misses; J. B. Merrill, A. Tobin, *Facebook's Screening for Political Ads Nabs News Sites Instead of Politicians*, ProPublica (June 15, 2018), https://www.propublica.org/article/facebook-new-screening-system-flags-the-wrong-ads-as-political (last visited September 13, 2018).

121 T. Lyons, *New Research Shows Facebook Making Strides Against False News*, Facebook Newsroom (October 19, 2018), https://newsroom.fb.com/news/2018/10/inside-feed-michigan-lemonde/.

shows that Facebook are making progress toward fulfilling their set goals, this does not show that their efforts are capable of addressing the problem of false news more generally.

As regards Google, in October 2017 it released a blogpost outlining their initiatives on information quality:[122]

"Over the past 18 months, we've undertaken a broad effort to highlight authoritative sources and minimize the spread of misinformation on our platforms. We are continuing these efforts:

1. Since the election we've made significant improvements to demote misleading and misrepresentative sites in search.

2. In 2016 we also introduced the Fact Check Label to provide useful context for people as they explore information online, which is now available globally in search and Google News.

3. *We are also concerned with sites abusing our ads systems by impersonating news organizations so we introduced a new policy against misrepresentative content for AdSense and Ad Exchange publishers and have since taken action against hundreds of publishers."*

Since then, criticism has mounted against Google's YouTube platform's capacity to polarize and radicalize.[123] Still, Google refused to attend the latest Senate Hearing on Social Media and Foreign Influence Operations on September 5th.

There are multiple weighty issues attached to the appropriateness and ability of large companies such as Facebook and Google to independently design their own disinformation policies and compliance programs, and also questions about the role of the state in facilitating or checking on private entities' self-regulatory behavior. On the one hand these companies have an incentive to self-regulate and comply to avoid excessive regulatory burdens and state interference. On the other hand, as it emerged from recent public hearings, they want regulators to intervene and clarify some of the basic rules of the game. In other words, companies want certainty about the rules that they must comply with, but flexibility on how they must comply, and of course they do not want penalties or public shaming. But these

[122] Google Blog: https://www.blog.google/topics/public-policy/security-and-disinformation-us-2016-election/

[123] Z. Tufekci, *YouTube, the Great Radicalizer*, New York Times (March 10, 2018) https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.

companies' perspective is only one side of the complex regulatory puzzle that must be taken into account in designing solutions to the disinformation problem. The other side of the puzzle, as we have explained throughout this paper, is the question of who is best placed to regulate speech: is it the state, which according to US First Amendment doctrine must refrain from regulating news content, or is it private actors who have a degree of independence from public oversight but may sometimes be seen as equally powerful and intrusive as the state?

The answer is probably to be found in new forms of democratic accountability for the actions of companies. This might require some minimal state regulation within and across territorial boundaries, combined with other forms of self-governance including those that Balkin has started to grapple with through his idea of 'information fiduciaries'. Recently a number of large technology companies including Apple, Google, Facebook, Sony and Intel have joined forces with NGOs and research centers to form the Partnership on AI. We must look to this kind of initiative with a critical eye, while maintaining a curious and open mind when it comes to assessing the results of such joint efforts.

5.3. European Efforts to Tackle Fake News

5.3.1. The German NetzDG

In March 2017 German legislators proposed a new "fake news" law, the *Netzwerkdurchsetzungsgesetz* (NetzDG),[124] which was adopted in September 2017, came into force on 1st January 2018 and is the first of its kind. It makes platforms liable to remove hate speech and other offensive content from their platforms within 24 hours in obvious cases or within 7 days in other cases. The law has been widely criticized for being overbroad and misconceived by social media companies[125] as well as a variety of NGOs, academics and other stakeholders.[126]

---

[124] For the original German version: http://www.buzer.de/s1.htm?a=1&g=NetzDG.

[125] S. Shead, *Facebook said Germany's plan to tackle fake news would make social media companies delete legal content* Business Insider (2017), http://uk.businessinsider.com/facebook-says-germany-fake-news-plans-comply-with-eu-law-2017-5 (last visited Jul 9, 2018).

[126] D. Sullivan, *Proposed German Legislation Threatens Free Expression Around the World Global Network Initiative* (2017),

The law applies to "social networks" defined in Section 1(1) as "*tele media providers who operate commercial platforms that are meant to enable users to exchange or share any kind of content with other users or to make such content available to the public*". This definition clearly includes Facebook, YouTube and Twitter. An initial draft also included instant messaging apps such as Whatsapp but the language was revisited and now appears to exclude individualized messaging services. The definition excludes platforms with journalist content for which the platform operator takes full responsibility, platforms that focus on 'specific topics' (such as LinkedIn, or gaming platforms), 'small' social networks with less than two million registered users.

The law imposes a duty on social networks to remove certain categories of content after being notified. Section 1(3) of the law identifies 21 different criminal offences that are eligible to be treated as removable "hate speech". Section 3(2) of the law imposes an obligation on "social networks" to take note of complaints and process them if applicable by removing or blocking the content within 24 hours for content that is 'obviously unlawful' under one of the 21 criminal offences and within 7 days for other content. "Notice" in the law occurs on receipt of the complaint whether or not the social network has actual knowledge of the infringement. This may raise compatibility issues with the European e-Commerce Directive, whose language under Articles 13 and 14, as discussed above, denies immunity only if there is 'actual knowledge' or 'awareness' on the part of the intermediary. "Social networks" must also provide effective and transparent mechanisms for addressing user-complaints. There is also a possibility for "social networks" to hand difficult cases to an independent body within 7 days. This measure has been contested because it still requires the platform to take an expedited decision on whether or not certain content is difficult content that needs to be referred.

The law has not been warmly welcomed. In an open letter to eight EU commissioners, a group of six civil society and industry associations argued that the law is in grave conflict with established EU law and would chill online speech by incentivizing

https://globalnetworkinitiative.org/proposed-german-legislation-threatens-free-expression-around-the-world/ (last visited Jul 9, 2018).

companies to drastically police and remove online content.[127] The law outsources decisions about free speech to private companies who are not well-placed to make such decisions, and it arguably imposes excessive fines of up to 50 million Euro for violations of rules that are not entirely clear. Technology companies for obvious reasons do not like the burden that the NetzDG imposes on them and would prefer for an independent public body to make determinations on the worthiness or accuracy of online content. Extremist right wing politician Beatrix von Storch was very unhappy to see her account suspended and some of her speech removed shortly after the law's coming into force.

A variety of other national efforts are emerging, inspired by the NetzDG,[128] also in countries without the same constitutional and procedural guarantees. Such laws have raised significant concerns from a freedom of speech perspective.

5.3.2. The European Union

In January 2018, the European Commission set up a high-level group of experts, "HLEG" as discussed above, to advise on policy initiatives to counter fake news and disinformation spread online. The Group issued a Report on 12 March 2018,[129] which advised the Commission to avoid narrow simplistic solutions and instead combine short and long-term solutions through a 'multi-dimensional' approach based on a number of parallel efforts organized along five main pillars: (1) a legal and regulatory effort to enhance the transparency of online news, including on data practices; (2) an educational effort to promote digital media literacy; (3) a technical effort to develop tools that empower readers and journalists and allow them to engage in positive public discourse; (4) a cultural effort to preserve and enhance the

---

127 Open Letter to Eight EU Commissioners, *Germany's Draft Network Enforcement Law is a threat to freedom of expression, established EU law and the goals of the Commission's DSM Strategy -- the Commission must take action* (May 22, 2017), https://edri.org/files/201705-letter-germany-network-enforcement-law.pdf.

128 *See* some of the efforts that have been taken around the world in this article: D. Funke, *A guide to anti-misinformation actions around the world*, Polynter (July 24, 2018), http://amp.poynter.org/news/guide-anti-misinformation-actions-around-world?__twitter_impression=true.

129 https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation

diversity and sustainability of the European news media ecosystem, and finally (5) an effort keep promoting research on and monitoring of disinformation in Europe.

A public consultation was also carried out, [130] which resulted in the announcement of a number of measures not already presented by the HLEG Report.[131] These measures include:

- A Code of Practice on Disinformation through which platforms would jointly commit to ensuring advertising transparency, in particular by restricting targeting options for political advertising and reducing revenues for purveyors of disinformation; greater algorithmic clarity and third-party verification; exposure to a plurality of viewpoints and information sources; elimination of fake accounts and bots; and continuous monitoring of online disinformation. A Code of Practice was approved in September 2018 by large online companies including Google and Facebook;[132]
- An independent European network of fact-checkers and a secure European online platform on disinformation both of which will enable common working methods, the exchange of best practices, and the broadest possible coverage of factual corrections across the EU;
- A Coordinated Strategic Communication Policy whose aim is to counter false narratives about Europe and tackle disinformation within and outside the EU.

Other initiatives include media literacy campaigns, elections support against cyber threats in EU Member States, the promotion of voluntary online identification systems, the promotion of media plurality and information quality. New efforts are being made as we write. We are convinced that the HLEG is correct in advocating in favor of a careful approach that seeks to tackle the problem from various angles. What we are less

130 *See* https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-fake-news-and-online-disinformation

131 *See* European Commission Press Release Tackling online disinformation: Commission proposes an EU-wide Code of Practice (26 April 2018), http://europa.eu/rapid/press-release_IP-18-3370_en.htm.

132 S. Writer, *Google, Facebook Agree on EU 'Fake News' Code of Conduct* (September 26th, 2018), https://theglobepost.com/2018/09/26/eu-fake-news-code/.

convinced about is the legitimacy of some of these suggested policies. The Commission's stated goal of ensuring '*the protection of European values and security*'[133] and the proposed '*Coordinated Strategic Communication Policy*' aimed at '*countering false narratives about Europe and tackling disinformation within and outside the EU*'[134] may indeed be inappropriate if carried out by a transnational entity that is not only very far from being a government, but is also under increasing amounts of criticism, lacks democratic legitimacy and is losing public support. In other words, the European Commission should weigh the benefits and costs before acting with excessive hubris in an area as politically sensitive as that of disinformation and fake news in the digital age.

6. Conclusion

In this paper we tried to offer a bird's eye view of the complex dynamics and legal constraints that shape the digital information ecosystem, of how the 'fake news' and disinformation debate fits within this broader picture, and how lawyers and policy-makers should think about possible solutions to the issues at hand. On the one hand, we believe that some action against disinformation is needed, and the best actions focus on the regulation of platforms rather than direct regulation by the state, eg. ensuring that platforms have effective mechanisms for eradicating fake account and coordinate disinformation efforts, ensuring greater transparency and traceability of disinformation and the financial incentives related to it, ensuring appropriate remedies for individuals affected. On the other hand, it seems that governments and institutions around the world, including some European countries, are so eager to regulate fake news that they might overstep their legitimacy bounds in doing so. In this delicate area, we echo the High Level Expert Group and recommend a high level of caution. Free speech values in each territory can be debated and questioned, but they remain fundamental limits to what governments and private entities are allowed to do to fight bad actors in a democracy. A healthier digital news ecosystem cannot be brought about unless we seek an

[133] *Id. supra.*
[134] *Id. supra.*

epistemic common ground, which can only be achieved by understanding the important role of infrastructure and gatekeepers in shaping discourse and by testing new regulatory possibilities.