Stanford – Vienna
Transatlantic Technology Law Forum

A joint initiative of
Stanford Law School and the University of Vienna School of Law

TTLF Working Papers

No. 73

Regulating Freedom of Speech on Social Media: Comparing the EU and US Approach

Marie-Andrée Weiss

2021

TTLF Working Papers

Editors: Siegfried Fina, Mark Lemley, and Roland Vogl

About the TTLF Working Papers

TTLF's Working Paper Series presents original research on technology, and business-related law and policy issues of the European Union and the US. The objective of TTLF's Working Paper Series is to share "work in progress". The authors of the papers are solely responsible for the content of their contributions and may use the citation standards of their home country. The TTLF Working Papers can be found at http://ttlf.stanford.edu. Please also visit this website to learn more about TTLF's mission and activities.

If you should have any questions regarding the TTLF's Working Paper Series, please contact Vienna Law Professor Siegfried Fina, Stanford Law Professor Mark Lemley or Stanford LST Executive Director Roland Vogl at the

Stanford-Vienna Transatlantic Technology Law Forum
http://ttlf.stanford.edu

Stanford Law School
Crown Quadrangle
559 Nathan Abbott Way
Stanford, CA 94305-8610

University of Vienna School of Law
Department of Business Law
Schottenbastei 10-16
1010 Vienna, Austria

About the Author

Marie-Andrée Weiss is a French-American attorney admitted in New York and in Strasbourg, France. Her solo practice focuses on intellectual property, privacy, data protection and social media law. Before becoming an attorney, she worked several years in the fashion and cosmetics industry in New York as a buyer and a director of sales and marketing. She enjoys blogging and is a guest blogger on several sites and maintains two of her own blogs.

General Note about the Content

The opinions expressed in this paper are those of the author and not necessarily those of the Transatlantic Technology Law Forum or any of its partner institutions, or the sponsors of this research project.

Suggested Citation

This TTLF Working Paper should be cited as:
Marie-Andrée Weiss, Regulating Freedom of Speech on Social Media: Comparing the EU and the US Approach, Stanford-Vienna TTLF Working Paper No. 73, http://ttlf.stanford.edu.

Copyright

Abstract

Social media platforms provide forums to share ideas, jokes, images, insults, and threats. These private companies form a contract with their users who agree in turn to respect the platform's private rules, which evolve regularly and organically, reacting sometimes to a particular event, just as legislatures may do.

As these platforms have a global reach, yet are, for the most part, located in the United States, the articulation between the platforms' terms of use and the laws of the states where the users are located varies greatly from country to country.

This article proposes to explore the often-tense relationships between the states, the platforms, and the users, whether their speech creates harm or they are a victim of such harm.

The first part of the article is a general presentation of freedom of expression law. This part does not attempt to be a comprehensive catalog of such laws around the world and is only a general presentation of the U.S. and the European Union laws protecting freedom of expression, using France as an example of a particular country in the European Union. While the principle is freedom of speech, the legal standard is set by international conventions, such as the United Nations Universal Declaration of Human Rights or the European Convention on Human Rights.

The second part of the article presents what the author believes to be the four main justifications for regulating free speech: protecting the public order, protecting the reputation of others, protecting morality, and advancing knowledge and truth. The protection of public order entails the protection of the flag or the king, and lèse-majesté sometimes survives even in a Republic. The safety of the economic market, which may dangerously sway if false information floats online, is another state concern, as is the personal safety of the public. Speech sometimes does harm, even kill, or place an individual in fear for her or his life. The reputation and honor of others is easily smeared on social media, whether by defamation, insults or hate speech, a category of speech not clearly defined by law, which yet is at the center of the debate on online content moderation, including whether there is a right to speak anonymously online. What is "morality" is another puzzling question, as blasphemy, indecency, even pornography, have different legal definitions around the world and private definitions by the platforms. Even truth is an elusive concept, and both states and platforms struggle to define what is "fake news," and whether what is clearly false information, such as denying the existence of the Shoah, should be allowed to be published online. Indeed, while four justifications for regulating speech are delineated in this article, the speech and conduct which should be considered an attack on values worthy to be protected is not equally considered by the different states and the different platforms, and how the barriers to speech are being placed provides a telling picture of the state of democracy.

The third part examines who should have the power to delete speech on social media. States may exert censorship on the platforms or even on the pipes to block access to

speech and punish, sometimes harshly, speakers daring to trespass the barriers to free speech erected by the states. For the sake of democracy, the integrity of the electoral process must not be threatened by false information, whether it spreads false information about the candidates or false information about alleged fraud, or even false information about the result of the vote.

Social media platforms must respect the law. In the United States, Section 230 of the Communications Decency Act of 1996 provides immunity to platforms for third-party content, but also for screening offensive content. Section 230 has been modified several times and many bills, from both sides of the political spectrum, aim at further reform. In the European Union, the E-commerce Directive similarly provides a safe harbor to social media platforms, but the law is likely to change soon, as the Digital Services Act proposal was published in December 2020. The platforms have their own rules, and may even soon have their own private courts, for example the recently created Facebook Oversight Board. However, other private actors may have a say on what can be published on social media, for instance employers or the governing bodies of regulated professions, such as judges or politicians. Even private users may censor the right of others to speak freely, using copyright laws, or may use public shaming to fear speakers into silence. Such fear may lead users to self-censor their speech, to the detriment of the marketplace of ideas, or they may choose to delete controversial messages. Public figures, however, may not have the right to delete social media posts or to block users.

The article was finished the last days of 2020, a year which saw attempts to use social media platforms to sway the U.S. elections by spreading false information, the semi-failed attempt of France to pass a law protecting social media users against hate speech, and false news about the deadly Covid-19 virus spreading online like wildfire, through malicious or naïve posts. A few days after the article was completed, the U.S. Capitol was attacked, on January 6, 2021, by a seditious mob seeking to overturn the results of the Presidential election, believing that the election had been rigged, a false information amplified by thousands of users on social media, including the then President of the United States. Several social media platforms responded by blocking the President's social media accounts, either temporarily or permanently, as did Twitter.

Table of Contents

Introduction

Professor Eugene Volokh noted in 1995 that "[*i*]*t's easier for the rich to speak than it is for the poor*."[1] A year later, John Perry Barlow wrote in *A Declaration of the Independence of Cyberspace* that web users were "*creating a world where anyone, anywhere may express his or her beliefs, no matter how singular, without fear of being coerced into silence or conformity*."[2] We will examine in this article who has the power, or at least the will, to silence and to deafen speakers in cyberspace, particularly on speech published on social media platforms. In 1997, the Supreme Court found the Internet "*not as "invasive" as radio or television.*"[3]

However, the platforms, including social media platforms, now "*provide the main access point to information and other content for most people on the internet today*."[4] Jochen Bittner, political editor for the German weekly newspaper *Die Zeit,* wrote in 2018: "*If the news media is the fourth estate, social media is the fifth, and a welcome check on the government, private sector and traditional journalists*"[5] This "fifth estate" is large and growing: while the world's population is currently 7.7 billion,[6] Facebook reported some

---

[1] Eugene Volokh, 104 YALE L.J. 1805 (1995).
[2] John Perry Barlow, *A Declaration of the Independence of Cyberspace*, https://www.eff.org/cyberspace-independence.
[3] Reno v. American Civil Liberties Union, 521 US 844, 869 (1997).
[4] *Tackling Illegal Content Online –Towards an enhanced responsibility of online platforms,* Communication COM(2017) 555 of September 2017 from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, p.2, https://ec.europa.eu/transparency/regdoc/rep/1/2017/EN/COM-2017-555-F1-EN-MAIN-PART-1.PDF.
[5] *See* Jochen Bittner, *Germans Are Getting on Twitter. Is That a Good Thing?* THE NEW YORK TIMES, (July 27, 2018), https://www.nytimes.com/2018/07/27/opinion/germans-are-getting-on-twitter-is-that-a-good-thing.html.
[6] *Growing at a slower pace, world population is expected to reach 9.7 billion in 2050 and could peak at nearly 11 billion around 2100*, (June 17, 2019), UNITED NATIONS, https://www.un.org/development/desa/en/news/population/world-population-prospects-2019.html, (last visited Dec. 30, 2020).

1.79 billion daily active users during the second quarter of 2020,[7] and Instagram, which is owned by Facebook, reached one billion monthly active users in June 2018.[8] Twitter had "only"330 million monthly active users in the first quarter of 2019.[9]

Justice Kennedy wrote, in 2017, in *Packingham v. North Carolina*:[10]

"*A fundamental First Amendment principle is that all persons have access to places where they can speak and listen, and then, after reflection, speak and listen once more. Today, one of the most important places to exchange views is cyberspace, particularly social media, which offers "relatively unlimited, low-cost capacity for communication of all kinds," Reno v. American Civil Liberties Union, 521 U.S. 844, 870, 117 S.Ct. 2329, 138 L.Ed.2d 874, to users engaged in a wide array of protected First Amendment activity on any number of diverse topics. The Internet's forces and directions are so new, so protean, and so far reaching that courts must be conscious that what they say today may be obsolete tomorrow. Here, in one of the first cases the Court has taken to address the relationship between the First Amendment and the modern Internet, the*

---

[7] *Number of daily active Facebook users worldwide as of 2nd quarter 2020*, STATISTA, https://www.statista.com/statistics/346167/facebook-global-dau, (last visited Dec. 30, 2020).

[8] *Number of monthly active Instagram users from January 2013 to June 2018*, STATISTA, https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users, (last visited Dec. 30, 2020).
According to a survey conducted by the Pew Research Center in 2019, 63% of adult Instagram users in the U.S. use Instagram daily, Brooke Auxier, *8 facts about Americans and Instagram*, PEW RESEARCH CENTER, (Oct. 21, 2020), https://www.pewresearch.org/fact-tank/2020/10/21/8-facts-about-americans-and-instagram.

[9] *Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019*, STATISTA, https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users, (last visited Dec. 30, 2020).

[10] Packingham was a case about a North Carolina law forbidding registered sex offenders "*to access a commercial social networking Web site where the sex offender knows that the site permits minor children to become members or to create or maintain personal Web pages*." The case is discussed further on in this article.

Court must exercise extreme caution before suggesting that the First Amendment provides scant protection for access to vast networks in that medium."[11]

Justice Kennedy noted further that "*social media users employ these websites to engage in a wide array of protected First Amendment activity on topics "as diverse as human thought*."[12] However, the low-cost of entry into the market also lead to a wealth of speech having no or little value.[13] But social media platforms give people, including poor and disenfranchised people, the power to easily produce speech as well as the power to easily receive it. In Zimbabwe, Pastor Evan Mawarine launched a popular social media campaign[14] against President Robert Mugabe, on Facebook and on Twitter, using the account @PastorEvanLive and the hashtag #ThisFlag.[15] In 2019, the image of Alaa Salah, 22-year-old student from Karthoum, dressed in a white *tobe* garment and chanting to protest against President Omar al-Bashir, became viral.[16] The #SudanUprising hashtag

---

[11] Packingham v. North Carolina, 137 S. Ct. 1730, 1735 (2017).
[12] Citing Reno, at 870.
[13] Keegan Hankes, a senior research analyst for the Southern Poverty Law Center, was quoted in this article as describing Twitter as "an absolute cesspool" of hate, *see* Jefferson Graham, *A 'direct correlation' between rise of hate on social media and attacks, says SPLC*, USA TODAY, https://www.usatoday.com/story/tech/talkingtech/2019/08/05/twitter-called-cesspool-for-hate-facebook-youtube-not-much-better/1924816001/.
[14] Bruce Mutsvairo, *Can Robert Mugabe be tweeted out of power?*, THE GUARDIAN, (July 26, 2016 10.02 EDT), https://www.theguardian.com/global-development-professionals-network/2016/jul/26/robert-mugabe-grassroots-protest-zimbabwe-social-media.
[15] This refers to the video shared online in 2016 by Mr. Mawarine, which started the whole movement. He posted a video on YouTube commenting about the colors of the Zimbabwean flag: "
"*This flag, this beautiful flag, they tell me that the green is for the vegetation and the crops. I don't see any crops in my country. The yellow is for all the minerals... I don't know how much is left. I don't know who they sold it to and how much they got for it. They tell me that the black is for the majority people like me and yet for some reason I don't feel like I am a part of it.*" *See* Kiri Rupiah, *Zimbabweans at home and globally take to social media in support of #ThisFlag protest* (May 19, 2016), MAIL & GUARDIAN, http://mg.co.za/article/2016-05-19-zimbabweans-at-home-and-globally-take-to-social-media-in-support-of-thisflag-protest.
[16] *See* @HindMakki , Twitter, (April 8, 2019, 3:35PM), https://twitter.com/HindMakki/status/1115337418301935616. *See* also Jason Burke, *'Inspiring' protester becomes symbol of resistance for Sudanese women*, THE GUARDIAN, (Apr.9 2019 13.18 EDT), https://www.theguardian.com/world/2019/apr/09/inspiring-protester-khartoum-becomes-symbol-of-resistance-for-sudanese-women.

became viral and contributed to direct a spotlight on Sudanese protests. The hashtag #TasgutBas ("just fall") had been used a few months earlier to call for the removal of President al-Bashir.[17] The founders of the app Periscope, which allows its users to broadcast event live on social media, had the idea for such an app while in Istanbul during the Taksim Square protests in 2013.[18] The hashtag #OccupyGezy became viral and allowed social media users around the world to follow the events.[19] Social media is credit by many for having led to the January 2011 overthrow of Tunisia President Ben Ali.[20]

Indeed, "[*g*]*reat advances in technology sometimes directly increase the power of ordinary people.*"[21] France's constitutional Council, the *Conseil constitutionnel* , whose mission is to decide whether laws are constitutional, took care to state in 2009 that:

> "[*i*]*n the current state of the means of communication and given the generalized development of public online communication services and the importance of the latter for the participation in democracy and the expression of ideas and opinions, this right implies freedom to access such services*."[22]

---

[17] *See* Aya Elmileik & Seena Khalil, *'Tasgut bas' to #SudanUprising: How social media told the story,* ALJAZEERA, (AUG. 12, 2019), https://www.aljazeera.com/indepth/features/bas-rihanna-tweets-sudan-protests-viral-190811103901411.html.
[18] Sophie Curtis, *Twitter's Periscope is a 'platform for truth', claims founder,* THE TELEGRAPH (July 5, 2015, 1:30 PM BST), http://www.telegraph.co.uk/technology/twitter/11717256/Twitters-Periscope-is-a-platform-for-truth-claims-founder.html.
[19] *Turkey protests spread online, and in the streets*, FRANCE24, (June 1, 2013 11:05), https://www.france24.com/en/20130601-turkey-protests-spread-online
[20] *Social Media Gets Credit For Tunisian Overthrow*, NPR WEEKDEND ADDITION SUNDAY, (Jan.16, 2011, 8:00 AM ET), https://www.npr.org/2011/01/16/132975274/Social-Media-Gets-Credit-For-Tunisian-Overthrow.
[21] THE OFFENSIVE INTERNET, 195 (JOHN DEIGH, FOUL LANGUAGE: SOME RUMINATIONS ON COHEN V. CALIFORNIA), Saul Levmore and Martha C. Nussbaum, eds., 2010), giving as examples the invention of telephone, automobiles, and information technology.
[22] Conseil constitutionnel [CC][Constitutional Court] decision No. 2009-580, June 10, 2009. Available in English at http://www.conseil-constitutionnel.fr/conseil-constitutionnel/root/bank/download/2009580DC2009_580dc.pdf .

But governments, platforms and even users have the power to police online speech. A government may choose to enact laws making a particular speech a crime, or at least a tort. Platforms may mirror these edicts in their own rules or may decide to ban speech legal in some countries, but not in others, thus defining their own online social order. Users may follow their own moral compasses when deciding to block some users or may decide to tolerate some mild abuse in the spirit of free speech. If the abuse becomes too much to bear alone, then the laws of their countries, if available to them, or the laws of the platforms, may provide help and redress. Twitter explains that it offers its services "*to give everyone the power to create and share ideas and information instantly, without barriers.*"[23] As "*Twitter's purpose is to serve the public conversation*," this power cannot be without limit and the Twitter rules "*are to ensure all people can participate in the public conversation freely and safely*."[24] Online speech is also regulated by the laws of the countries where the social media user publishes its speech. Such laws may also be used to censor some speech after publication. For instance, France declared a state of emergency in November 2015 following the Paris terrorist attacks on the night of November 13th, which gave, inter alia, power to the government to take "any measures" necessary to block websites condoning or inciting terrorism or terrorist act.[25] North Korea blocked gambling and "sex and adult websites"[26] in March 2016. Same action, blocking, but to protect very different values:

---

[23] *Twitter, our services, and corporate affiliates*, TWITTER, https://help.twitter.com/en/rules-and-policies/twitter-services-and-corporate-affiliates, (last visited Dec. 30, 2020).
[24] THE TWITTER RULES, https://help.twitter.com/en/rules-and-policies/twitter-rules, (last visited Dec. 30, 2020).
[25] FRENCH GOVERNMENT, *State of emergency in France: what are the consequences?* https://www.gouvernement.fr/en/state-of-emergency-in-metropolitan-france-what-are-the-consequences (last visited Dec. 30, 2020).
[26] Eric Talmadge for AP in Pyongyang, *North Korea announces blocks on Facebook, Twitter and YouTube*, THE GUARDIAN, (Apr. 1 2016, 07.56 EDT), https://www.theguardian.com/world/2016/apr/01/north-korea-announces-blocks-on-facebook-twitter-and-youtube.

homeland security for France, moral for North Korea. Terrorism is a threat to democracy, while gambling and pornography are not, even if they have poor societal values.

One of the justifications of the almost limitless scope of the First Amendment is the need of a robust marketplace of ideas. John Stuart Mills argued in 1859 that:

"the peculiar evil of silencing the expression of an opinion is, that it is robbing the human race; posterity as well as the existing generation; those who dissent from the opinion, still more than those who hold it. If the opinion is right, they are deprived of the opportunity of exchanging error for truth: if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error."[27]

More than 150 years later, the (NTIA) wrote in its July 27, 2020 *Petition for Rulemaking of the National Telecommunications and Information Administration, In the Matter of Section 230 of the Communications Act of 1934,* that "*social media and its growing dominance present troubling questions on how to preserve First Amendment ideals and promote diversity of voices in modern communications technology*."[28] Are social media platforms a threat to the First Amendment?

[27] JOHN STUART MILLS, ON LIBERTY, (The Walter Scott Publishing Co., Ltd), (Projet Gutenberg eBook #34901, 2011), 31, available at https://www.gutenberg.org/files/34901/34901-h/34901-h.htm#Page_28.
[28] Petition for rulemaking of the national telecommunications and Information Administration, In the Matter of Section 230 of the Communications Act of 1934 (July 27, 2020), available at https://www.ntia.gov/files/ntia/publications/ntia_petition_for_rulemaking_7.27.20.pdf.

1st Part: General Presentation of Freedom of Expression Laws

In the U.S. as in other Western countries, freedom of expression is the rule, but there are limits to this freedom. How the states place these barriers and whether such barriers can be overcome or not may offer clues to the state of democracy of a particular country.

I. The Principle: Freedom of Speech

A. International Laws

The right to produce and to receive speech is protected by several international conventions. For instance, Article 9 of the African Charter on Human and Peoples' Rights states that every individual has the right to receive information and "*to express and disseminate his opinions within the law*."[29] Article 13-1 of the American Convention on Human Rights, adopted on November 22, 1969, proclaims the right of freedom of thought and expression to everyone, which "*includes freedom to seek, receive, and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing, in print, in the form of art, or through any other medium of one's choice*."[30] The American Declaration of the Rights and Duties of Man, which was adopted in 1948 in Bogotá, Colombia, by the Ninth International Conference of American States, states in its article IV that "*[e]very person has the right to freedom of investigation, of opinion, and of the expression and dissemination of ideas, by any medium whatsoever*."[31] This Declaration was the first international human

[29] *African Charter on Human and Peoples' Rights*, adopted June 27, 1981, OAU Doc. CAB/LEG/67/3 rev. 5, 21 I.L.M. 58 (1982), entered into force October 21, 1986), available at https://www.achpr.org/legalinstruments/detail?id=49.
[30] *American Convention on Human Rights*, adopted at the Inter-American Specialized Conference on Human Rights, San José, Costa Rica, November 22, 1969. Available at https://www.cidh.oas.org/basicos/english/basic3.american%20convention.htm.
[31] American Declaration of the Rights and Duties of Man, adopted by the Ninth International Conference of American States, Bogotá, Colombia, 1948), available at http://www.cidh.oas.org/Basicos/English/Basic2.American%20Declaration.htm .

rights instrument of a general nature,[32] as it precedes by few months the Universal Declaration of Human Rights, proclaimed by the General Assembly United Nations on December 10, 1948. According to Article 19 of the Universal Declaration of Human Rights, "[*e*]*veryone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.*"[33] Article 19 of the International Covenant on Civil and Political Rights of the United Nations provides that freedom of expression includes the "*freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice*." There are some limits to this right, but only "*as are provided by law and are necessary… [f]or respect of the rights or reputations of others [and] [f]or the protection of national security or of public order (ordre public), or of public health or morals*."[34] Therefore, these limitations to free speech must be "provided by law" as well as necessary and proportional to their goals, namely protecting the reputation of others and national security.

On September 12, 2001, the UN Human Rights Committee adopted General Observation n°34 on article 19 of the Universal Declaration of Human Rights protecting freedom of opinion and freedom of expression.[35] Its paragraph 9 states:

---

[32] See What is the IACHR? OAS, http://www.oas.org/en/iachr/mandate/what.asp, (last visited Dec. 30, 2020).
[33] Universal Declaration of Human Rights, G.A. Res. 217 (III) A, U.N. Doc. A/RES/217(III) (Dec. 10, 1948). It should be noted that the Declaration is not legally binding, even for its signatories.
[34] International Covenant on Civil and Political Rights Adopted and opened for signature, ratification and accession by General Assembly resolution 2200A (XXI), Dec. 16, 1966.
[35] Human Rights Council General comment 34, 102nd Sess., July 11-29 2001, CCPR/C/GC/34(Sep.12, 2011).

"*Paragraph 1 of article 19 requires protection of the right to hold opinions without interference. This is a right to which the Covenant permits no exception or restriction. …No person may be subject to the impairment of any rights under the Covenant on the basis of his or her actual, perceived or supposed opinions. All forms of opinion are protected, including opinions of a political, scientific, historic, moral or religious nature. It is incompatible with paragraph 1 to criminalize the holding of an opinion. The harassment, intimidation or stigmatization of a person, including arrest, detention, trial or imprisonment for reasons of the opinions they may hold, constitutes a violation of article 19, paragraph 1*."

International law recognizes, however, that freedom of speech can be limited. For instance, the Human Rights Committee of the United Nations explained in 1994 in its *Womah Mukong v. Cameroon* Communication that:

"[*u*]*nder article 19, everyone shall have the right to freedom of expression. Any restriction of the freedom of expression pursuant to paragraph 3 of article 19 must cumulatively meet the following conditions: it must be provided for by law, it must address one of the aims enumerated in paragraph 3(a) and (b) of article 19, and must be necessary to achieve the legitimate purpose*."[36]

Article 20.2 of the UN International Covenant on Civil and Political Rights, adopted on December 16, 1966, states that "[*a*]*ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law*."[37]

[36] Womah Mukong v. Cameroon, Communication No. 458/1991, U.N. Doc. CCPR/C/51/D/458/1991 (1994).
[37] G.A. United Nations, Treaty Series, vol. 999, p. 171 (Dec. 16, 1966).

All of these declarations are a testimony of the will of the countries, around the world, to reaffirm their commitment to human rights after the horrors of World War II. However, they are declaration of intent with no legal authority. European countries were especially aware that human rights are the safeguards of democracy and, as such, needed to be provided some teeth.

B. European Union law

Article 11-1 of the Charter of Fundamental Rights of the European Union states that "[*e*]*veryone 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,*" and its article 11-2 states that" [*t*]*he freedom and pluralism of the media shall be respected*." The European Convention on Human Rights (ECHR)[38] was adopted on November 4, 1950, by the members of the Council of Europe, which must, under Article 1 of the Convention, "*secure to everyone within their jurisdiction the rights and freedoms defined in Section I of* [*the*] Convention, including its article 10 stating that everyone has the right to freedom of expression, which includes "*freedom to… receive and impart information and ideas without interference by public authority."*

Article 10.2 of the ECHR explains how the general right to freedom of expression may be subjected to formalities or conditions, or even restricted altogether. This may happen if a law prescribes it, if doing so is "*necessary in a democratic society*" to achieve other worthy goals enumerated by article 10.2. They are:

[38] Convention for the Protection of Human Rights and Fundamental Freedoms, Nov. 4, 1950, Europ.T.S. No. 5; 213 U.N.T.S. 221

"*the interests of national security, territorial integrity or public safety, … the prevention of disorder or crime… the protection of health or morals …the protection of the reputation or rights of others…* [*the prevention of*] *the disclosure of information received in confidence*, [*the maintenance of*] *the authority and impartiality of the judiciary*."

Unlike the Universal Declaration of Human Rights, the European Convention is enforceable and enforced. The Convention established the European Court of Human Rights (ECtHR), which famously held in *Handyside v. The United Kingdom*[39] that:

> "*freedom of expression constitutes one of the essential foundations of such [a democratic society], one of the basic conditions for its progress and for the development of every man. … {I]t is applicable not only to "information" or "ideas" that are favorably 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". This means, amongst other things, that every "formality", "condition", "restriction" or "penalty" imposed in this sphere must be proportionate to the legitimate aim pursued."*[40]

[39] Handyside v. the United Kingdom, 1 EHRR 737 (1976). The speech at stake in Handyside was *The Little Red Schoolbook*, a book written by two Danish schoolteachers for teenagers, which addressed teenager sexuality and gave this advice on masturbation: "*If anybody tells you it's harmful to masturbate, they're lying. If anybody tells you you mustn't do it too much, they're lying too, because you can't do it too much. Ask them how often you ought to do it. They'll usually shut up then*." The book had been first published in Denmark in 1969 without any issues, then translated and published without challenges in Belgium, Finland, France, West Germany, Greece, Iceland, Italy, the Netherlands, Norway, Sweden and Switzerland. However, when published in Great Britain, it had been deemed obscene under the Obscene Publication Act.
[40] ECHR, Handyside v. The United Kingdom, §49.

The Court explained since in several cases how it determines whether an interference is "*necessary in a democratic society*":

> "(i) *Freedom of expression constitutes one of the essential foundations of a democratic society and one of the basic conditions for its progress and for each individual's self-fulfilment. Subject to paragraph 2 of Article 10, 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. Such are the demands of pluralism, tolerance and broadmindedness without which there is no 'democratic society'. As set forth in Article 10, this freedom is subject to exceptions, which ... must, however, be construed strictly, and the need for any restrictions must be established convincingly ...*
>
> (ii) *The adjective 'necessary', within the meaning of Article 10 § 2, implies the existence of a 'pressing social need'. The Contracting States have a certain margin of appreciation in assessing whether such a need exists, but it goes hand in hand with European supervision, embracing both the legislation and the decisions applying it, even those given by an independent court. The Court is therefore empowered to give the final ruling on whether a 'restriction' is reconcilable with freedom of expression as protected by Article 10*.
>
> (iii) *The Court's task, in exercising its supervisory jurisdiction, is not to take the place of the competent national authorities but rather to review under Article 10 the decisions they delivered pursuant to their power of appreciation. This does not mean that the supervision is limited to ascertaining whether the respondent State exercised*

its discretion reasonably, carefully and in good faith; what the Court has to do is to look at the interference complained of in the light of the case as a whole and determine whether it was 'proportionate to the legitimate aim pursued' and whether the reasons adduced by the national authorities to justify it are 'relevant and sufficient' ... In doing so, the Court has to satisfy itself that the national authorities applied standards which were in conformity with the principles embodied in Article 10 and, moreover, that they relied on an acceptable assessment of the relevant facts ..."[41]

The ECHR cannot, however, be used as a shield and a weapon. Its Article 17 allows the ECtHR to declare inadmissible an application if it considers that one of the parties to the proceedings invokes the Convention to engage in activities violating the very rights and freedoms protected by the Convention. For example, in the *Garaudy v France* case, which will be discuss further on, the ECtHR explained why applicant, the writer of a book denying the Holocaust, could not invoke article 10,[42] as:

"*the main content and general tenor of the applicant's book, and thus its aim, are markedly revisionist and therefore run counter to the fundamental values of the Convention, as expressed in its Preamble, namely justice and peace. It considers that the applicant attempts to deflect Article 10 of the Convention from its real purpose by using his right to freedom of expression for ends which are contrary to the text and spirit of the Convention. Such ends, if admitted, would contribute to the destruction of the rights and freedoms guaranteed by the Convention*."

[41] For instance, Hertel v. Switzerland, 25 August 1998, § 46, Reports of Judgments and Decisions 1998-VI; Steel and Morris v. the United Kingdom, no. 68416/01, § 87, ECHR 2005-II.
[42] Garaudy v. France, Application No. 65831/01, ECtHR (2003)

Article 10(2) of the ECHR provides that freedom of expression may be subject to "*restrictions or penalties as are prescribed by law and are necessary in a democratic society… for the prevention of disorder or crime… [or] for the protection of the reputation of the rights of others…*"[43] As noted by Professor Michel Verpeaux, "[*f*]*reedom of expression is…sanctioned in various forms or names, in all the countries constituting the Council of Europe. It is often treated as one of the most fundamental freedoms, and constitutional courts take inspiration from European case law, but it is designed, like any other freedoms, as not being absolute."*[44]

There cannot be a vibrant freedom of expression without pluralism of ideas, which is acknowledged in a 1999 Recommendation of the Committee of Ministers to the Member States "*on measures to promote media pluralism*", stating as a "*general principle*" that "*Member States should consider possible measures to ensure that a variety of media content reflecting different political and cultural views is made available to the public, bearing in mind the importance of guaranteeing the editorial independence of the media and the value which measures adopted on a voluntary basis by the media themselves may also have*."[45]

The European Union does not generally protect speech hurtful to individuals, especially if they belong to minorities. The Recommendation No. R (97) 20 on Hate Speech,

---

[43] Convention for the Protection of Human Rights and Fundamental Freedoms, Nov. 4, 1950, art. 10, available at http://www.echr.coe.int/Documents/Convention_ENG.pdf.
[44] Michel Verpeaux, *La liberté d'expression dans les jurisprudences constitutionnelles*, Nouveaux Cahiers du Conseil constitutionnel n° 36, June 2012, p. 3, available at http://www.conseil-constitutionnel.fr/conseil-constitutionnel/root/bank/pdf/conseil-constitutionnel-114765.pdf. The article reviews how the different European countries are viewing freedom of speech by studying their respective constitutional courts freedom of speech case law.
[45] Council of Eur., Comm. Of Ministers, Recommendation No. R (99) 1 to Member States on Measures to Promote Media Pluralism , adopted on January 19, 1999, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016804fa377.

adopted on October 30, 1997 by the Council of Europe Committee of Ministers, had defined hate speech as "*covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, antisemitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin*."[46] The European Union Council Framework Decision 2008/913/JHA of November 28, 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law[47] asked Member States to criminalize "*conduct publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin*." The purpose of Framework was "*to ensure that certain serious manifestations of racism and xenophobia are punishable by effective, proportionate and dissuasive criminal penalties throughout the European Union.*"[48]

The Framework defined hate speech as:

"*public incitement to violence or hatred directed against a group of persons or a member of such a group defined on the basis of race, colour, descent, religion or belief, or national or ethnic origin;-the above-mentioned offence when carried out by the public dissemination or distribution of tracts, pictures or other material; publicly condoning, denying or grossly trivialising crimes of genocide, crimes against humanity*

[46] Council of Eur., Comm. Of Ministers, Recommendation No. R (97) 20 to Member States on Hate Speech, adopted on October 30, 1997, http://www.coe.int/t/dghl/standardsetting/media/doc/cm/rec%281997%29020&expmem_EN.asp

[47] European Union Council Framework Decision 2008/913/JHA of November 28, 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law, available at http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:l33178.

[48] Council Framework Decision on combating certain forms and expressions of racism and xenophobia by means of criminal law 2008/913/JHA, 2008 O.J. (L 328).

and war crimes as defined in the Statute of the International Criminal Court (Articles 6, 7 and 8) and crimes defined in Article 6 of the Charter of the International Military Tribunal, when the conduct is carried out in a manner likely to incite violence or hatred against such a group or a member of such a group" and further notes that "[*i*]*nstigating, aiding or abetting in the commission of the above offences is also punishable*."

The seventy-year ECHR applies to speech on social media. The ECtHR noted in its 2016 *Cengiz v. Turkey* case:

"*the importance of Internet sites in the exercise of freedom of expression… [as][i]n the light of its accessibility and its capacity to store and communicate vast amounts of information, the Internet plays an important role in enhancing the public's access to news and facilitating the dissemination of information in general"* and noting further that "*political content ignored by the traditional media is often shared via YouTube, thus fostering the emergence of citizen journalism*."[49]

In 2011, the Committee of Ministers acknowledged, in another Recommendation,[50] the importance of social media platforms in the current marketplace of ideas:

"*People, notably civil society representatives, whistleblowers and human rights defenders, increasingly rely on social networks, blogging websites and other means of*

[49] Cengiz v. Turkey, Application No. 48226/10, ECtHR, (2015), § 52.
[50] Council of Eur., Comm. Of Ministers, Declaration of the Committee of Ministers on the protection of freedom of expression and freedom of assembly and association with regard to privately operated Internet platforms and online service providers, adopted on December 7, 2011, https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016805cb844.

mass communication in aggregate to access and exchange information, publish content, interact, communicate and associate with each other. These platforms are becoming an integral part of the new media ecosystem. Although privately operated, they are a significant part of the public sphere through facilitating debate on issues of public interest; in some cases, they can fulfil, similar to traditional media, the role of a social "watchdog" and have demonstrated their usefulness in bringing positive real-life change."

C. National Laws

Outside of these regional laws, many individual States have laws generally proclaiming freedom of speech while setting limits. We will see how the United States of America and France limit freedom of speech.

a. U.S. Laws

Free speech is protected since 1791 by the First Amendment to the U.S. Constitution, which is part of the Bill of Rights. Expressive conduct is also protected by the First Amendment, whether it is displaying a red flag in public,[51] burning a flag,[52] burning a draft card in public,[53] or liking the Facebook Page of a political campaign.[54] The First Amendment prevents Congress from making laws abridging the freedom of speech or freedom of the press: "*Congress shall make no law …abridging the freedom of speech, or of*

[51] Stromberg v. California, 283 U.S. 359 (1931).
[52] Texas v. Johnson, 491 U.S. 397 (1989).
[53] U.S. v. O'Brien, 391 U.S. 367 (1968).
[54] Bland v. Roberts, 730 F.3d 368, 386 (4th Cir. 2013) : "[*a*]*side from the fact that liking the Campaign Page constituted pure speech, it also was symbolic expression. The distribution of the universally understood "thumbs up" symbol in association with Adams's campaign page, like the actual text that liking the page produced, conveyed that Carter supported Adams's candidacy*."

the press…." It is applicable to the States, as it was incorporated in 1925 in the due process clause of the Fourteenth Amendment as held by the Supreme Court in 1925 *in Gitlow v. New York*.[55] It is also applicable to the Internet, as held by the Supreme Court in 1997 in *Reno v. ACLU*,[56] and to speech published on social media platforms, since, as noted by the Supreme Court in *Brown v. Entertainment Merchants Ass'n*:

> "whatever *the challenges of applying the Constitution to ever-advancing technology, "the basic principles of freedom of speech and the press, like the First Amendment's command, do not vary" when a new and different medium for communication appears*."[57]

"*Congress shall make no law …abridging the freedom of speech, or of the press…*." However, some U.S. laws do "abridge" freedom of speech or of the press. Abridging is diminishing,[58] and the scope of this freedom is not indefinite. How the barriers are set, and

---

[55] Gitlow v. New York, 268 U.S. 652, 666 (1925). "*We may and do assume that freedom of speech and of the press — which are protected by the First Amendment from abridgment by Congress — are among the fundamental personal rights and "liberties" protected by the due process clause of the Fourteenth Amendment from impairment by the States*."

[56] Reno v. American Civil Liberties Union, 521 US 844, 870 (1997): "…the Internet can hardly be considered a "scarce" expressive commodity. It provides relatively unlimited, low-*cost capacity for communication of all kinds. … This dynamic, multifaceted category of communication includes not only traditional print and news services, but also audio, video, and still images, as well as interactive, real-time dialogue. Through the use of chat rooms, any person with a phone line can become a town crier with a voice that resonates farther than it could from any soapbox. Through the use of Web pages, mail exploders, and newsgroups, the same individual can become a pamphleteer. As the District Court found, "the content on the Internet is as diverse as human thought."…We agree with its conclusion that our cases provide no basis for qualifying the level of First Amendment scrutiny that should be applied to this medium.*" This reasoning obviously applies to social media sites which provide a powerful soap box to town criers and haters alike.

[57] Brown v. Entertainment Merchants Ass'n, 564 US 786, 790 (2011), citing Joseph Burstyn, Inc. v. Wilson, 343 U.S. 495, 503 (1952).

[58] Merriam-Webster Dictionary, *Definition of abridge*, https://www.merriam-webster.com/dictionary/abridge?src=search-dict-box (last visited Dec. 30, 2020).

based on which standards, is of great interest, as the rationale for abridging such an important right is telling of a society's aspirations and values.

Even though the Bill of Rights was enacted in 1791, the First Amendment, as we know it, a constitutional right knowing little limitations, emerged from the Supreme Court jurisprudence only after World War I. Justice Holmes wrote in 1919 the Court's unanimous decision in *Schenk v. United States* that "*the prohibition of laws abridging the freedom of speech is not confined to previous restraints, although to prevent them may have been the main purpose.*"[59] Indeed, the First Amendment was first interpreted as merely preventing prior restraint on speech.[60] Justice Holmes explained in *Schenck* that only a "*clear and present danger*" can limit speech. A few months later, Justice Holmes wrote his now famous dissent in *Abrams v. United States*, in which Justice Brandeis concurred, which is considered to be the founding stone of the modern free speech jurisprudence, as it offered an explanation on why freedom of expression should know little if no bounds.[61] Justice Holmes famously advocated "*free trade in ideas*" because "*the best test of truth is the power of the thought to get itself accepted in the competition of the market*."[62] Allowing free flow of outrageous and shocking ideas is justified as it provides an open and free marketplace of ideas where ideas are competing against one another for people's attention, approval or disapproval.[63]

[59] Schenk v. United States, 249 U.S. 47, 51-52 (1919).
[60] Paterson v. Colorado, 205 U.S. 454, 462 (1907).
[61] Abrams v. United States, 250 US 616 (1919).
[62] Abrams, at 630.
[63] Abrams, at 630.

Almost sixty years later, Justice Powell wrote for the majority in *Gertz v. Robert Welch*, that "[*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 not on the conscience of judges and juries but on the competition of other ideas*."[64] In 1989, Justice Brennan writing the *Texas v. Johnson* Supreme Court opinion, noted that "*if there is a bedrock principle underlying the First Amendment, it is that the government may not prohibit the expression of an idea simply because society finds the idea itself offensive or disagreeable*."[65] In that case, the Court had found that a Texas law which made a crime to desacralize the flag of the United States was unconstitutional because burning the flag is expressive conduct protected by the First Amendment and the law was related to suppress this expression. One year earlier, in 1988, the Supreme Court had held in *Hustler Magazine v. Falwell* case that:

> "[*a*]*t the heart of the First Amendment is the recognition of the fundamental importance of the free flow of ideas and opinions on matters of public interest and concern. The freedom to speak one's mind is not only an aspect of individual liberty -- and thus a good unto itself -- but also is essential to the common quest for truth and the vitality of society as a whole."*[66]

The First Amendment is often presented as having almost no limit. Justice Brennan, delivering the opinion of the Court in Roth v. United States, took care to state that "[*a*]*ll ideas having even the slightest redeeming social importance—unorthodox ideas,*

[64] Gertz v. Robert Welch, Inc., 418 U.S. 323, 339 (1974).
[65] Texas v. Johnson, 491 US 397, 414 (1989).
[66] Hustler Magazine, Inc. v. Falwell, 485 US 46, 50 (1988).

controversial ideas, even ideas hateful to the prevailing climate of opinion—have the full protection of the guaranties, unless excludable because they encroach upon the limited area of more important interests, before stating that obscenity does not have that redeeming social importance which would grant it this protection.[67] Justice Brandeis, in a concurring opinion, in which he was joined by Justice Holmes, wrote in *Whitney v. California* that, while the right to free speech is fundamental, it is not absolute, and its exercise is subject to restriction.[68] Such abridgment cannot only be to bar expression of ideas which are not acknowledged as being ideas fitting a particular party line or reflecting the values of the society, at least at a particular moment in time. As explained by the Supreme Court in *Texas v. Johnson*[69], "[*i*]*f there is a bedrock principle underlying the First Amendment, it is that the government may not prohibit the expression of an idea simply because society finds the idea itself offensive or disagreeable.*"

The U.S. do not recognize many exceptions to free speech, but there are some, which have been listed by the Supreme Court in its 1942 *Chaplinsky v. New Hampshire* case:

"[*t*]*here are certain well-defined and narrowly limited classes of speech, the prevention and punishment of which have never been thought to raise any Constitutional problem. These include the lewd and obscene, the profane, the libelous, and the insulting or "fighting" words — those which by their very utterance inflict injury or tend to incite an immediate breach of the peace*."[70]

[67] Roth v. United States, 354 U.S. 476, 484-485 (1957).
[68] Whitney v. California, 274 U.S.357, 373 (1927).
[69] Texas v. Johnson, 491 U. S. 397, 414 (1989).
[70] Chaplinsky v. New Hampshire, 315 U.S. 568, 571-572 (1942).

Justice Scalia wrote in *R. A. V. v. City of St. Paul* that the First Amendment does restrict "*speech in a few limited areas, which are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality*."[71] The government can regulate speech to promote a compelling interest and if it chooses the least restrictive means to further this interest.[72] It has been recognized by the courts that the Government has a compelling interest to regulate speech of lower social value such as true threats, incitement of imminent lawless action, obscenity, defamation, speech integral to criminal conduct, fighting words, child pornography and fraud.[73] So the four main limitations to free speech in the U.S. are obscene speech, defamatory speech, speech advocating immediate violent of illegal action, and desecrating or vulgar speech.

b. French Laws

Article 10 of the French Declaration of the Rights of Man and of the Citizen (DRMC), enacted by France's National Assembly on August 26, 1789, a few weeks after the start of the French Revolution, states that "[*n*]*o one can be disturbed on account of his opinions, even religious ones*." Its article 11 declares that "*free communication of thoughts and opinions is one of the most precious rights of man: any citizen may therefore speak, write and publish freely, subject to responsibility for the abuse of this freedom as shall be defined by law*." Both the principle of freedom of expression and of its limits are thus expressed in the same phrase. Only the law may determine what constitutes abuse of freedom of expression. The

[71] R. A. V. v. City of St. Paul, 505 U.S. 377, 382-383 (1992).
[72] Sable Communications of Cal. Inc. v. FCC, 492 U.S. 115, 126 (1989).
[73] United States v. Alvarez, 132 S. Ct. 2537, 2544 (2012).

freedom of expression right of Article 11 of the French Declaration of Human Rights extends to online communication, as explained by the French Constitutional Council in 2009:

> *"In the current state of the means of communication and given the generalized development of public online communication services and the importance of the latter for the participation in democracy and the expression of ideas and opinions, this right implies freedom to access such services*."[74]

Article 4 of the DRMC further states that "[*f*]*reedom is being able to do anything that does not harm others [and that] the exercise of the natural rights of each man has no limits except those which ensure to the other members of society the enjoyment of these same rights. These limits can only be determined by law*." These rights are natural, meaning that lawmakers did not create them but, instead, are inherent to men. However, the 1789 Declaration is not merely a 'feel-good' document, stating lofty values that France should aim to attain. Instead, it has normative value, as it belongs to the 'constitutionality block,' the *bloc de constitutionalité*, which is constituted by the 1958 French Constitution, its Preamble, the Preamble of the 1946 French Constitution and the DRMC. The *Conseil d'État* (Council of State), France's highest administrative court, recognized the DRMC as being part of this bloc in 1960.[75] The *Conseil constitutionnel* (the constitutional Council), which reviews the constitutionality of laws, recognized it in 1973[76] and considers since that a law

[74] Conseil constitutionnel [CC] [Constitutional Court] decision Decision n° 2009-580 of June 10, 2009, paragraph 12.
[75] CE Sect.Feb.12, 1960, Soc. Eky, Rec. Lebon 101
[76] Conseil constitutionnel [CC] [Constitutional Court] decision No. 73-51 DC, Dec. 27, 1973, Rec.25 (Fr.)

violating the DRMC is unconstitutional.[77] For instance, the Constitutional Council stated in its July 1, 2004 decision, quoting article 11 of the DRMC, that "*the pluralism of schools of thoughts and opinions is in itself a constitutional objective [and] respect for their expression is a prerequisite for democracy*."[78] In order for all of these opinions to be expressed, there must be a plurality of publications. The Constitutional Council stated that the plurality of political and general daily information is a constitutional objective, adding that:

> " *the free communication of ideas and of opinions, as guaranteed by Article 11 of the [DRMC] would not be effective if the public to which these newspapers cater was unable to have a sufficient number of publications, in various trends and characters; that ultimately the objective to reach is that readers who are among the key recipients of the freedom proclaimed by Article 11 of the 1789 [DRM] are able to exercise their freedom of choice without having either private or governments interests being substitute to their own decisions, nor that they can be made the object of a market."*[79]

The constitutional Council could not have forecast in 1984 the emergence of the web 2.0 and social media, as such technology now allows for an ever-expanding freedom of choice in the source of publication. Article 34 of the French Constitution provides that only laws can « *determine the rules concerning… civic rights and the fundamental guarantees granted to citizens for the exercise of their civil liberties; freedom, pluralism and the independence of the media*" and thus the government cannot issue decrees in the field of

[77] See for example, Conseil Constitutionnel (CC) (Constitutional Court) decision No.87-237 DC, Dec. 30, 1987, Rec. 63 (Fr.).
[78] Conseil constitutionnel [CC] [Constitutional Court] decision No. 2004-497 DC, July 1, 2004, §23, Rec.107 (Fr.)
[79] Conseil constitutionnel [CC] [Constitutional Court] decision No. 1984 DC, Oct. 11, 1984, §38, Rec.78 (Fr.)

freedom of expression and media. The Constitution has been modified several times, including in 2008 when the constitutional law (*loi constitutionnelle*) n°2008-724 of July 23, 2008, added a third paragraph to article 4 of the Constitution stating that "[*t*]*he law guarantees the pluralistic expression of opinions and the equitable participation of political parties and groups in the democratic life of the Nation*." The plurality and diversity of the marketplace of ideas is thus protected by the Constitution.

The French law protecting and limiting freedom of expression is the *Loi sur la liberté de la presse*, the Freedom of the Press Law of July 29, 1881 (French Press Law),[80] which has been modified many times since its enactment. It does not only regulate the movable type press and its current bits and bytes incarnations, but also "*speech made public*" and the "*free communication of thoughts and opinions"* as protected by article 11 of the DRMC. As such, the French Press Law generally protects freedom of any speech, as it does not only apply to the Press, but, much more broadly, to public speech including speech on social media, as its article 23 states that that, to be within the scope of the law, speech must have been made public by one of the means of publication, which it enumerates.

When the French constitutional Council decided in January 2016 that article 24 bis of the French Press Law incriminating denying crimes against humanity was constitutional, it took care to state that "*freedom of expression and communication is all the more precious since its exercise is a condition of democracy and one of the guarantees of respect for others*

[80] The Recommendation number 14 2015 report from the *Commission de réflexion sur le droit et les libertés à l'âge du numérique*, submitted to then President of the national Assembly Claude Bartolone, recommended to change the name of the French Press law to "Freedom of expression law" ("*loi sur la liberté d'expression*") in order to "*put an end to the widely held view that the scope of the* [*Press Law*] *is restricted to the press*."*Numériques et Libertés: Un nouvel âge démocratique. Rapport n° 3119,* http://www.assemblee-nationale.fr/14/pdf/rapports/r3119.pdf.

'rights and freedoms; … it follows that the interference with the exercise of this freedom must be necessary, appropriate and proportionate to the objective pursued.[81] Speech can nevertheless be limited, including online speech.

Article 1 of the September 30, 1986 French law on freedom of communication law[82] states that "[*c*]*ommunication to the public by electronic means is free*," adding, however, that:

> "[*t*]*he exercise of this freedom may be limited only to the extent required, on the one hand, by respect for the dignity of the person, the freedom and property of others, the pluralistic character of thoughts and opinions, and on the other hand by the protection of children and adolescents, the maintenance of public order, the needs of national defense, the requirements of public service, the technical constraints inherent to means of communication, as well as the need for the audiovisual services to develop audiovisual production*."

The law thus identifies two different set of values, which must both be protected: the protection of the persons and their property on one hand, and public order on the other hand. Defamation and insults are crimes in France, showing the high value in which protection of reputation is held there. Hate speech, an assault on human dignity, is also a threat to public order, including online. The *Commission Nationale Consultative des Droits*

---

[81] Conseil constitutionnel [CC] [Constitutional Court] decision No 2015-512 QPC, (Jan 8 2016).
[82] Loi n° 86-1067 du 30 septembre 1986 relative à la liberté de communication [Law 86-1067 of September 30, 1986 on freedom of communication, JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], Oct. 1, 1986, p. 11749, as modified by Loi n° 2004-669 du 9 juillet 2004 relative aux communications électroniques et aux services de communication audiovisuelle [Law n° 2004-669 of July 9, 2004 on electronic communication], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], July 10, 2004, p. 12483.

de l'Homme, (National Consultative Commission on Human Rights) wrote in its 2015 *Opinion on the fight against online hate speech* that it:

"*reiterate*[*d*] *its recommendation designed to encourage widespread reflection on the potential definition of a form of 'digital public order' based on the notion that the Internet must remain a platform for exercising freedoms where fundamental rights and liberties are respected and not a platform for impunity*."[83]

This first part of this article was meant to be a short general presentation on freedom of expression laws. This right is generally respected and protected in developed countries, yet it is the boundaries which are nevertheless placed to limit such right which are telling of the state of democracy and the concerns of a country.

2nd Part: The Four Main Justifications for Regulating Free Speech

Even if a country recognizes freedom of speech, it may also recognize some limitations to that right to protect equal or superior interests, whether online or "in real life." Indeed, "[*w*]*hat is illegal offline is also illegal online*."[84] However, online speech may be anonymous, or rather, pseudonymous, as platforms are likely to have information sufficient to learn the identity of a particular user. Most social media platforms make it easy to speak anonymously, and it should be saluted, as "*the ability to speak anonymously on the Internet*

[83] *Avis sur la lutte contre les discours de haine sur internet*, JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], July 10, 2015, n°125, https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000030862432. a*vailable* in English: http://www.cncdh.fr/sites/default/files/15.02.12_avis_lutte_discours_de_haine_internet_en.pdf (p.6).

[84] Tackling Illegal Content Online –Towards an enhanced responsibility of online platforms, Communication COM(2017) 555 of September 2017 from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, p.2, https://ec.europa.eu/transparency/regdoc/rep/1/2017/EN/COM-2017-555-F1-EN-MAIN-PART-1.PDF

promotes the robust exchange of ideas and allows individuals to express themselves freely without "fear of economic or official retaliation ... [*or*] *concern about social ostracism*.""[85]

The U.S. Supreme Court explained in *McIntyre v. Ohio Elections Comm'n* that *"the interest in having anonymous works enter the marketplace of ideas unquestionably outweighs any public interest in requiring disclosure as a condition of entry*."[86] This case was cited in 2017 by Twitter in its complaint against the U.S. Department of Homeland Security and other defendants, in the United States District Court of the Northern District of California, alleging that Defendants had no right to compel disclosure of the identity of the person(s) behind the pseudonymous @ALT_USCIS Twitter account.[87] This account was one of several created after the inauguration of President Donald Trump, on January 21, 2017, and which were presented as being "alternative" accounts of several U.S. agencies, such as the Department of Labor (@alt_labor), the federal Bureau of Land Management (@blm_alt), or the U.S. National Park Services (@AltNatParkSer), each posting messages critical of the then-new administration. Defendants issued and delivered a summons to Twitter in March 2017, demanding that the platform provide records associated with the @ALT_USCIS Twitter account, which could then be used to learn the identity of the user(s) behind the account. The summons cited 19 U.S.C. § 1509 as the sole legal basis of the summons, which allow the U.S. Customs to require production of any record which may be relevant in an

[85] In re Anonymous Online Speakers, 661 F. 3d 1168,1173 (9th Circ.2011), citing McIntyre v. Ohio Elections Comm'n, 514 U.S. 334, 341-42 (1995).
[86] McIntyre v. Ohio Elections Comm'n, 514 US 334, 342 (1995).
[87] Complaint, Twitter Inc. v. U.S. Department of Homeland Security, No. 3:17-cv-01916 (N.D. Calif. filed April 6, 2017). Defendants withdrew their summons and Twitter filed a notice of voluntary dismissal of the case under FRE 41 on April 7, 2017, one day after filing the original complaint.

"*investigation or inquiry conducted for the purpose of ascertaining the correctness of any entry, for determining the liability of any person for duty, fees and taxes due or duties, fees and taxes which may be due the United States, for determining liability for fines and penalties, or for insuring compliance with the laws of the United States administered by the United States Customs Service*."

Twitter stated in its complaint that defendants "*could not plausibly establish that they issued the… summons… in any investigation or inquiry relating to the import of merchandise*."[88] Twitter also argued that compelling the disclosure of the pseudonymous account would have a chilling effect on speech, and that a "*court generally permit an organization or business to assert* [*free speech*] *rights on behalf of its members or customers*," citing *Virginia v. American Booksellers Ass'n, Inc*. where the Supreme Court held that booksellers can assert the First Amendment rights of buyers of adult-oriented books, writing that "*in the First Amendment context*, "`[*I*]*itigants . . . are permitted to challenge a statute not because their own rights of free expression are violated, but because of a judicial prediction or assumption that the statute's very existence*."[89]

The Committee of Ministers of the Council of Europe adopted in May 2003 a Declaration on freedom of communication on the Internet, which principle 7 on Anonymity states that "*to ensure protection against online surveillance and to enhance the free expression of information and ideas, member states should respect the will of users of the*

[88] Paragraph 57 of the complaint.
[89] Virginia v. American Booksellers Assn., Inc., 484 US 383 (1988), citing Secretary of State of Maryland v. J. H. Munson Co., 467 U. S. 947, 956-957 (1984).

Internet not to disclose their identity."[90] The Grand Chamber of the European Court of Human Rights stated in *Delfi AS v. Estonia* that it "*is mindful of the interest of Internet users in not disclosing their identity. Anonymity has long been a means of avoiding reprisals or unwanted attention. As such, it is capable of promoting the free flow of ideas and information in an important manner, including, notably, on the Internet***."**[91] The Grand Chamber observed further "*that different degrees of anonymity are possible on the Internet*."[92] Indeed, remaining anonymous is a decision which can be made for several reasons, and "*may be motivated by fear of economic or official retaliation, by concern about social ostracism, or merely by a desire to preserve as much of one's privacy as possible*", as explained by the U.S. Supreme Court in *Watchtower Bible* and in *McIntyre*.[93]

It is sometimes necessary to speak anonymously or pseudonymously, as speaking may be illegal. How a country balances freedom of speech with other rights is very telling about how democratic this country really is. Not every speech is protected by law but separating the wheat from the chaff is done following different methods on either side of

---

[90] Declaration on freedom of communication on the Internet,(Adopted by the Committee of Ministers on 28 May 2003 at the 840th meeting of the Ministers' Deputies), available at https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016805dfbd5. Principle 7 specifies however that member states are not prevented from taking measures and co-operating to find criminals, in accordance with national law, the Convention for the Protection of Human Rights and Fundamental Freedoms and other international agreements in the fields of justice and the police.

[91] Delfi SA v. Estonia, Grand Chamber, paragraph 145.

[92] Delfi AS v. Estonia, Grand Chamber, paragraph 148 : " *An Internet user may be anonymous to the wider public while being identifiable by a service provider through an account or contact data that may be either unverified or subject to some kind of verification – ranging from limited verification… to secure authentication, be it by the use of national electronic identity cards or online banking authentication data allowing rather more secure identification of the user. A service provider may also allow an extensive degree of anonymity for its users, in which case the users are not required to identify themselves at all and they may only be traceable – to a limited extent – through the information retained by Internet access providers. The release of such information would usually require an injunction by the investigative or judicial authorities and would be subject to restrictive conditions. It may nevertheless be required in some cases in order to identify and prosecute perpetrators*."

[93] Watchtower Bible and Tract Soc'y of New York, Inc. v. Village of Stratton, 536 U.S. 150,166 (2002), citing McIntyre v. Ohio Elections Comm'n, 514 U. S., at 341-342. We will discuss anonymity online further on in this article.

the Atlantic. We will discuss four main reasons for the states to limit speech: (I) protecting public order, (II) protecting the reputation of others, (III) protecting the morality and (IV) advancing knowledge and truth.

I. Protecting Public Order

Governments may have different goals when limiting speech. Thy may protect the honor of the country and the leaders, when they enact laws protecting the flag (A) or making a crime to mock the King (or the President…) (B). They may the economic interests in mind when these laws are enacted to protect the market (C). Finally, limiting freedom of speech is seen as a way to protect the safety and security of people (D).

A. Protecting the Flag

Hong Kong legislature introduced a bill in October 2017 which would have made disrespecting the national anthem a crime carrying a three-year prison term. This bill was triggered by football fans booing the Chinese anthem played in the beginning of the games.[94] On June 4, 2020, the day of the 31st anniversary of Tiananmen, the Hong Kong's legislature passed a law making disrespecting China's national anthem a crime punishable by up to three years in prison.[95]

In the U.S., then-President-Elect Donald Trump tweeted on November 29, 2016: "*Nobody should be allowed to burn the American flag - if they do, there must be consequences*

[94] *See Chinese citizens who 'disrespect' national anthem may face up to 3 years in prison*, HONG KONG FREE PRESS, https://www.hongkongfp.com/2017/10/31/chinese-citizens-disrespect-national-anthem-may-face-3-years-prison/

[95] Austin Ramzy, Tiffany May and Javier C. Hernández, *On Tiananmen Anniversary, Hong Kong Makes Mocking China's Anthem a Crime,* THE NEW YORK TIMES, (June 4, 2020, 8:54 a.m. ET). https://www.nytimes.com/2020/06/04/world/asia/tiananmen-hong-kong-china.html.

- *perhaps loss of citizenship or year in jail!*"[96] During a reelection rally in June 2020, President Trump called for making burning the flag a crime punishable by one-year imprisonment.[97] While Section 349 of the Immigration and Nationality Act enumerates all the instances where a natural born U.S. citizen or a naturalized U.S. citizen can lose his or her U.S. nationality, the list does not include 'burning the flag', which is not surprising, as burning the flag is speech protected by the First Amendment, and is not a criminal offense.

Indeed, the Supreme Court held in *Texas v. Johnson*[98] that burning the flag is expressive conduct protected by the First Amendment and noted that attaching a peace sign to the flag,[99] refusing to salute the flag,[100] and even treating flag "contemptuously" by wearing trousers having a small cloth flag sewn on their seat[101] are all expressive conducts. In the *Texas v, Johnson* case, Mr. Johnson had burned an American flag in front of the Dallas City Hall while participating in protests taking place during the 1984 Republican National Convention. He was charged and found guilty of intentionally desecrating the flag, in violation of Tex. Penal Code Ann. § 42.09(a)(3).[102] The Texas Penal Code defined at the time

---

[96] @realDonaldTrump, Twitter (Nov. 29, 2016, 6:55 am), https://twitter.com/realDonaldTrump/status/803567993036754944.
[97] Andrew Solender, *Trump Says He Wants To Punish Flag Burning With A Year In Prison*, FORBES, (Jun 20, 2020,09:20pm EDT), https://www.forbes.com/sites/andrewsolender/2020/06/20/trump-says-he-wants-to-punish-flag-burning-with-a-year-in-prison/#399509f34046.
[98] Texas v. Johnson, 491 U.S. 397, (1989).
[99] Spence v. Washington, 418 US 405, 409-410 (1974).
[100] West Virginia Bd. of Ed. v. Barnette, 319 US 624, 632 (1943).
[101] Smith v. Goguen, 415 U. S. 566, 588 (1974) (WHITE, J., concurring in judgment).
[102] The Texas Penal Code Ann. § 42.09 (1989) provided:
"Desecration of Venerated Object
"(a) A person commits an offense if he intentionally or knowingly desecrates:
"(1) a public monument;
"(2) a place of worship or burial; or
"(3) a state or national flag.
"(b) For purposes of this section, `desecrate' means deface, damage, or otherwise physically mistreat in a way that the actor knows will seriously offend one or more persons likely to observe or discover his action.
"(c) An offense under this section is a Class A misdemeanor."

desecrating a flag as to "*deface, damage, or otherwise physically mistreat in a way that the actor knows will seriously offend one or more persons likely to observe or discover his action.*"[103] Several witnesses testified that they indeed had been offended by the flag-burning. Mr. Johnson was convicted and sentenced to one year in prison and a $2,000 fine.

Mr. Johnson argued, on appeal and at the Supreme Court, that the Texas statute violated the Constitution, as burning the flag is a political protest and thus protected speech. The Supreme Court agreed, noting that Mr. Johnson "*was not… prosecuted for the expression of just any idea; he was prosecuted for his expression of dissatisfaction with the policies of this country, expression situated at the core of our First Amendment value*."[104] The Supreme Court found that Texas had no interest than suppression of speech when passing its flag desecration statute, adding in a footnote that the Statute was thus not narrowly drawn.[105] Therefore, the prosecution of a tired person dragging the flag in mud, without any intention to desecrate the flag, and thus not engaging in an expressive conduct, "*would pose a different case, and … this case may be disposed of on narrower grounds*." As such, the Court only addressed the claim that the Texas Statute "*as applied to political expression like his violates the First Amendment*." The Court held it did, and that Johnson had engaged in protected expressive conduct. While the State may have an interest to prevent breaches of the peace, "*Johnson's conduct did not threaten to disturb the peace. Nor does the State's interest in preserving the flag as a symbol of nationhood and national unity justify his criminal conviction for engaging in political expression*."[106] The Supreme Court concluded

[103] Tex. Penal Code Ann. § 42.09 (1989).
[104] Johnson v. Texas, at 411.
[105] Johnson v. Texas, footnote 3.
[106] Johnson v. Texas, at 420A.

that its decision was "*a reaffirmation of the principles of freedom and inclusiveness that the flag best reflects, and of the conviction that our toleration of criticism… is a sign and source of our strength*."

Other countries do not share this position. It is illegal to burn the German flag, and a bill was introduced in 2019 to extend the law to foreign flags, including the flag of the European Union.[107] The law was triggered by the burning of Israeli flags in Berlin in 2017.[108] Article L 433-5-1 of the French penal Code makes it a crime, punished by 7,500 euros fine, to publicly insult the national anthem or the national flag during an event organized or regulated by the public authorities. It can be argued that this "*public outrage"* can be done by publishing an insulting social media post, streaming the burning of the flag, or posting a video or a picture of it after the event. It is also illegal, under article R. 645-15 of the French criminal Code, to destroy, deteriorate, or use the national flag in a degrading manner, in a public place or a place open to the public, if done "*under conditions likely to disturb public order and with the intention of insulting the flag*."[109] If such act is committed in a private place, it is a crime only if broadcasted, which would include streaming it live on social media. Article R. 645-15 of the French criminal Code was created by decree in July 2010 after an award was given to a photograph showing using the French flag to clean his buttocks.[110] The legality of the decree was challenged by the French League of Human

[107] Elliot Douglas, *German lawmakers push to ban flag burning*, DW.COM, (Jan, 15, 2020), https://www.dw.com/en/german-lawmakers-push-to-ban-flag-burning/a-52016871.
[108] Michaela Küfner, *Israeli flag burning prompts German Foreign Minister Sigmar Gabriel to back outlawing it,* DW.COM, (Dec. 15, 2017), https://www.dw.com/en/german-lawmakers-push-to-ban-flag-burning/a-52016871
[109] Art. R. 645-15 of the French criminal Code (Decree. no 2010-835 July 21, 2010).
[110] See *Outrage sparked by French flag bottom-wiping photo*, FRANCE 24, (April 21, 2010, 19:18), https://www.france24.com/en/20100421-controversial-French-flag-photo-stirs-passions-nice.

Rights, which claimed that the State has exceeded its powers and that the decree had violated the French Declaration of Human Rights, which has the same rank than the Constitution. The Council of State (*Conseil d'État*), acting as the highest administrative court, held however that the decree did not violate the Declaration of Human Rights,[111] reasoning that the decree only incriminates "*physical or symbolic degradations of the flag likely to cause serious disturbances to public peace and security and committed with the sole intention of destroying, damaging or degrading the flag.*" The decree does not make expressing political or philosophical ideas illegal, nor does it prevent artistic performances except if "*this mode of expression cannot, under the supervision of the criminal judge, be regarded as a work of the mind.*" But this means that criminal Courts are thus given the power to determine if a particular speech is indeed a "work of the mind," which is an intellectual property concept. The Council of State concluded that:

> "*in spite of the generality of the definition of the incriminated acts, the decree attacked does not carry, … an excessive attack on the freedom of expression guaranteed by the Declaration of Human and Citizen Rights and the European Convention for the Protection of Human Rights and Fundamental Freedoms*."

B. Mocking the King

Section 66A of the Information Technology Act 2008 of India was used for arresting a man in 2012 for having posted on Facebook a comment criticizing India Prime Minister Narendra Modi, and again in 2014 , as a man who allegedly published abusive social media

[111] CE Sect. July 19, 2011, n° 343430, available at https://www.legifrance.gouv.fr/affichJuriAdmin.do?idTexte=CETATEXT000024390173.

posts about India Independence Day was arrested and tried.[112] Section 112 of the Thai Criminal Code states that "[*w*]*hoever defames, insults or threatens the King, Queen, the Heir apparent or the Regent, shall be punished with imprisonment of three to 15 years*"[113] and the current Thai Constitution, promulgated in 2007, states in its Section 8 that "[*t*]*he King shall be enthroned in a position of revered worship and shall not be violated. No person shall expose the King to any sort of accusation or action."*[114] The 2007 Computer Crime Act (CCA) is also used to prosecute individuals who use the web to criticize the Thai monarchy.[115] For instance, Thai blogger Suwicha Thakhor was sentenced to ten years in jail in 2009 under the CCA for having posted online doctored images of the Thai royal family. He was pardoned by the King in 2010.[116] In 2015, a man was sentenced to thirty years in jail for having insulted the monarchy on Facebook.[117] The same year, Thanakorn Siripaiboon was charged under the Thai *lèse-majesté* law for allegedly posting online sarcastic comments about… the King's dog.[118] Pongsak Sriboonpeng was sentenced to 30 years in jail for six Facebook posts, and a woman was sentenced to 28 years in jail for seven Facebook posts

---

[112] *India: End use of archaic sedition law to curb freedom of expression*, AMNESTY INTERNATIONAL (Sept. 2, 2014), https://www.amnesty.org/en/latest/news/2014/09/india-end-use-archaic-sedition-law-curb-freedom-expression.

[113] The Thai Criminal Code is available http://www.ilo.org/dyn/natlex/natlex4.detail?p_lang=en&p_isn=82844 (last visited Dec. 30, 2020).

[114] https://www.constituteproject.org/constitution/Thailand_2007.pdf

[115] https://www.article19.org/data/files/medialibrary/1739/11-03-14-UPR-thailand.pdf p 3 and 4.

[116] Reporters Without Borders had asked in December 2009 King Bhumibol Adulyadej to pardon "*Internet users who are in jail or who are being prosecuted in connection with the dissident views they allegedly expressed online,"* see *King asked to pardon Internet users prosecuted on lese majesté or national security charges,* https://rsf.org/en/news/king-asked-pardon-internet-users-prosecuted-lese-majeste-or-national-security-charges

[117] *See* Agence France-Presse in Bangkok*. Man jailed for 30 years in Thailand for insulting the monarchy on Facebook*, THE GUARDIAN, (Fri 7 Aug 2015 04.45 EDT) https://www.theguardian.com/world/2015/aug/07/man-jailed-for-30-years-in-thailand-for-insulting-the-monarchy-on-facebook.

[118] Thomas Fuller, *Thai Man May Go to Prison for Insulting King's Dog*, THE NEW YORK TIMES, Dec. 14, 2015), http://www.nytimes.com/2015/12/15/world/asia/thailand-lese-majeste-tongdaeng.html.

which insulted the royal family.[119] Frank La Rue, then the United Nations Special Rapporteur on the right to freedom of opinion and expression, urged Thailand in October 2011 to amend its *lèse majesté* law[120] and:

> *"to hold broad-based public consultations to amend section 112 of the penal code and the 2007 Computer Crimes Act so that they are in conformity with the country's international human rights obligations*."
>
> Mr. La Rue added that "[*t*]*he recent spike in lèse majesté cases pursued by the police and the courts shows the urgency to amend them*."[121]

In April 2020, Taiwan refused to deport immigrant worker from the Philippines who had allegedly insulted President Rodrigo Duterte on her Facebook page.[122] In September 2020, a Turkish journalist who had published a tweet making fun of an historical drama series was jailed for having insulted Ertuğrul Ghazi, a sultan who died around 1280.[123]

Even developed countries, such as the U.S. (a) or European countries (b), have laws limiting what can be said about their Presidents.

---

[119] *Thailand: 2 Punished for Insulting King*, THE NEW YORK TIMES, (Aug. 8, 2015), p. A5.
[120] *Thailand / Freedom of expression: UN expert recommends amendment of lèse majesté laws*, THE UNITED NATIONS, (Oct. 10, 2011), http://newsarchive.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=11478&LangID=E#sthash.sBEmGagh.dpuf (last visited Dec. 30, 2020).
[121] Ibid.
[122] Mong Palatino, *Taiwan refuses to deport caregiver who 'insulted' Philippine president on Facebook*, (May 7, 2020 15:40 GMT), https://globalvoices.org/2020/05/07/taiwan-refuses-to-deport-caregiver-who-insulted-philippine-president-on-facebook/.
[123] *Turkish journalist Oktay Candemir charged with 'insulting' deceased sultan in satirical tweet*, COMMITTEE TO PROTECT JOURNALISTS, (Sept. 8, 2020 1:42 PM EDT), https://cpj.org/2020/09/turkish-journalist-oktay-candemir-charged-with-insulting-deceased-sultan-in-satirical-tweet.

a. U.S. Laws

18 U.S.C. §871(a) makes it a crime, punishable by a fine and up to five years in jail, to:

> "*knowingly and willfully deposits for conveyance in the mail or for a delivery from any post office or by any letter carrier any letter, paper, writing, print, missive, or document containing any threat to take the life of, to kidnap, or to inflict bodily harm upon the President of the United States, the President-elect, the Vice President or other officer next in the order of succession to the office of President of the United States, or the Vice President-elect, or knowingly and willfully otherwise makes any such threat against the President, President-elect, Vice President or other officer next in the order of succession to the office of President, or Vice President-elect."*[124]

The Supreme Court upheld the constitutionality of this federal law in 1969, explaining in *Watts v. U.S.* that "[*t*]*he Nation undoubtedly has a valid, even an overwhelming, interest in protecting the safety of its Chief Executive and in allowing him to perform his duties without interference from threats of physical violence."*[125] In this case, Mr. Watts had said during a public meeting: "*They always holler at us to get an education. And now I have already received my draft classification as 1-A and I have got to report for my physical this Monday coming. I am not going. If they ever make me carry a rifle the first man I want to get in my sights is L. B. J.*" He was convicted for having violated a 1917 federal law prohibiting anyone to "*knowingly and willfully*… [*making*] *any threat to take the life of or to inflict bodily*

[124] 18 U.S. C. § 871 - Threats against President and successors to the Presidency.
[125] Watts v. U.S., 394 U.S. 705,707 (1969).

harm upon the President of the United States." The U.S. Court of Appeals for the District of Columbia Circuit affirmed, but the Supreme Court reversed.[126]

Justice Douglas explained in his concurring opinion in *Watts*[127] that the 1917 law[128] on which the current statute is based "*traces its ancestry to the Statute of Treasons (25 Edw. 3) which made it a crime to "compass or imagine the Death of . . . the King.*" The Court found this statute "*constitutional on its face*, [*as*] [*t*]*he Nation undoubtedly has a valid, even an overwhelming, interest in protecting the safety of its Chief Executive and in allowing him to perform his duties without interference from threats of physical violence*."[129] While the federal law was constitutional on its face, as "*the Nation undoubtedly has a valid, even an overwhelming, interest in protecting the safety of its Chief Executive,*" the Supreme Court nevertheless noted that "[*w*]*hat is a threat must be distinguished from what is constitutionally protected speech*."[130] The Court explained that the statute requires the Government to prove a true "threat." But that it did "*not believe that the kind of political hyperbole indulged in by petitioner*" was such a threat, even though it agreed with Petitioner that it was "a *kind of very crude offensive method of stating a political opposition to the President*."[131]

The Supreme Court articulated in *Watts* a three-factor test to find out whether a threat against the President is protected speech: (1) the context is a political speech; (2) the

---

[126] Watts v. United States, 394 US 705 (1969.
[127] Watts v. United States, 394 US 705,709 (1969)
[128] H. R. 15314, Feb. 14, 1914, punishing any individual who would knowingly and willfully threaten to kill the President of the U.S. or to inflict him bodily harm, see https://uscode.house.gov/statviewer.htm?volume=39&page=919.
[129] Watts v. United States, at 707, referring to See H. R. Rep. No. 652, 64th Cong., 1st Sess. (1916).
[130] Watts, at 707.
[131] Watts, at 708.

statement was "expressly conditional"; and (3) "the reaction of the listeners" who, in the facts leading to *Watts*, had "*laughed after the statement was made*." The third part of the *Watts* test seems to require that the recipient of the message, not the sender, defines the line between true threat and political hyperbole. In *Watts*, the speaker had threatened Lyndon Johnson at an anti-war rally, which, presumably was attended by members of the public who shared Watt's opinions. On social media, members of the public, who read a particular message, may have more divided opinions.

Justice Antonin Scalia noted in dicta *in R.A.V. v. St. Paul* that "*the federal Government can criminalize only those threats of violence that are directed against the President, see 18 U. S. C. § 871—since the reasons why threats of violence are outside the First Amendment* (*protecting individuals from the fear of violence, from the disruption that fear engenders, and from the possibility that the threatened violence will occur*) *have special force when applied to the person of the President*."[132] Justice Scalia had written five years earlier in an article:

> "[*i]t is generally agreed… that what might be called "political speech"- the expression of views on matters that are, or would be, the subject of governmental action- is entitled to the highest degree of protection from official interference."*[133]

He also noted that the "*clear and present danger*" test was originally applied to political speech, and as such, only imminent threat of harm could bar its expression.

---

[132] R.A.V. v. St. Paul, RAV v. St. Paul, 505 US 377, 388 (1992).
[133] Antonin Scalia, *A House with Many Mansions: Categories of Speech Under the First Amendment*, in James Brewer Stewart, THE CONSTITUTION, THE LAW, AND FREEDOM OF EXPRESSION 1787-1987, P. 1, Southern Illinois University Press, 1987.

18 U.S.C. §871(a) also applies to social media. President Obama joined Twitter on May 18, 2015, in his President of the United States capacity,[134] under the handle @POTUS, and users took then the opportunity to send him messages, many of them inappropriate. A man was sentenced to prison for six months for in 2013 after having posted on Twitter several threats against President Obama[135] One of them read "*Well Ima Assassinate president Obama this evening !... Gotta get this monkey off my chest while he's in town* -_- ". His Twitter handle was @DestroyLeague_D. Another Twitter user tweeted in the Summer of 2012 "Let's Go Kill the President." Interrogated by FBI agents, he said he was drunk and apologized. After he tweeted a series of threat against the President a few weeks later, he was arrested and sentenced to one year in federal prison.[136] After comedian Kathy Griffin posed in 2017 with a plastic head of Donald Trump covered in ketchup, making it appear she was holding the severed head of the President, she was allegedly investigated by the secret services,[137] and placed on the no-fly list.[138]

b. European Countries

Article 36 of the French Press Law used to incriminate insulting a foreign head of State. However, the European Court of Human Rights found in 2002, in *Colombani v.*

---

[134] Barack Obama has his own, private Twitter account, @BarackObama, since March 2007 and still uses it.
[135] Alex Fitzpatrick, *Man Jailed After Threatening President Obama on Twitter*, MASHABLE, (May 29, 2013), https://mashable.com/2013/05/29/obama-threats-twitter.
[136] Robbie Brown*, 140 Characters Spell Charges and Jail*, THE NEW YORK TIMES (July 2, 2013), https://www.nytimes.com/2013/07/03/us/felony-counts-and-jail-in-140-characters.html.
[137] Sopan Deb, *Kathy Griffin Is Being Investigated by the Secret Service, Her Lawyers Say*, THE NEW YORK TIMES (June 2, 2017), https://www.nytimes.com/2017/06/02/arts/television/kathy-griffin-donald-trump-news-conference.html.
[138] Sam Sanders, *Kathy Griffin: Life After The Trump Severed Head Controversy*, NPR MORNING EDITION (April 23, 2019 5:09 AM ET), https://www.npr.org/2019/04/23/716258113/kathy-griffin-life-after-the-trump-severed-head-controversy.

France,[139] that article 36 violated article 10 of the European Convention on Human Rights, and it was abrogated by law on March 9, 2004.

On March 31, 2016, German comedian Jan Böhmermann made fun of Turkish President Recep Tayyip Erdogan on his television show, "*Neo Magazin Royal*", shown on public channel ZDF. He recited a poem implying that Erdogan has been engaged in child pornography and bestiality. Germany has a long tradition of satiric comedy, but Section 103 of the German Criminal Code incriminates defamation of organs and representatives of foreign states. It states that:

> "*(1) Whosoever insults a foreign head of state, or, with respect to his position, a member of a foreign government who is in Germany in his official capacity, or a head of a foreign diplomatic mission who is accredited in the Federal territory shall be liable to imprisonment not exceeding three years or a fine, in case of a slanderous insult to imprisonment from three months to five years.*
>
> *(2) If the offence was committed publicly, in a meeting or through the dissemination of written materials (section 11(3)) Section 200 shall apply. An application for publication of the conviction may also be filed by the prosecution service."*[140]

The Mainz public prosecutor received a complaint from the Turkish President and thus had to start investigating the issue, but the charges were dropped after a 6-months

[139] Columbani v. France, Case No. 51279/99 (June 25, 2002), available at https://hudoc.echr.coe.int/tur#{%22itemid%22:[%22001-60532%22]}.
[140] The translation is from https://ec.europa.eu/anti-trafficking/sites/antitrafficking/files/criminal_code_germany_en_1.pdf

investigation.[141] Chancellor Angela Markel had called the poem "deliberately offensive" during a phone call with Turkish Prime Minister Ahmet Davutoglu, but later expressed regrets for having characterized it as such.[142] The German law nevertheless made the Mainz prosecutors act as if they were Turkish prosecutors, who themselves had prosecuted a former Turkish beauty queen who had published on her Instagram account a poem criticizing President Erdogan.[143]

Article 26 of the French Press Law used to incriminate insulting the President of the Republic, but did not define what constituted such offense, merely stating that insulting the President of the Republic by one of the means set forth in Article 23, that is, " *by uttering speeches, shouts or threats in a public place or public meeting, or by writing, printed text, drawing, engraving, painting, emblem, image, or using other written, spoken or image sold or distributed, offered for sale or exhibited in a public place or public meeting, or by a poster or notice displayed in a place where it can be seen by the public*," was an offense punishable by 45,000 Euro fine. While this fine is quite hefty when compared to France's average salary, it was not comparable to the punishment faced by individuals found guilty of *lèse-majesté* during France's *Ancien Régime*, which carried then a death penalty sentence. The French courts interpret article 26 as meaning "*any insulting or disparaging expression or any defamatory insinuation which is liable to undermine the President's honor or dignity either in*

---

[141] Susanne Spröer, *Opinion: Böhmermann vs. Erdogan: Criminal satire? Not in Germany*, DW, (Oct. 5, 2016), https://www.dw.com/en/opinion-b%C3%B6hmermann-vs-erdogan-criminal-satire-not-in-germany/a-35960062.

[142] *Merkel admits 'mistake' in Böhmermann satire case*, DW, https://www.dw.com/en/merkel-admits-mistake-in-b%C3%B6hmermann-satire-case/a-19209284.

[143] Glen Johnson, *Former beauty queen convicted of insulting Turkey's president by sharing poem on social media*, LOS ANGELES TIMES, (May 31, 2016, 3:18 PM), https://www.latimes.com/world/middleeast/la-fg-turkey-beauty-queen-20160531-snap-story.html.

the performance of his presidential duties or in his private life, or in his public life prior to being elected."[144] Article 26 was truly a crime of *Lèse Majesté*, as it was not possible to use truth as a defense (*exceptio veritatis*) for this crime: the head of the State was infallible.

Article 26 was, however, abolished by law in 2013,[145] after the ECtHR held on March 14, 2013, in *Éon v. France*, that France had violated article 10 of the ECHR when convicting under article 26 Hervé Éon, a French citizen who had insulted then-President Nicolas Sarkozy.[146] Mr. Éon had been sentenced in August 2008 to a 30 Euros fine for having held a placard, which read "*casse toi pauvre con*" ("*get lost you stupid cunt*") on the side of a road where Nicolas Sarkozy, then President of France, was scheduled to pass by. The sentence was upheld on appeal, and the French civil Supreme Court, the *Cour de cassation*, refused to examine the case. Mr. Éon then applied to the ECtHR, which held on that France had violated article 10 of the ECHR protecting the right to freedom of expression. Article 10§2 of the ECHR does allow a public authority to interfere with freedom of expression, but only if (1) such interference is prescribed by law, (2) is necessary in a democratic society, and (3) pursues one of the legitimate goals enumerated by article 10(2), among them *"the prevention of disorder or crime* [*and*] *the protection of the reputation or rights of others.*" The French government had argued that Mr. Éon's conviction had been "*necessary in a democratic society for the prevention of disorder, given the need to protect the institutional*

[144] Cour de cassation [Cass.] [supreme court for judicial matters] crim., Dec.21, 1966, Bull. crim. No 302 (Fr).

[145] Loi 2013-711 du 5 août 2013 portant diverses dispositions d'adaptation dans le domaine de la justice en application du droit de l'Union européenne et des engagements internationaux de la France [Law 2013-711 of August 5, 2013 implementing various adaptation in the field of justice pursuant to the law of the European Union and international commitments of France], art. 21, JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF France],

[146] Eon v. France, App. No. 2618/10, (Eur. Ct. H.R. marc. 14, 2013).

representative embodying one of the highest State authorities from verbal and physical attacks liable to undermine the State institutions themselves."[147] The ECHR was not convinced, noting, as it had done in previous cases, that:

> "*here is little scope under Article 10 § 2 for restrictions on freedom of expression in the area of political speech or debate – where freedom of expression is of the utmost importance – or in matters of public interest. The limits of acceptable criticism are wider as regards a politician as such than as regards a private individual. Unlike the latter, the former inevitably and knowingly lays himself open to close scrutiny of his every word and deed by both journalists and the public at large, and he must consequently display a greater degree of tolerance*.[148]

It should be noted that the head of the State may protect his or her reputation by other means, even though *lèse-majesté* is no longer a crime in France. In January 2019, the French *Office central de la lutte contre la criminalité informatique* (*central Office for the fight against computer crime*), which depends from the Minister of Police, required Google + to take down a photograph of General Pinochet surrounded by military officers in uniforms, where Pinochet's face had been replaced by the face of French President Emmanuel Macron, while then-Prime Minister Édouard Philippe, and then-Minister of Police Christophe Castaner, played the role of the military officers.[149] The montage had been posted on Google + under the title "*La dictature en marche,*" further mocking the President

---

[147] Eon, §41.

[148] Eon, §59, citing Lingens v. Austria, §42; Vides Aizsardzības Klubs v. Latvia, §40.

[149] Marc Rees, *Quand l'Office de lutte contre la cybercriminalité exige le retrait d'un photomontage visant Macron*, NEXTINPACT, (Jan.29, 2019, 5:47), https://www.nextinpact.com/article/29130/107547-quand-office-lutte-contre-cybercriminalite-exige-retrait-dun-photomontage-visant-macron.

by alluding to the name of the political movement he created "*La République en marche.*" The Office asked for the satirical image to be taken down, which is surprising, to say the least, as posting such picture online is no crime in France, but rather, protected political speech.

C. Protecting the Market

On September 27, 2018, the U.S. Securities and Exchange Commission (SEC) filed a civil complaint against Elon Musk, Chief Executive Officer of Tesla, following a series of tweets sent in August 2018.[150] On August 7, 2018, Musk posted on Twitter: ""*Am considering taking Tesla private at $420. Funding secured*," followed by more tweets about this proposed move,[151] which could be interpreted by his then-over 22million followers[152] as making Tesla a private company. Indeed, Tesla's stock price increased by more than 6% that day, closing up 10.98% from the previous day.[153] As noted in the SEC complaint, the purchase price quoted in the tweet "*reflected a substantial premium over Tesla stock's then-current share price*." The tweet was published at 12:48 p.m. EDT on August 7, 2018, during trading tie, as noted by the complaint, and was followed by more tweets:

- "*My hope is *all* current investors remain with Tesla even if we're private. Would create special purpose fund enabling anyone to stay with Tesla.*"

---

[150] U.S. Securities and Exchange Commission v. Elon Musk, 1:18-cv-08865 (S.D.N.Y). https://www.sec.gov/litigation/complaints/2018/comp-pr2018-219.pdf.

[151] "My hope is *all* current investors remain with Tesla even if we're private. Would create special purpose fund enabling anyone to stay with Tesla." "Shareholders could either to [sic] sell at 420 or hold shares & go private." "Investor support is confirmed. Only reason why this is not certain is that it's contingent on a shareholder vote."

[152] There are now more than 37 million to follow @elonmusk.

[153] Nasdaq halted trading in Telsa (TSLA) shares around 2:08 PM EDT, then resumed trading around 3:45 PM EDT. The stock price closed at $379.57, which is "*up over 6% from the time Musk first tweeted about taking Tesla private earlier that day*", see U.S. Securities and Exchange Commission v. Telsa, Inc., 1:18-cv-08947 (S.D.N.Y).

- "*Shareholders could either to* [sic] *sell at 420 or hold shares & go private*."
- "*Investor support is confirmed. Only reason why this is not certain is that it's contingent on a shareholder* vote."

The SEC called these tweets "*false and misleading*," as Elon Musk knew at the time that "*he had never discussed a going-private transaction at $420 per share with any potential funding source, had done nothing to investigate whether it would be possible for all current investors to remain with Tesla as a private company via a "special purpose fund," and had not confirmed support of Tesla's investors for a potential going-private transaction."* Section 10(b) of the Exchange Act[154] prohibits material misrepresentations and misleading omissions in connection with the purchase or sale of securities.

On September 29, 2018, the SEC filed a civil complaint against Telsa, Inc., alleging that the company had failed *"to implement disclosure controls or procedures to assess whether information disseminated by its Chief Executive Officer, Elon Musk, via his Twitter account was required to be disclosed in reports*," pursuant to the Exchange Act, within the time periods specified in the Commission's rules and forms.[155] Indeed, Testa had publicly filed a Form 8-K with the SEC in November 2013, stating that company would use Elon Musk's Twitter account to announce "*material information to the public about Tesla and its products and services and has encouraged investors to review the information about Tesla published by Musk via his Twitter account*."[156] The SEC alleged that failure to implement

[154] 15 U.S. Code § 78j.
[155] U.S. Securities and Exchange Commission v. Telsa, Inc., 1:18-cv-08947 (S.D.N.Y), https://www.sec.gov/litigation/complaints/2018/comp-pr2018-226.pdf.
[156] This for mis available at https://www.sec.gov/Archives/edgar/data/1318605/000119312513427630/d622890d8k.htm.

such controls and procedures had violated Rule 13a-15 of the Securities Exchange Act of 1934 (Exchange Act)[157] which requires public companies to maintain disclosure controls and procedures, defined as controls and other procedures "*designed to ensure that information required to be disclosed by the issuer in the reports that it files or submits under the Act… is recorded, processed, summarized and reported, within the time periods specified in the Commission's rules and forms.*"[158]

Both Elon Musk and Tesla agreed to settle the charges[159] and the settlement was approved by the Southern District of New York on October 16, 2018.[160] Elon Musk and Tesla agreed to pay a separate $20 million penalty, to be distributed under a court-approved process to harmed investors. Elon Musk agreed to step down as Tesla's Chairman and to be replaced by an independent Chairman, and Tesla agreed to appoint two new independent directors to its board, to establish a new committee of independent directors, and to "*put in place additional controls and procedures to oversee Musk's communications*."[161] However, in an interview on December 9, 2018,[162] journalist Lesley Stahl asked Elon Musk "*So your tweets are not supervised?"* He answered that "[*t*]*he only tweets that would have to be say reviewed would be if a tweet had a probability of causing a*

---

[157] 17 C.F.R. § 240.13a-15.
[158] 17 C.F.R. § 240.13a-15 (e).
[159] *Elon Musk Settles SEC Fraud Charges; Tesla Charged With and Resolves Securities Law Charges*, U.S. Securities and Exchange Commission, (Sept. 29, 2018), https://www.sec.gov/news/press-release/2018-226.
[160] Consent Motion for Entry of Final Judgment, US Securities and Exchange Commission v. Musk, Inc., No. 1:18-CV-8865 (S.D.N.Y. Oct. 16, 2018.), Entry of Final Judgment, US Securities and Exchange Commission v. Tesla, Inc., No. 1:18-CV-08947 (S.D.N.Y. Oct. 16, 2018.)
[161] The Final Judgment ordered Elon Musk, inter alia, to " *comply with all mandatory procedures implemented by Tesla, Inc. (the "Company") regarding (i) the oversight of communications relating to the Company made in any format, including, but not limited to, posts on social media (e.g., Twitter), the Company's website (e.g., the Company's blog), press releases, and investor calls, and (ii) the pre-approval of any such written communications that contain, or reasonably could contain, information material to the Company or its shareholders*."
[162] Lesley Stahl, *Tesla CEO Elon Musk: The 60 Minutes Interview*, CBS NEWS, (Dec. 9, 2018), https://www.cbsnews.com/news/tesla-ceo-elon-musk-the-2018-60-minutes-interview/.

movement in the stock." Lesley Stahl pressed on, "*And that's it?"*, to which Musk replied "*Yeah, I mean otherwise it's, "Hello, First Amendment." Like Freedom of Speech is fundamental*."

This statement reflects the general understanding of the public of the First Amendment as a law so powerful it knows no limits. However, the First Amendment has limits, and the capital market takes precedence over the marketplace of ideas when securities and disclosure laws must be implemented and enforced. The First Amendment does not always protect false statements of fact[163], and Rule 10b-5 of the Exchange Act prohibits making misleading or untrue statement of material facts, or omissions of material facts "*necessary in order to make the statements made, in the light of the circumstances under which they were made, not misleading*." This is the reason why many Twitter users add to their profile a statement such as *"This is a personal account, my views do not reflect the view of my employer*." The user may have chosen to publish the statement, but often it is done at the behest of the employer's social media policy, requiring the use of such disclaimer.

Broker-dealer members of the Financial Industry Regulatory Authority (FINRA) follow their own social media guidelines, FINRA Regulatory Notice 10-06, Guidance on Blogs and Social Networking Web Sites, issued in January 2010,[164] and Regulatory Notice 11-39,

---

[163] *See* for example Eugene Volokh, *Amicus Curiae Brief: Boundaries of the First Amendment's False Statements of Fact Exception*, 6 Stan. J. C.R. & C.L. 343, 348 (2010), explaining that "[*t*]*he boundaries of the "false statements of fact "exception to First Amendment protection are not well-defined*… but some false statements of fact are immune from liability, even if they are knowingly false", giving as example false statements about the Holocaust of global warming. The U.S. Supreme Court noted in *Gertz v. Robert Welch* that "[*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 not on the conscience of judges and juries but on the competition of other ideas. But there is no constitutional value in false statements of fact*." Gertz v. Robert Welch, Inc., 418 US 323, 339-340, (1974).

[164] FINRA Regulatory Notice 10-06 Guidance on Blogs and Social Networking Web Sites, available at https://www.finra.org/rules-guidance/notices/10-06.

Guidance on Social Networking Websites and Business Communications,[165] published in August 2011. Regulatory Notice 10-06 specifies that firms must supervise interactive electronic communications by the firm, or its registered representatives, on blog and social media sites, as the content provisions of FINRA's communications rules apply to interactive electronic communications sent this way by the firm. However, "*prior principal approval is not required under Rule 2210 for interactive electronic forums* [*on communications with the public*][166] [*but*] *firms must supervise these interactive electronic communications under NASD Rule 3010[167] in a manner reasonably designed to ensure that they do not violate the content requirements of FINRA's communications rules.*" Regulatory Notice 10-06 clearly states that firms must have a social media policy "*reasonably designed to ensure that their associated persons who participate in social media sites for business purposes are appropriately supervised, have the necessary training and background to engage in such activities, and do not present undue risks to investors*." They also must have a policy "*prohibiting any associated person from engaging in business communications in a social media site that is not subject to the firm's supervision* [*and*]*also must require that only those associated persons who have received appropriate training on the firm's policies and procedures regarding interactive electronic communications may engage in such communications*." Regulatory Notice 11-39 specified that a firm or an associated person could not "*sponsor a social media site or use a communication device that includes technology which automatically erases or*

[165] FINRA Regulatory Notice 11-39 Guidance on Social Networking Websites and Business Communications, available at https://www.finra.org/rules-guidance/notices/11-39.
[166] *Rule 2210. Communications with the Public,* available at https://www.finra.org/rules-guidance/rulebooks/finra-rules/2210.
[167] NASD Rule 3010 has been superseded by FINRA Rules 3110 and 3170, see https://www.finra.org/rules-guidance/rulebooks/retired-rules/3010. It required that each firm establishes and maintains "*a system to supervise the activities of each associated person that is reasonably designed to achieve compliance with applicable federal securities laws and FINRA rules*", see FINRA Regulatory Notice 11-39.

deletes the content" because a [*t*]*echnology that automatically erases or deletes the content of an electronic communication would preclude the ability of the firm to retain the communications in compliance with their obligations under SEA Rule 17a-4. Accordingly, firms and associated persons may not sponsor such sites or use such devices*." This rule was released a month after the introduction of the *Snapchat* app, which allows its users to publish messages which are only available for a short period of time. Social media policies must be enforced to be efficient: FINRA *2019 Report on Examination Findings and Observations*[168] noted that, while some firms had prohibited using texting, messaging, social media or collaboration applications, such as *WhatsApp*, for business-related communication with customers, they had not "*maintain a process to reasonably identify and respond to red flags that registered representatives were using impermissible personal digital channel communications in connection with firm business*." FINRA suspended a Florida broker for 30 days and fined him $5,000 in 2020 because he had used *WhatsApp* to conduct securities-related business with three of his firm's customers. He agreed to sign a letter of acceptance.[169] The broker had been one since 1994 and had never had any prior disciplinary history. Use of new communication technologies are so prevalent that messages sent this way may be regarded as sent using a customarily channel.

Lesley Stahl had also asked Elon Musk during the December 2018 interview: "*But how do they know if it's going to move the market if they're not reading all of* [*your tweets*] *before*

---

[168] *FINRA 2019 Report on Examination Findings and Observations*, *Digital Communications*, (Oct. 16, 2019), https://www.finra.org/rules-guidance/guidance/reports/2019-report-exam-findings-and-observations/digital-communication.

[169] The letter is available at https://www.finra.org/sites/default/files/fda_documents/2018059746001%20Paul%20A.%20Falcon%20CRD%202464566%20AWC%20sl.pdf.

you send them?" to which Musk had replied "*Well, I guess we might make some mistakes. Who knows?*" Stahl pressed on: "*Are you serious?"* to which Musk rather petulantly answered: "*Nobody's perfect*." This last exchange was quoted in bold in the motion for an order to show cause why Elon Musk should not be held in contempt for violating the terms of the Court's October 16, 2018 Final Judgment, filed by the SEC on February 25, 2018.[170] It claimed that Musk should be held in contempt of the Court's final judgment for having violated its terms, which required Musk to stop "*recklessly disseminating false or inaccurate information about Tesla*." Musk had published a tweet at 7:15 PM ET on February 19, 2019 which read that "*Tesla made 0 cars in 2011, but will make around 500K in 2019*," without seeking nor receiving prior approval. This statement was inaccurate but had nevertheless been sent to Musk's Twitter followers, which were 24 million at the time, and had not been preapproved. The SEC alleged that, after the tweed had been posted, Tesla's securities counsel met with Musk, and they drafted a corrective tweet, which was posted at 11:41 PM EST and read "*Meant to say annualized production rate at end of 2019 probably around 500k, ie 10k cars/week. Deliveries for year estimated to be around 400k."* The case settled, but SEC Commissioner Robert Jackson made his "dissenting statement" public, writing: "*Those who settle cases with the SEC must be held to the bargain they struck Given Mr. Musk's conduct, I cannot support a settlement in which he does not admit what is crystal clear to*

[170] Motion for Order to Show Cause Why Defendant Elon Musk Should Not Be Held in Contempt and Memorandum of Law in Support. Document filed by United States Securities and Exchange Commission, United States Securities and Exchange Commission v. Elon Musk, 1:18-cv-8865-AJN-(Feb. 25, 2019, S.D.N.Y.)

anyone who has followed this bizarre series of events: Mr. Musk breached the agreement he made last year with the Commission—and with American investors."[171]

D. Protecting the Safety and Security of the Public

a. Threatening Speech

Speech which is of a nature to create a clear and present danger is not protected by the First Amendment. Justice Holmes famously stated in 1919, in *Schenck v. United States*, that "[*t*]*he most stringent protection of free speech would not protect a man in falsely shouting fire in a theatre and causing a panic*,"[172] adding that the First Amendment "*does not even protect a man from an injunction against uttering words that may have all the effect of force*."[173] Justice Holmes then articulated what is now known as the clear and present danger test:

> "[*t*]*he question in every case is whether the words used are used in such circumstances and are of such a nature as to create a clear and present danger that they will bring about the substantive evils that Congress has a right to prevent. It is a question of proximity and degree*.[174]

Six years after *Schenck*, Justice Holmes elaborated further on the clear and present danger doctrine when dissenting in *Gitlow v. New York,* writing that "[*e*]*very idea is an incitement*," and noting that "[*t*]*he only difference between the expression of an opinion and*

---

[171] David Marino-Nachison, *Elon Musk's Spat With the SEC Is Over. Not Everyone Is Happy About That,* BARRONS, (May 1, 2019 12:21 pm ET), https://www.barrons.com/articles/tesla-ceo-elon-musk-sec-judge-51556727457.
[172] Schenck v. United States, 249 U.S. 47, 52 (1919).
[173] Id.
[174] Id.

an incitement in the narrower sense is the speaker's enthusiasm for the result."[175] However, what society considers to be a "*clear and present danger"* has greatly varied since *Schenck*, a case about whether protesters who had distributed leaflets advocating draft-resistance had violated the 1917 Espionage Act, by causing insubordination in the U.S. military and naval forces, and by obstructing recruitment efforts, at a time when the U.S. had just entered World War I. However, the test set forth by Justice Holmes, examining both the degree and the proximity of the danger, is still used today. The Supreme Court clarified the clear and present danger doctrine in 1951, when stating in *Dennis v. United States* that in these cases, the courts "*must ask whether the gravity of the evil, discounted by its improbability, justifies such invasion of free speech as is necessary to avoid the danger*."[176]

Fighting words are not protected by the First Amendment either. The Supreme Court recognized in 1942 in *Chaplinsky v. New Hampshire* that:

> "[*t*]*here are certain well-defined and narrowly limited classes of speech, the prevention and punishment of which has never been thought to raise any Constitutional problem*. *These include the lewd and obscene, the profane, the libelous, and the insulting or "fighting" words — those which by their very utterance inflict injury or tend to incite an immediate breach of the peace*."[177]

Thirty years later, the Supreme Court defined "*fighting words*" in *Cohen v. California* as "*those personally abusive epithets which, when addressed to the ordinary citizen, are, as a*

[175] Gitlow v. New York, 268 U.S. 652,673 (1925).
[176] Dennis v. United States, 341 U.S. 494,510 (1951).
[177] Chaplinsky v. New Hampshire, 315 U. S. 568, 571-572 (1942).

matter of common knowledge, inherently likely to provoke violent reaction."[178] The Court gave as examples "*direct personal insult*" or speech "*intentionally provoking a given group to hostile reaction."*[179] Insults are regularly proffered on social media[180] and appears to be so mainstream that Pope Francis even asked in 2020 Catholics engaged in such abuse to give up insults on social media for Lent.[181] Indeed, it is easy to post on Facebook or Twitter in the heat of the moment, even when free on bail, as was disgraced pharmaceutical executive Martin Shkreli when he posted on his Facebook page a $5,000 offer for a strand of Hillary Clinton's hair, then on a book tour to promote her memoirs. Judge Kiyo A. Matsumoto from the Eastern District of New York revoked his bail, as the social media post could be perceived as a true threat, arguing that it was *"a solicitation to assault in exchange for money that is not protected by the First Amendment."*[182]

Indeed, true threats are not protected speech. In *Virginia v. Black et al.*, the Supreme Court defined "*true threats*" as "*statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals…" The speaker need not actually intend to carry out the*

---

[178] Cohen v. California, 403 U.S. 15, 20 (1971).
[179] Cohen, at 20.
[180] See for example Maxime Gautier, *Percentage of French having already be the target of insults or rude comments on social media in 2019*, (July 14, 2020), STATISTA, https://www.statista.com/statistics/972422/share-persons-already-victim-insults-social-media-by-age-france, reporting that 4 percent of the people who had been polled had been insulted on social media in the past year. Insults may be made under the cover of anonymity, but not always, as Donald Trump insulted people on social media so regularly that it warranted an article in The New York Times, see Jasmine C. Lee, and Kevin Quealy, *The 598 People, Places and Things Donald Trump Has Insulted on Twitter: A Complete List*, THE NEW YORK TIMES,(Updated May 24, 2019), https://www.nytimes.com/interactive/2016/01/28/upshot/donald-trump-twitter-insults.html.
[181] *Pope to Catholics: For Lent, give up trolling*, REUTERS, (Feb. 26, 2020 6:51 AM), https://www.reuters.com/article/us-pope-generalaudience-lent-insults/pope-to-catholics-for-lent-give-up-trolling-idUSKCN20K1O7.
[182] As quoted by the New York Times. See Stephanie Clifford, *Martin Shkreli Is Jailed for Seeking a Hair From Hillary Clinton*, THE NEW YORK TIMES, (Sep. 13, 2017), https://www.nytimes.com/2017/09/13/business/dealbook/martin-shkreli-jail.html

threat."[183] The Court held that a Virginia's statute banning cross burning with "*an intent to intimidate a person or group of persons*" violated the First Amendment, as the statute treated any cross burning as *prima facie* evidence of intent to intimidate. A statute banning cross burning with the intent to intimidate would be, however, consistent with the First Amendment. In *Virginia v. Black*, Black had burned a cross during a Ku Klux Klan meeting. And had been charged under the Virginia statute. The jury had been instructed that "*intent*" meant "*the motivation to intentionally put a person or a group of persons in fear of bodily harm*."

The Supreme Court reviewed the history of the Klan and of cross burning to conclude that the burning of a cross is a "*symbol of hate, and that it sometimes is used to intimidate*."[184] The Court explained that "true threats" are "*statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals*."[185] If applying this definition to social media networks, a threat can be directed at a particular individual, if publishing the true threat using a feature of the platforms allowing to communicate privately. It can also be directed at one individual, but posted publicly, by publicly directing the message at one individual. The individual message can then be further disseminated by being republished by third parties, after having become aware of its content and thus be aware to the "expression of an intent to commit an act of unlawful violence" contained in the message they have republished. It is not necessary that the speaker intends to carry out the threat

[183] Virginia v. Black, 538 U.S. 343, 359-360 (2003).
[184] Virginia v. Black, at 357.
[185] Virginia v. Black, at 359.

for a particular speech to be a true threat, as the purpose of prohibiting true threats is to protect from the fear of violence and from the disruption engendered by fear.[186] The Court explained in *Virginia v. Black* that "*[i]ntimidation in the constitutionally proscribable sense of the word is a type of true threat, where a speaker directs a threat to a person or group of persons with the intent of placing the victim in fear of bodily harm or death.*"[187]

The Terry Jones case provides an example of incitement to violence which used social media to find a large public and successfully foment violence.[188] Jones had planned to burn a Quran on September 11, 2010 to protest the edification of an Islamic center near the World Trade Center site in Manhattan. He backed down following international pressure but finally did so in March 2011, streaming the event live on social media.[189] The 2011 burning had been also delayed because the city attorney of Gainesville, Florida, where Jones was a pastor and planned to accomplish his deed, changed the fire code to prevent him from burning the Quran outdoors. Jones then accomplished his deed indoors, and a fire department official observed the event.[190] Professor Lyrissa Barnett Lidsky wrote that "*Jones's speech was not a true threat to do violence to another because he never communicated any intent whatsoever to commit a violent act against another; nor did he*

[186] Virginia v. Black, at 360.
[187] Virginia v. Black, at 360.
[188] See for example Lyrissa Barnett Lidsky, *Incendiary Speech and Social Media*, 44 Tex. Tech L. Rev. 147 (2011).
[189] Florida pastor Terry Jones arrested on way to burn Qur'ans, THE GUARDIAN, (Sept. 12, 2013 09.25 EDT), https://www.theguardian.com/world/2013/sep/12/florida-pastor-terry-jones-qurans. He was also charged with unlawful open-carry of a firearm and Sapp was charged for failure to have a valid registration for the trailer. Jones was again arrested in 2013 as he was about to reiterate his act and was charged for unlawful conveyance of fuel, as he was driving a truck with a barbecue grill and Qurans soaked in kerosene.
[190] Kevin Sieff, THE WASHINGTON POST, id.

urge others to commit violence on his behalf."[191] Indeed, in *Black v. Virginia*, Justice O'Connor explained that "*while a State, consistent with the First Amendment, may ban cross burning carried out with the intent to intimidate*, [*a statute*] *treating any cross burning as prima facie evidence of intent to intimidate"* would make it unconstitutional.[192] It is not an easy task to assess whether a particular social media message is an illegal incitement to violence, but it however be argued, yet not proven, that Jones' act directly led to violence[193] as protesters enraged by this act thwarted security at an U.N. compound in Afghanistan and killed seven staff members.[194] Words have an impact, and researchers from the Brookings Institution analyzed in October 2020 three of President Trump tweets, described as "*attacking politicians on Twitter while remaining in-bounds of platforms' content moderation policies*."[195] The three tweets attacked Michigan Governor Gretchen Whitmer, Virginia Governor Ralph Northam, and Ohio Congressman Tim Ryan. The researchers analyzed random samples of tweets about each of the three politicians, posted after the President's tweets, and found that "*in the immediate aftermath of Trump's tweets, levels of severe toxicity and threats increased in all three cases*."

---

[191] Lyrissa Barnett Lidsky, id., at 152, citing *Virginia v. Black*, 538 U.S. 348, 359, where the Supreme Court defined "true threats" as "*encompass*[*ing*] *those statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals*."
[192] Virginia v. Black, at 348.
[193] Kevin Sieff, *Florida pastor Terry Jones's Koran burning has far-reaching effect*, THE WASHINGTON POST, (April 2, 2011), https://www.washingtonpost.com/local/education/florida-pastor-terry-joness-koran-burning-has-far-reaching-effect/2011/04/02/AFpiFoQC_story.html?noredirect=on&utm_term=.e884156f8185.
[194] Mohammad Bashir, *Worst attack on U.N. in Afghanistan kills at least 7,* REUTERS.COM, (April 1, 2011, 3:51 PM), https://www.reuters.com/article/us-afghanistan/worst-attack-on-u-n-in-afghanistan-kills-at-least-7-idUSTRE7306JP20110401.
[195] Megan Brown and Zeve Sanderson, *How Trump impacts harmful Twitter speech: A case study in three tweets*, BROOKINGS, (Oct. 22, 2020), https://www.brookings.edu/techstream/how-trump-impacts-harmful-twitter-speech-a-case-study-in-three-tweets.

It is a federal crime to transmit in interstate commerce "*any communication containing any threat… to injure the person of another*."[196] True threats are "*generally not entitled to First Amendment protection*."[197] The Supreme Court published on June 1, 2015, its *Elonis v. U.S.* opinion.[198] *Elonis* is not a First Amendment but a criminal law case which discussed the mental state required by Section 875 (c) of the U.S. Code. It is a crime under a federal law first enacted in 1932, 18 U.S.C. §875(c), to *"transmits in interstate or foreign commerce any communication containing any threat to kidnap any person or any threat to injure the person of another."* These are "true threats" and are thus outside of the scope of protection of the First Amendment. The issue was whether the defendant must have transmitted the communication for the purpose of issuing a threat, or if the law requires the defendant to have known that the communication will be viewed as a threat.

Anthony Elonis was indicted on several counts by a grand jury under Section 875 (c) because he had posted multiple time messages on his Facebook page, in the form of rap lyrics, which could be interpreted as threats to his estranged wife, to visitors of the amusement park where he was employed at the time, and to the FBI agent who had come to his home to investigate. He was later convicted by a jury of four counts of violating 18 U.S.C. § 875(c) for posting these messages on Facebook. The jury instructions had explained that "[*a*] *statement is a true threat when a defendant intentionally makes a statement in a context or under such circumstances wherein a reasonable person would foresee that the statement would be interpreted by those to whom the maker communicates*

[196] 18 U.S.C. § 875(c).
[197] United States v. Keyser, 704 F.3d, 631, 638 (9th Cir. 2012).
[198] Elonis v. U.S., 575 U.S. (2015).

the statement as a serious expression of an intention to inflict bodily injury or take the life of an individual."

Elonis filed post-convictions motions, arguing that, under *Virginia v. Black*, a subjective intent to threaten is required under the true threat exception to the First Amendment. The Eastern District Court of Pennsylvania denied theses motions, reasoning that Third Circuit's precedents required the Government only to prove that the defendant `*acted knowingly and willfully'* when making the threatening communication, and that the communication was reasonably perceived as threatening bodily injury.[199] Elonis appealed to the Third Circuit Court of Appeals, presenting the case as a true threat exception to the First Amendment case, asking "*whether the true threats exception to speech protection under the First Amendment requires a jury to find the defendant subjectively intended his statements to be understood as threats*,"[200] and arguing that he did not subjectively intend his Facebook posts to be threatening. The Third Circuit Court of Appeals upheld Elonis' convictions. The Court considered whether *Virginia v. Black*,[201] required a subjective intent to threaten. In 1991, the Court had defined true threat" in *United States v. Kosma*[202] as requiring that " *the defendant intentionally make a statement, written or oral, in a context or under such circumstances wherein a reasonable person would foresee that the statement would be interpreted by those to whom the maker communicates the statement as a serious expression of an intention to inflict bodily harm upon or to take the life of the President, and that the statement not be the result of mistake, duress, or coercion."* As such, it is the person

[199] US v. Elonis, 897 F. Supp. 2d 335, 339, citing U.S. v. Voneida, 337 Fed. Appx. 246, 247 (3d Cir.2009), quoting U.S. v. Himelwright, 42 F.3d 777, 782 (3d Cir.1994).
[200] US v. Elonis, 730 F. 3d 321, 323, (3d Cir. 2013).
[201] Virginia v. Black, at 359.
[202] United States v. Kosma, 951 F.2d 549, 557 (3d Cir.1991).

or group of persons receiving the communications whose subjective intent is considered, which is a delicate task when the content has been published on social media. The Third Court reasoned that "[*l*]*imiting the definition of true threats to only those statements where the speaker subjectively intended to threaten would fail to protect individuals from "the fear of violence" and the "disruption that fear engenders," because it would protect speech that a reasonable speaker would understand to be threatening*" and that "*the true threats exception requires a subjective intent to threaten*."[203]

Elonis petitioned for a brief of certiorari to the Supreme Court, asking "[*w*]*hether, consistent with the First Amendment and Virginia v. Black,… conviction of threatening another person requires proof of the defendant's subjective intent to threaten, as required by the Ninth Circuit … or whether it is enough to show that a "reasonable person" would regard the statement as threatening, as held by other federal courts of appeals and state courts of last resort*. "The Supreme Court granted the petition, reversed, and remanded in an opinion authored by Chief Justice Roberts, who avoided however to address the First Amendment issues, focusing instead on the required mental state. For the Supreme Court, it had been error to instruct the jury that the Government only needed to prove that a reasonable person would regard Elonis's communications as threat. At issue is the classic criminal issue of mental state*, mens rea*, but "[*t*]*he fact that the [Section 875(c)] does not specify any required mental state, however, does not mean that none exists."*[204] Chief Justice Roberts reasoned that "*Section 875(c)… requires proof that a communication was transmitted and that it contained a threat"* and that "[*t*]*he mental state requirement must therefore apply to*

[203] U.S. v. Elonis, at 330.
[204] Elonis v. US, 135 S. Ct. 2001, 2009 (2015).

the fact that the communication contains a threat."[205] Elonis had been convicted "*on how his posts would be understood by a reasonable person*," a standard akin to a negligence standard, which, as noted by Chief Justice Roberts, is "a *familiar feature of civil liability in tort law, but is inconsistent with "the conventional requirement for criminal conduct."*[206] The jury had been erroneously instructed that the Government had only to prove that a reasonable person would regard Elonis's communications as threats, whereas "[*f*]*ederal criminal liability generally does not turn solely on the results of an act without considering the defendant's mental state."*[207] Section 875(c)'s required mental state "*is satisfied if the defendant transmits a communication for the purpose of issuing a threat."*[208] Justice Kennedy noted during the *Elonis v. U.S.* oral arguments that he was "*not sure that the [Supreme] Court did either the law or the English language much of a good service when it said "true threat." It could mean so many things*. *It could mean that you really intend to carry it out, A; you really intend to intimidate the person; or that no one could possibly believe it*."[209] *Indeed, a message social media can "mean so many thing."*

In a pre-*Elonis* case, a fan of the Saint Louis baseball team, the Cardinals, posted several tweets during the 2013 World Series, where his favorite team played against the Boston Red Socks:

---

[205] Elonis, at 2011.
[206] Elonis, at 2011.
[207] Elonis, at 2012.
[208] Elonis, at 2012.
[209] Oral Arguments, Elonis v. United States, December 1, 2014, available at Elonis v. United States, OYEZ.COM, https://www.oyez.org/cases/2014/13-983.

(October 21, 2013): "Going to be tailgating with a #PressureCooker during games 3-4-5 in #STL during #WorldSeries. #STLStrong #GoCards #postseason from Springfield, MO."

(October 22, 2013): "*Putting my loft up for ridiculous "Boston-only" rate on @airbnb for "the #WorldSeries. Pressure cooker sold separately*."

(October 22, 2013): "*The #WorldSeries will be another finish line not crossed by #Boston*."

(October 25, 2013): "*Listening to the Offspring's "Bad Habit" and the lyrics just ring true of what will go down very soon*."

Defendant was charged with making a terrorist threat under Section 574.115 of the Revised Missouri Statutes, under which:

> "*A person commits the crime of making a terrorist threat if such person communicates a threat to cause an incident or condition involving danger to life, communicates a knowingly false report of an incident or condition involving danger to life, or knowingly causes a false belief or fear that an incident has occurred or that a condition exists involving danger to life:*
>
> *(1) With the purpose of frightening ten or more people;*
>
> *(2) With the purpose of causing the evacuation, quarantine or closure of any portion of a building, inhabitable structure, place of assembly or facility of transportation; or*
>
> *(3) With reckless disregard of the risk of causing the evacuation, quarantine or closure of any portion of a building, inhabitable structure, place of assembly or facility of transportation; or*

(4) With criminal negligence with regard to the risk of causing the evacuation, quarantine or closure of any portion of a building, inhabitable structure, place of assembly or facility of transportation."

The scope of Section 574.115 is quite large as even though its title is "*Making a terrorist threat, penalty*," it can be used to charge an individual who had merely the intent to do a prank, or wants his school evacuated to avoid taking a test and posts a message on social media stating that he will blow up the school up today. Such an act could indeed be described as "*communicating a knowingly false report of an incident or condition involving danger to life."* Defendant filed a motion to dismiss and argued that the tweets were protected speech, and that his "*sarcastic posts on Twitter did not constitute 'true threats' as a matter of law and cannot be punished by the State*." The Circuit Court of the City of St. Louis dismissed, and the Missouri Appeals Court affirmed.[210] The Appeals Court cited *Virginia v. Black* but noted that the Supreme Court had "*provided minimal guidance to courts tasked with the challenge of distinguishing "true threats" from protected speech*." Citing the definition of true threats provided by the Supreme Court, *"statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals. The speaker need not actually intend to carry out the threat. "*[211] The Court of Appeals further noted that federal courts, following *Black*, have found communicating a statement that "*a reasonable jury could find… expressed an intent to injure in the present or future*" were true threats under 18 U.S.C.

[210] State of Missouri v. Metzinger, No. ED101165 (Feb. 24, 2015), available at https://www.courts.mo.gov/file.jsp?id=83896.
[211] Black, 538 U.S. at359-60.

875(c), which incriminates transmitting threatening communications in interstate commerce.[212] The tweets "*were not serious expressions of an intent to cause injury to another*" and "*were made in the context of sports rivalry, an area often subject to impassioned language and hyperbole*."[213]

Two years after *Elonis*, a man who had allegedly sent three tweets to U.S. Senator Joni Ernst, was charged by a grand jury with transmitting threatening communications in violation of Section 875(c).[214] The Defendant moved to dismiss, claiming that the tweets were either "*impolite criticism*" or threats impossible to execute. The motion was denied, based on the Report and Recommendation of Judge Williams, who addressed whether the indictment has violated First Amendment. Noting that true threats are protected by the First Amendment, and citing *Virginia v. Black*, Judge Williams found improper to dismiss the indictment based on defendant's First Amendment argument, as defendant's claim that his tweets were protected speech, not threats, is a factual question. Judge William argued that "*a jury may conclude that defendant's statements did not constitute a threat. A jury could also conclude that calling the senator a "bitch," and stating that he intended to "f her up" and "beat" her, constituted true threats*. "[215]

---

[212] "*Whoever transmits in interstate or foreign commerce any communication containing any threat to kidnap any person or any threat to injure the person of another, shall be fined under this title or imprisoned not more than five years, or both*."

[213] Citing Howard M. Wasserman, *Fans, Free Expression, and the Wide World of Sports*, 67 U.PITT.L.REV. 525, 579(2006) ("*Nor can we forget cheering speech's dependence on humor, satire, and rhetorical hyperbole and overstatement, none of which is intended or reasonably capable of being taken literally*.").

[214] US v. Dierks, (Dist. Court, ND Iowa) (2017).

[215] The three tweets had been sent to Senator Joni K. Ernst at her Twitter accounts, @SenJoniErnst and @JoniErnst on August 16, 2017 and read (1)"*u r sn army bitch and I'll @USMC u tf up :)(:",* (2) "*I'll f u up seriously in my sleep*" and (3) "*I'll beat ur ass in front of ur widow I promise that*."

In Great Britain, Section 127(1) (a) of the Communications Act 2003 prohibits sending "*by means of a public electronic communications network a message or other matter that is grossly offensive or of an indecent, obscene or menacing character*" and its Section 127(3) provides that a person guilty of such offense can be sentenced to no more than six months in prison and to a fine. British accountant Paul Chambers was tried under Section 127(1) (a) for a tweet sent on January 6, 2010. He was scheduled that day to fly within a week to Northern Ireland from the Robin Hood airport in Yorkshire and learned that the airport was closed. This upset him and he vented his frustration on Twitter on such terms:

> "*Crap! Robin Hood Airport is closed. You've got a week and a bit to get your shit together otherwise I am blowing the airport sky high!!"*

This message led to his arrest for a few days later for suspicion of involvement in a bomb hoax. During his interrogation by the police, he assured that the message was a joke, but was nevertheless charged under Section 127 of the Communications Act 2003. The House of Lords had been asked in 2006, on appeal by the Director of Public Prosecutions (DPP), to consider the meaning and application of Section.[216] In this case, a man had called a MP office on the phone several times and had expressed messages which were not "*of menacing character*" but were "*grossly offensive*" within the meaning of Section 127(1)(a). Lord Bingham, at the time Senior Law Lord, that is, the President of the Supreme Court, noted that "[*t*]*he purpose of the legislation which culminates in section 127(1)(a) was to prohibit the use of a service provided and funded by the public for the benefit of the public for the transmission of communications which contravene the basic standards of our society*."

[216] DPP v. Collins, UKHL 40 (July 19, 2006).

In the *Chambers* case, the Crown Court of Doncaster uphold Chambers' conviction of having sent a message of a "*menacing character by an electronic communication network*," and held "*that the required mens rea … is that the person sending the message must have intended the message to be menacing, or be aware that it might be taken to be so*." The Crown Court found Chambers was aware that his message was of a menacing character.

Chambers appealed to the High Court, which held that the tweet posted he had posted was not of a menacing character, because:

> "*it is unusual for a threat of a terrorist nature to invite the person making it to ready identified, as this message did*. … [*A*]*though we are accustomed to very brief messages by terrorists to indicate that a bomb or explosive device has been put in place and will detonate shortly, it is difficult to image a serious threat in which warning of it is given to a large number of tweet "followers" in ample time for the threat to be reported and extinguished*."

The High Court noted that the message:

> "*was posted on "Twitter" for widespread reading, a conversation piece for the appellant's followers, drawing attention to himself and his predicament. Much more significantly, although it purports to address "you", meaning those responsible for the airport, it was not sent to anyone at the airport or anyone responsible for airport security, or indeed any form of public security. The grievance addressed by the message is that the airport is closed when the writer wants it to be open. The language and punctuation are inconsistent with the writer intending it to be or to be taken as a serious warning*."

The High Court, touched briefly upon the issue of intent, noting that, as Section 127(1)(a) makes no express provision for *mens rea*, and it is therefore an offence of basic intent. This intent, which Chambers did not have in this case, is "*to have intended that the message should be of a menacing character (the most serious form of the offence) or alternatively, if [the offender] is proved to have been aware of or to have recognised the risk at the time of sending the message that it may create fear or apprehension in any reasonable member of the public who reads or sees it.*"

Social media messages can send the speaker to jail. When there, is it possible to continue to be an active social media user?

b. Preventing Prisoners from Using Social Media

The Supreme Court noted in 1974 that "*a prison inmate retains those First Amendment rights that are not inconsistent with his status as a prisoner or with the legitimate penological objectives of the corrections system*"[217] and Justice Marshall wrote in a concurring opinion, in another case the same year, that "[*a*]*prisoner does not shed such basic First Amendment rights at the prison gate*."[218] Does that mean that prisoners may freely use social media platforms?

Facebook disclosed in 2016 in a *Transparency Report* that it had disabled 53 U.S. prisoner accounts and 74 U.K. prisoner accounts in the second half of 2015, after "*governmental authorities identified either unlawful access to our service or safety issues*."[219]

---

[217] Pell v. Procunier, 417 US 817, 822 (1974).
[218] Procunier v. Martinez, 416 US 396, 422 (1974) (Marshall, J., concurring).
[219]Dave Maass, *Report Inmate Social Media Takedowns to OnlineCensorship.org*, EFF, (May 5, 2016), https://www.eff.org/deeplinks/2016/05/report-inmate-social-media-takedowns-onlinecensorshiporg.

Indeed, inmates all around the world generally do not have the right to access Internet, nor even to have a smart phone. Inmates could use social media to perpetrate illegal activities or to actively prepare crimes, by recruiting accomplices or gathering information. This would be the safety issues alluded to by Facebook. What about unlawful access? Which are the rules prohibiting prisoners to access a social media site?

Section 14 -11-70 of the Alabama Code prohibits inmate from establishing or maintaining a social media account, which is defined by Section 14-11-70(b) as "*an Internet-based website that has any of the following capabilities: (1) Allows users to create web pages or profiles about themselves that are available to the general public or to any other users. (2) Offers a mechanism for communication among users, such as a forum, chat room, electronic mail, or instant messaging*." This misdemeanor is punishable by a fine which cannot exceed five hundred dollars. Under Rule 905 "*Creating and/or Assisting With A Social Networking Site*" of the South Carolina Department of Corrections, "[*t*]*he facilitation, conspiracy, aiding, abetting in the creation or updating of an internet web site or social networking site*" is a level 1 offense of violation of prison conduct policies, just as sexual assault, homicide or hostage taking.[220] According to documents obtained by the Electronic Frontier Foundation under a South Carolina's Freedom of Information Act request, inmates are charged for each day he posts on a social media site: each day is a count.[221]

[220] See https://www.eff.org/files/2015/02/12/scdc_social_media_discipline_policies.pdf, p. 27 and Dave Maas, *Hundreds of South Carolina Inmates Sent to Solitary Confinement Over Facebook*, EFF, (Feb. 12, 2015), https://www.eff.org/deeplinks/2015/02/hundreds-south-carolina-inmates-sent-solitary-confinement-over-facebook.

[221] The documents obtained by the EFF are available at https://www.eff.org/files/2015/02/12/scdc_social_media_discipline_policies.pdf (last visited Dec. 30, 2020).

The freedom of speech of inmates' family members is also restricted. For instance, the sister of an inmate incarcerated in Indiana who claimed his innocence filed suit in February 2015 against the Commissioner of the Indiana Department of Correction (DOC), claiming that the DOC, had violated her freedom of speech when blocking her from communicating online with her brother because of one of her Facebook posts. The sister of the inmate had reposted on her own Facebook account a videogram sent to her by her brother through the DOC's "J-Pay" system, a private electronic communication system used by inmates to send email, messages and videograms to persons who have been approved by the DOC. The inmate had forwarded to his sister a videogram thanking persons who have supported his claim of innocence and urging them to attend upcoming court hearings. The sister reposted it on her Facebook 'event' page, adding as a comment: "*PACK THE COURT! MAKE IT TREMBLE! Justice for* [*my brother*]." The DOC discovered this Facebook post and then disciplined the inmate The DOC also blocked the inmate's sister from using J-Pay to communicate with her brother and informed to her that her communication privileges has been revoked because of unauthorized communication via social media.[222] In Texas, the Department of Criminal Justice updated in April 2016 its "General Information Guide for Families of Offenders."[223] It stated that offenders were "*prohibited from maintaining active social media accounts, including Facebook, Twitter, Instagram and similar social media, for the purposes of soliciting, updating, or engaging others, through a third party or otherwise*." The families were instructed not to assist inmates to bypass this policy

[222] Kristine Guerra, Inmate's sister in trouble for Facebook post, USA TODAY (Feb. 5, 2015, 9:03 PM), https://www.usatoday.com/story/news/nation/2015/02/05/inmates-sister-in-trouble-for-facebook-post/22955289/.
[223] Texas Department of Criminal Justice, *General Information Guide for Families of Offenders* (April 2016), https://www.tdcj.texas.gov/documents/General_Information_Guide_for_Families_of_Offenders.pdf (last visited Dec. 30, 2020).

by operating an account on behalf of their family member. However, third parties had the right to operate a social media account about an inmate or discussing an inmate, unless the account was presented as being personally maintained by the inmate. While inmates do not have direct access to social media sites while in jail, since they have no Internet access, unless they possess a contraband cell phone, some nevertheless maintained an active social media presence by forwarding their updates by mail to people outside of the jail, which then updated the inmate social media account accordingly. Why this change of policy? A spokesperson for the Texas Department of Criminal Justice explained to a journalist that it is because "[*o*]*ffenders have used social media accounts to sell items over the internet based on the notoriety of their crime, harass victims or victim's families, and continue their criminal activity*."[224]

In 2017, the Supreme Court invalidated in *Packingham v. North Carolina* a North Carolina statute restricting sex offenders to access social media because it violated the First Amendment by being too broad[225] Chief Justice Roberts, writing for the Court, emphasized that the case was "*one of the first* [*the*] *Court has taken to address the relationship between the First Amendment and the modern Internet. As a result, the Court must exercise extreme caution before suggesting that the First Amendment provides scant protection for access to vast networks in that medium*."[226] The North Carolina General Statutes Chapter 14, Section 205.5 (Section 205.5) made it unlawful for a sex offender "*to access commercial social networking* [*w*]*ebsite where the sex offenders knows that the site permits minor children to*

[224] Casey Tolan, *Texas is banning inmates from having social media accounts*, SPLINTERNEWS (April 13, 2016, 11:43AM), https://splinternews.com/texas-is-banning-inmates-from-having-social-media-accou-1793856124.
[225] Packingham v. North Carolina, 137 S. Ct. 1730 (2017).
[226] Packingham, at 1736.

become members or to create or maintain personal [*w*]*ebpages."* Sex offenders were however authorized to access websites "[*p*]*rovid*[*ing*] *only one of the following discrete services: photo-sharing, electronic mail, instant messenger, or chat room or message board platform*"[227]and websites having as "*primary purpose the facilitation of commercial transactions involving goods or services between* [*their*] *members or visitors*."[228] Mr. Packingham, a North Carolina resident and a former sex offender, was thus barred by the statute from accessing social media sites. He shared however on Facebook in April 2010, using a fictitious name, his happiness for having had a traffic ticket dismissed.[229] Durham Police Corporal found the post while searching, using his own Facebook account, for sex offenders using the social networking site. Packingham was arrested, indicted under the statute, convicted, and sentenced to a suspended prison sentence. The North Carolina Court of Appeals unanimously overturned the conviction, finding that Section 205.5 violated the First Amendment, as "*probibit*[*ing*] *an enormous amount of expressive activity on the internet,*"[230] including speech "*unrelated to online communication with minors.*" The North Carolina Supreme Court however reversed, finding that Section 205.5 is a "*limitation on conduct*," not on speech, as it only prevents "*accessing*" the social media sites and the burden on sex offenders' ability "*to engage in speech on the Internet*" is only "*incidental*."[231]

[227] § 14-202.5(c)(1).
[228] § 14-202.5(c)(2).
[229] The post read: "*Man God is Good! How about I got so much favor they dismissed the ticket before court even started? No fine, no court costs, no nothing spent……Praise be to GOD, WOW! Thanks JESUS*."
[230] State v. Packingham, 748 SE 2d 146.
[231] State v. Packingham, 777 SE 2d 738,744 (2015): "*This limitation on conduct only incidentally burdens the ability of registered sex offenders to engage in speech after accessing those Web sites that fall within the statute's reach. Thus we conclude that section 14-202.5 is a regulation of conduct.*"

Packingham had noted in his certiorari brief the prevalence of social media use, passing even one billion for Facebook. During the debates, his attorney, David T. Goldberg, argued that Section 205.5 "*does not operate in some sleepy First Amendment quarter…* [*but*] *operates and forbids speech on the very platform on which today Americans are most likely to communicate, to organize for social change, and to petition the government*."[232] North Carolina argued that the statute had been enacted to prevent registered sex offenders to use social media sites to obtain children's personal information. The brief generally assimilated registered sex offenders as "*predators*," thus appearing to take the view that once a sex offender, always a predatory sex offender.[233] During the debates, Justice Kagan asked North Carolina's counsel if the statute would prevent a sex offender from accessing President's Trump's Twitter account, and he answered it would. Justice Kagan then noted that all governors, senators, and members of the House have a Twitter account. Chief Justice Roberts, writing for the Court, stated that cyberspace, "*and social media in particular*" are "*the most important places (in a spatial sense) for the exchange of views*," further noting that Facebook's 1.79 billion active users represent "*about three times the population of North America*,"[234] and that social media users can "*gain access to information and communicate with one another about it on any subject that might come to mind."*[235] The

[232] Transcript of Oral Argument, Packingham v. North Carolina, available at OYEZ, https://www.oyez.org/cases/2016/15-1194 (last visited Dec. 30. 2020).
[233] Brief for Respondent at 1-2, Packingham v. North Carolina, 137 S. Ct. 1730 (2017) (No.15-1194). Respondent argued further that "[*s*]*exual predators' use of the Internet has created special challenges to society as it attempts to protect its most vulnerable members. The Internet does not merely allow predators to communicate more easily with children whom they stalk. It also allows them to gain intimate information about children's social lives, families, hobbies, and hangouts. Predators then use that information to target an unwitting victim, either in person or online, under the guise of familiarity or shared interests*" and that was thwart that Section 205.5 was enact to "*thwart that conduct*."
[234] Packingham, at 1735.
[235] Packingham, at 1737.

statute does not only prevent registered sex offenders to access information, but it also prevents them to use social media to speak:

> "*By prohibiting sex offenders from using those websites, North Carolina with one broad stroke bars access to what for many are the principal sources for knowing current events, checking ads for employment, speaking and listening in the modern public square, and otherwise exploring the vast realms of human thought and* knowledge. *These websites can provide perhaps the most powerful mechanisms available to a private citizen to make his or her voice heard*," adding that "*convicted criminals — and in some instances especially convicted criminals — might receive legitimate benefits from these means for access to the world of ideas, in particular if they seek to reform and to pursue lawful and rewarding lives*."[236]

In *Doe v. Prosecutor, Marion County, Indiana*, the Seventh Circuit acknowledged that sex offenders can use social media sites to search out minors to later solicit them, but added that entirely barring sexual predators from social media would "*allo*[*w*] *law enforcement to swoop in and arrest perpetrators before they have the opportunity to send an actual solicitation* is a "*speculative*" argument.[237] The Court added that "*perhaps such a law could apply to certain persons that present an acute risk — those individuals whose presence on social media impels them to solicit children*."[238] In *Packingham*, the Supreme Court reasoned, even "[t]*hough the issue* [*was*] *not before the Court … that the First Amendment permits a State to enact specific, narrowly tailored laws that prohibit a sex offender from*

[236] Packingham, at 1737.
[237] Marion County, Indiana, at 701.
[238] Marion County, Indiana, at 702.

engaging in conduct that often presages a sexual crime, like contacting a minor or using a website to gather information about a minor."[239] The State had not met the burden to prove that the "*sweeping*" statute was "*necessary or legitimate to serve* [*the*] purpose [of keeping convicted sex offenders away from vulnerable victims.]"[240]

A few years before *Packingham*, the Seventh Circuit Court of Appeals had found that an Indiana statute prohibiting most registered sex offenders from using social networking websites, instant messaging services, and chat programs,[241] was unconstitutional.[242] The Court found the statute to be overbroad, noting that Indiana agreed that "*there is nothing dangerous about* [*Plaintiff, a registered sex offender*]*'s use of social media as long as he does not improperly communicate with minors*" and that "*there is no disagreement that illicit communication comprises a minuscule subset of the universe of social network activity*." Therefore, the Indiana statute was overbroad as it "*targeted substantially more activity than the evil it seeks to redress*."[243] In *Packingham*, the Court found that the statute was broadly worded and this could prevent registered sex offenders to access websites such as Amazon.com, Washingtonpost.com, and Webmd.com.[244]

The French press reported in 2015 about a Facebook page, 'MDR o Baumettes', which appeared to have been created by inmates of famous Marseilles' jail Les Baumettes.

---

[239] Packingham, at 1737.
[240] Packingham, at 1737.
[241] Indiana Code § 35-42-4-12 prohibited certain sex offenders from "*knowingly or intentionally us*[*ing*] *a social networking web site*" or "*an instant messaging or chat room program*" that "*the offender knows allows a person who is less than eighteen (18) years of age to access or use the web site or program*." Violation of the statute was a Class A misdemeanor, but subsequent violations were Class D felonies. The law applied to anyone required to register as sex offender under Indiana Code § 11-8-8, et seq.
[242] Doe v. Prosecutor, Marion County, Indiana, 705 F. 3d 694(7th Circ. 2013).
[243] Marion County, Indiana, at 699.
[244] Packingham, at 1736.

The page featured images of inmates taken from inside the jail, several of them in possession of items which inmates are not allowed to possess, such as cell phones or cash, or even illegal, such as marijuana. This publicity led to an administrative inquiry inside the jail, and the page was taken down. Indeed, French prisoners do not have the right to "*capture, fix or record or attempt to capture, fix or record, by any means whatsoever, images or sounds in an* [*penal*] *establishment or* to *broadcast or attempt to broadcast, by any means whatsoever, fixed images or sounds captured in an* [*penal*] *establishment, or to participate in such capture, fixation, recording or broadcast*."[245] As such, they do not have a right to post on social media images taken inside the jail. They do not have the right to own cell phones either. Article 434-35 of the criminal Code incriminates communicating with an inmate "*by any means… except in cases permitted by the regulations*," a crime punishable by one-year imprisonment and 15,000 Euros fine. Would it be illegal under article 434-35 to communicate with a French inmate on social media? It seems that it is not necessary that the accused had knowledge of the fact that social media communications are forbidden to inmates, and the crime is punishable merely by communicating with an inmate using an illegal way.

c. Protecting the Right of Law Enforcement Officers

French representative Éric Ciotti presented on May 26, 2020 a bill proposing to add an article 35 quinquies to the French Press Law to forbid publishing, by any means whatsoever, and on any support whatsoever, the image of police officers, militaries, and

[245] CODE DE PROCÉDURE PÉNALE [C. P.P.], Article R57-7-45 15 °(Fr.).

customs employees.[246] The fine could not be less than 10,000 Euros and the prison sentence could not be less than six months. However, the courts would have been able to issue lower penalties, but only by "*a specially reasoned decision… in consideration of the circumstances of the offense, the personality of its author or the guarantees of insertion or reinsertion presented*." The timing of the bill is interesting as it was published the day after, in the U.S., George Floyd was murdered by a police officer in Minneapolis, applying his boot on Mr. Floyd's neck for almost nine minutes while three other police officers watched the scene without intervening.[247]A bystander video had captured the scene, where Mr. Floyd was heard saying *"I can't breathe, please, I can't breathe*." The video, posted on social media, became viral and led to protests in the U.S. and around the world, leading to stronger public support of the *Black Lives Matter* movement.[248] France is not immune to police violence, which led to multiple protests in 2020.[249] The French May 2020 bill explained that its purpose was to protect police officers against "police bashing", which is "*dangerously developing,*" giving as example the *Urgence violences policières* app (Emergency police violence app), which "*has the effect of stigmatizing police officers but also to circulate, particularly on social media, erroneous information about them*." The app was created by a

[246] *Proposition de loi N° 2992 visant à rendre non identifiables les forces de l'ordre lors de la diffusion d'images dans l'espace médiatique*, http://www.assemblee-nationale.fr/dyn/15/textes/l15b2992_proposition-loi#.
[247] Christine Hauser, Derrick Bryson Taylor and Neil Vigdor, *'I Can't Breathe': 4 Minneapolis Officers Fired After Black Man Dies in Custody*, THE NEW YORK TIMES, (May 26, 2020), https://www.nytimes.com/2020/05/26/us/minneapolis-police-man-died.html.
[248] Nate Cohn and Kevin Quealy, *How Public Opinion Has Moved on Black Lives Matter*, THE NEW YORK TIMES, (June 10, 2020), https://www.nytimes.com/interactive/2020/06/10/upshot/black-lives-matter-attitudes.html.
[249] *See e.g. France backs away from chokehold ban following police protests*, BBC, (June 16, 2020), https://www.bbc.com/news/world-europe-53062357; Eleanor Beardsley, *George Floyd's Death In Minneapolis Forces Change In France*, NPR, (July 21, 2020 5:04 AM ET), https://www.npr.org/2020/07/21/893406647/george-floyds-death-in-minneapolis-forces-change-in-france.

non-profit organization, the National Observatory of Police Practices and Violence,[250] which aims at protecting the interest of family members of victims of police violence and allows users to film police interventions. The images are kept on a secure server and can later be used as evidence in a court of law.[251] The bill further explained that "[*t]he circulation of abusive images and comments against numerous police officers or military police officers very often places them in a climate of insecurity. It has become common for police officers and their families to be threatened, or even followed and assaulted, right up to their front doors*."

A few months later, article 24 of the « *sécurité globale* » bill (global security bill) would have added an article 35 quinquies to the French Press law to make a crime, punishable by, to publish images of on-duty police officers, or members of the military police, with the intent of harming heir "physical or psychological integrity."[252] Professor Evan Raschel noted[253] that such dispositions would favor police anonymity, even though a 2008 administrative note (*circulaire*) of the Ministry of Police clearly stated that police officers did not benefit from a special protection of their images[254] After protests in France and the

[250] *Observatoire national des pratiques et des violences policières* (ONPVP).
[251] Anne Brigaudeau, *Quatre questions à propos de l'application Urgence violences policières*, France TV INFO, (Originally published march 14, 202007:00, Updated April 2, 2020 10 :36), https://www.francetvinfo.fr/faits-divers/police/violences-policieres/quatre-questions-a-propos-de-l-application-urgence-violences-policieres_3861427.html.
[252] "*Is punished by one year of imprisonment and a fine of 45,000 euros the dissemination, by any means whatsoever and whatever the medium, with the aim of harming his or her physical or psychic integrity, the image of the face or any other element of identification of an agent of the national police or the national gendarmerie other than his individual identification number when acting within the framework of a police operation*."
[253] Evan Raschel, Pénalisation de la diffusion d'images des forces de l'ordre : une proposition de loi inutile et dangereuse, D 2020, 2298.
[254] Circulaire n°2008-8433-0, adoptée par le Ministre de l'Intérieur, de l'Outre-Mer et des Collectivités territoriales le 23 décembre 2008, relative à l'enregistrement et la diffusion éventuelle d'images et de paroles de fonctionnaires de police dans l'exercice de leurs fonctions [Circular n ° 2008-8433-0, adopted by the Minister of the Interior, Overseas Territories and Local Authorities on December 23, 2008, relating to the recording and possible dissemination of images and words of officials police in the performance of their

much-publicized police beating of a Black music producer in Paris, the government announced that article 24 would be rewritten.[255]

This bill aimed at protected the privacy, but also the reparation of the police. Social media platforms may easily be used to harm someone's reputation. We will now examine how this is addressed by the States.

II. Protecting the Reputation of Others

"*Sticks and stones may break my bones, but words will never hurt me*" is rather an optimist statement for social media users dealing with defamation (A), insults (B) or hate speech (C), the later a notion which definitions and limits differs country to country.

A. Defamation

An Australian court awarded in 2013 $105 000 to a music teacher who had been defamed on *Twitter* and *Facebook* by a young man who wrongly believed that the teacher had been responsible for the early retirement of his father.[256] The defamed music teacher had replaced the defendant's father as a teacher after he had to step down due to ill health. The defamed teacher was so affected by these defamatory social media posts that she had to immediately take a leave from her teaching position after having become aware of them.

---

duties], cited by *Point Droit Filmer les forces de l'ordre*, LIGUE DES DROITS DE L'HOMME, https://www.ldh-france.org/wp-content/uploads/2019/04/Point-Droit-Filmer-les-FDO-et-diffusion-enregistrement.pdf (last visited ?), "*The police does not benefit from any particular protection in terms of image rights [...] Freedom of information, whether it is the act of the press or of a private individual, takes precedence over the right to respect image or private life as long as this freedom is not distorted by an attack on the dignity of the person or the secrecy of the investigation or instruction* ".

[255] Aurelien Breeden, *France to Rewrite Contentious Security Bill on Sharing Images of Police*, THE NEW YORK TIMES (Nov. 30, 2020), https://www.nytimes.com/2020/11/30/world/europe/paris-police-beating-charges.html.

[256] *Mickle v. Farley* [2013] WSWDC 292 (Austl.). Available at http://www.austlii.edu.au/cgi-bin/sinodisp/au/cases/nsw/NSWDC/2013/295.html?stem=0&synonyms=0&query=title%28%222013%20NSWDC%20295%22%29 .

When awarding damages, the judge wanted "*to stress that when defamatory publications are made on social media it is common knowledge that they spread. They are spread easily by the simple manipulation of mobile phones and computers. Their evil lies in the grapevine effect that stems from the use of this type of communication. I have taken that into account in the assessment of damages that I previously made."*[257]

Another Australian court awarded in 2014 12,500Australian dollars in damages to a man whose estranged wife had posted on her Facebook page: "*'separated from* [*M.D.*] *after 18 years of suffering domestic violence and abuse. Now fighting the system to keep my children safe*."[258] The judgment is interesting as it narrates the testimonies of the brother and of a female friend of the plaintiff, who stated in court their own doubts about the true personality of "M.D." after reading the Facebook posts, before the court ruled them to be defamatory, even though nothing in the past ever gave them ground to believe that their brother and friend was a domestic abuser. This shows the negative impact a defamatory statement posted on social media may have on someone's reputation, especially conserving how easy it may be to repost it for further publication.

Indeed, it is easy to defame someone on social media as publishing a tweet or a post is quick and simple, and there is no filter applied before the post is made public.[259] It may be difficult to assess the country from which the message was posted, and which

---

[257] *Mickle v. Farley,* paragraph 21.
[258] *Dabrowski v. Greeum* [2014] WADC 175 (Austl.). Available at http://decisions.justice.wa.gov.au/district/disdcsn.nsf/PDFJudgments-WebVw/2014WADC0175/%24FILE/2014WADC0175.pdf .
[259] See Ellyn M. Angelotti, *Twibel Law: What Defamation and Its Remedies Look like in the Age of Twiter*, 13 J. High Tech. L. 430, 466-467 (2013), noting that "*Twitter users are able to publish in a vacuum with no one responsible or assigned to correcting or fact-checking theirs posts."*

defamation law should apply. Defamation laws differ widely around the world, and an U.S. social media user may be sued for defamation in a foreign country. It should be noted that the "Securing the Protection of Our Enduring and Established Constitutional Heritage Act (Speech Act)"[260] protects U.S. residents from defamation judgments obtained in foreign countries by requiring that, for a foreign defamation judgment to be enforced in the U.S., the defamation law applied by the foreign court must provide at least as much protection as the First Amendment to the U.S. Constitution and the constitution of the state in which the domestic court is located. However, the foreign judgment can be enforced, even if the defamation law applied abroad does not provide such protection, if the defendant would have been found liable for defamation by a domestic court applying the first amendment to the Constitution of the United States and the constitution and law of the State in which the domestic court is located. For instance, while defamation is a tort in the U.S., it is a crime in France, as we will now see.

a. USA

Proverb XXII 1: "*A good name is rather to be chosen than great riches*."

There is no federal defamation law, and each state has its own statute. Most of these statutes follow the guidance of the Restatement (Second) of Torts, which paragraph 558 defines the elements of a defamation tort, as such: "(*a) a false and defamatory statement concerning another;(b) an unprivileged publication to a third party; (c) fault amounting at least to negligence on the part of the publisher [with respect to the act of publication]; and (d) either actionability of the statement irrespective of special harm or*

[260] 28 U.S.C. § 4101.

the existence of special harm caused by the publication."[261] Its paragraph 559[262] defines a "*defamatory communication*" as one "*tend*[*ing*]… *to harm the reputation of another as to lower him in the estimation of the community or to deter third persons from associating or dealing with him.*" A defamatory communication may adversely affect adversely the third party personal or financial reputation,[263] but this is not a necessary condition, and communication which may lead to "*social aversion*" are also defamatory.[264]

In New York, courts define defamation as "the *making of a false statement of fact which tends to expose the plaintiff to public contempt, ridicule, aversion or disgrace*."[265] The elements of a cause of action for defamation in New York are : (1) a false statement, (2) published without privilege or authorization to a third party, (3) which is a constituting fault under, at a minimum, a negligence standard, and (4) which had either caused special harm or constitutes defamation per seb.[266] A false statement is a necessary element of a defamation cause of action and, as only facts can be proven false, only statements alleging facts can be the subject of a defamation action.

Under Virginia law, the three elements of the tort of defamation are (1) publication, (2) of an actionable statement, and (3) intent.[267] An "*actionable statement is one that involves a false statement of fact and tends to injure the plaintiff's reputation*."[268] The

---

[261] RESTATEMENT (SECOND) OF TORTS § 558 (1977).
[262] RESTATEMENT (SECOND) OF TORTS § 559 (1977).
[263] RESTATEMENT (SECOND) OF TORTS § 559, comment b.
[264] RESTATEMENT (SECOND) OF TORTS § 559, comment c, giving as examples "*the imputation of certain physical and mental attributes such as disease or insanity*" as "*they tend to deter third persons from associating with the person so characterized*."
[265] Rinaldi v. Holt, Rinehart & Winston, 42 N.Y.2d 369, 379, 397 N.Y.S.2d 943, 366 N.E.2d 1299 (1977], cert. denied 434 U.S. 969, (1977).
[266] See for example Kindred v. Colby, 50 N.Y.S.3d 26, 54 Misc.3d 1205(A), (A).
[267] Wilson v. Miller Auto Sales, Inc., 47 Va.Cir. 153, 161 (Winchester 1998).
[268] Tomlin v. International Business Machines Corp., 84 Va.Cir. 280, 284 (2012).

Virginia Supreme Court held in 1998 that "*speech which does not contain a provably false factual connotation, or statements which cannot reasonably be interpreted as stating actual facts about a person cannot form the basis of a common law defamation action*."[269] U.S. Representative Devin Nunes (R-California) filed a defamation suit in 2019 in Virginia against Twitter and against the anonymous person behind the @DevinNumesMom and the @DevinCow Twitter account, he alleged that the user behind the @DevinNumesMom account had "*falsely accused Nunes "of putting a "Fake News MAGA" sign outside a Texas Holocaust museum*."[270] This is a statement of fact, and its alleged truth can be proven, if it occurred.[271] However, the complaint also alleged that the person behind the @DevinNunesMom "*falsely stated that @DevinNumes is DEFINITEVELY a feckless cunt*" is not a statement of fact that can be proven. It can be proven that this tweet has been posted, but it is not possible to prove that Representative Nunes is what the Twitter user has accused him to be, just as it is impossible to prove it about any other human being. "Feckless cunt" is an insult, not a defamation, and everybody can have his or her private definition of what is necessary to meet the standard of indeed being of "feckless cunt." On June 24, 2020, the defamation lawsuit filed by Representative Devin Nunes against the alleged authors of the two Twitter parody accounts, @DevinNunesMom and @DevinCow, was dismissed by Judge John Marshall from the Henrico County Circuit Court in Virginia.[272]

---

[269] Yeagle v. Collegiate Times, 497 S.E.2d 136, 137 (1998).
[270] Devin Nunes v. Twitter, Inc., Elizabeth L. "Liz" Mair, Mair Strategies, LLC et al, No 19-1715-00, Complaint, at 16.
[271] Truth is a defense in defamation suits. It should be noted that, would the Plaintiff have to prove that a particular had not occurred, he or she would face the bigger hurdle of having to prove a negative fact.
[272] Nunes v. Twitter, Inc., CL19-1715-00 (Va. Cir. Ct. June 24, 2020). See Eric Goldman, *Section 230 Protects Twitter from the "Devin Nunes' Cow" Lawsuit–Nunes v. Twitter*, THE TECHNOLOGY AND MARKETING LAW BLOG, (June 25, 2020), https://blog.ericgoldman.org/archives/2020/06/section-230-protects-twitter-from-the-devin-nunes-cow-lawsuit-nunes.htm.

Representative Nunes had also named Twitter as defendant, and had sought 250 million dollars in damages, a significant sum, and a threat likely to chill speech. The @DevinNunesMom Twitter account had been suspended by Twitter before the complaint had been filed.[273] Devin Nunes alleged in the complaint that "*Devin Nunes' Mom… is a person who, with Twitter's consent, hijacked Nunes's name, falsely impersonated Nunes' mother, and created and maintained an account on Twitter (@DevinNumesMom) for the sole purpose of attacking, disparaging and demeaning Nunes*."[274] Setting aside the (involuntarily) comic effect of affirming that "Devin Nunes' Mom" is defaming him, it should be noted that the @DevinNunesMom account had a white middle age female as profile image. Whether the image is indeed the likeness of Representative Nunes' mother is relevant when assessing if the account is impersonating her. Twitter's impersonation policy states that accounts "*pos*[*ing] as another person, brand, or organization in a confusing or deceptive manner may be permanently suspended under Twitter's impersonation policy*."[275] Under this policy, "[*a*]*ccounts with similar usernames or that are similar in appearance (e.g., the same profile image) are not automatically in violation of the impersonation policy.*"

Devin Nunes, as an U.S. Representative, is a public official and would have to prove that @DevinNunesMom and @DevinCow made their statement with "*actual malice*" as the Supreme Court famously held in 1964, in *New York Times Co. v. Sullivan*, that public officials may only recover for publication of a defamatory falsehood relating to their official conduct if they prove that "*the statement was made with "actual malice" -- that is, with knowledge*

[273] The suspended account had been since replaced by @NunesAlt, see *Don't forget to tip your cow!*, GOFUNDME, https://www.gofundme.com/f/21x5q5a4ao (Last visited Dec. 30, 2020).
[274] Complaint at 12.
[275] *Impersonation policy*, TWITTER, https://help.twitter.com/en/rules-and-policies/twitter-impersonation-policy, (Last visited Dec. 30, 2020).

that it was false or with reckless disregard of whether it was false or not."[276] Justice Brennan wrote that there is "a *profound national commitment to the principle that debate on public issues should be inhibited, robust, and wide-open, and that it may well include vehement, caustic, and sometimes unpleasantly sharp attacks on government and public officials.*"[277] The Supreme Court extended the actual malice standard in 1967, in *Curtis Publishing Co. v. Butts*, to "*public figures*", who are "*individuals… who do not hold public office…* [*but*] *are nevertheless intimately involved in the resolution of important public questions or, by reason of their fame, shape events in areas of concern to society at large*,"[278] and again, in 1971, in *Rosenbloom v. Metromedia,* to private individuals, if the defamatory speech is about an event of public or general interest.[279] What mattered was the speech, not the quality of the person defamed.[280] However, in 1974, the Supreme Court held in *Gertz v. Robert Welch* that the actual malice standard applies only to public officials and public figures, not to private individuals.[281]

Justice Clarence Thomas wrote in February 2019, in a concurrence in a certiorari denial, that "*New York Times* [*v. Sullivan] and the Court's decisions extending it were policy-driven decisions masquerading as constitutional law*" and called for an end of "*reflexively apply*[*ing*] *this policy-driven approach to the Constitution",* stating that "[*i*]*nstead, we should carefully examine the original meaning of the First and Fourteenth Amendments. If the*

[276] New York Times Co. v. Sullivan, 376 U.S. 254, 279 (1964),
[277] New York Times Co. v. Sullivan, at 270.
[278] Curtis Publishing Co. v. Butts, 388 U. S. 130, 164 (1967).
[279] Rosenbloom v. Metromedia, Inc., 403 US 29 (1971).
[280] As explained three years later by Justice Powell in *Gertz v. Robert Welch,* at 337, Justice Brennan "*abjured the… distinction between public officials and public figures on the one hand and private individuals on the other. He focused instead on society's interest's interest in learning about certain issues."*
[281] Gertz v. Robert Welch, Inc., 418 US 323 (1974).

Constitution does not require public figures to satisfy an actual-malice standard in state-law defamation suits, then neither should we."[282] This statement reflects Justice Thomas adherence to the belief that the Constitution should be interpreted as the forefathers meant it to be, the so-called "originalism" theory favored by the late Justice Scalia. Justice Thomas detailed the history of libel in the United States, from the colonies thereon:

> "[*l*]*aws authorizing the criminal prosecution of libel were both widespread and well established at the time of the founding. …And they remained so when the Fourteenth Amendment was adopted, although many States by then allowed truth or good motives to serve as a defense to a libel prosecution*."

In these days, public figures were more protected by defamation laws not less, notes Justice Thomas, citing Blackstone: "*Words also tending to scandalize a magistrate, or person in a public trust, are reputed more highly injurious than when spoken of a private man*". Justice Thomas noted, however, that common law provided the right to comment "*on public questions and matters of public interest*," a privilege which extended to the "*public conduct of a public man*," which was a "*matter of public interest*" that could "*be discussed with the fullest freedom*" and "*made the subject of hostile criticism*." Justice Thomas ended his concurring opinion by stating that "[*t*]*he States are perfectly capable of striking an acceptable balance between encouraging robust public discourse and providing a meaningful remedy for reputational harm. We should reconsider our jurisprudence in this area*." Considering that the States are now customarily being labeled as being "red" or "blue", but that speech can be instantly published on social media in every single state, without regard

[282] Kathrine Mae McKee v. William H. Cosby, Jr., 586 U. S. ____ (2019).

of the political opinion of the majority of its citizens, a single state could by itself establish the limit of what be said about a public figure. This state standard would become universal by the virtue of social media use, since a social media message published in one state could be viewed, and tried, in another.

Were we to revert to a pre-*Sullivan* jurisprudence, defamation suits filed by public figures alleging that they had been defamed on social media would likely flourish. This would be particularly concerning as merely "retweeting" a message may be considered defamatory. The Second Circuit quoted the Restatement, Second, Torts § 578 , which states that "*one who repeats or otherwise republishes defamatory matter is subject to liability as if he had originally published it* " in *Cianci v. New TimesPublishing Co*.,[283] a case which, in turn, was quoted by the Southern District of New York in 1991 in *Cubby, Inc. v. CompuServe, Inc*.,[284] one of the cases which led Congress to pass passing the Communications Decency Act (CDA).[285] Section 230 of the CDA provides a safe harbor to social media platforms, which, as providers of interactive computer services are not treated "*as the publisher or speaker of any information provided by another information content provider*,"[286] but not to users republishing a defamatory statement posted by a third party.

Justice Brennan argued in *Rosenbloom* that "[*d*]*rawing a distinction between "public" and "private" figures males no sense in terms of the First Amendment guarantees"* and that the *New York Times* standard was applied to encourage debate of public issues, not because

[283] Cianci v. New Times Pub. Co., 639 F. 2d 54 (2nd Circ. 1980).
[284] Cubby, Inc. v. CompuServe Inc., 776 F. Supp. 135 (S.D. New York 1991).
[285] Communications Decency Act of 1996, (CDA), Pub. L. No. 104-104 (Tit. V), 110 Stat. 133 (Feb. 8, 1996), codified at 47 U.S.C. §§223, 230.
[286] 47 U.S. C. § 230 (c)(1).

the public official has less interest in his reputation than a private citizen.[287] Would *Rosenbloom* not have been rejected by *Gertz*, or would the *New York Times* standard be abandoned entirely, as wished by Justice Thomas, private citizens using social media to express their opinions about a public official or a public figure could face defamation lawsuits where plaintiff would not need to prove knowledge of the falsity of the speech or reckless disregard of whether it are false or not. Could social media, which offers users a free, easy to use, platform and allows a message to instantaneously reach millions of individuals around the globe, be one day lead to the demise of the *New York Times* standard? Writing for the majority in *Gertz*, Justice Powell had noted that:

> "[*t*]*he first remedy of any victim of defamation is self-help—using available opportunities to contradict the lie or correct the error and thereby to minimize its adverse impact on reputation,* noting further that "[*p*]*ublic officials and public figures usually enjoy significantly greater access to the channels of effective communication and hence have a more realistic opportunity to counteract false statements than private individuals normally enjoy. Private individuals are therefore more vulnerable to injury, and the state interest in protecting them is correspondingly greater*."[288]

Many private individuals are social media celebrities, some having thousands of followers, often more than their elected local officials. Social media may also create its own public figures, which may only be famous on these platforms, and famous only to their followers. Should the *New York v. Sullivan* standard be applied to them? Public officials are held to this standard when suing for defamation, and public figures are deemed to be such

[287] Rosenbloom, at 46.
[288] Gertz, at 344.

figures "*by reason of the notoriety of their achievements or the vigor and success with which they seek the public's attention*,"[289] so it could certainly be argued.

b. France

French law professor Bernard Beignier wrote about defamation: "*the law condemns it; public morals accept it*."[290] How does French law condemn it?

Defamation is a crime in France, not a tort but proposals to decriminalize defamation are regularly made. For instance, on March 30, 2008, the Commission Guinchard, presided by a renowned French Law Professor Serge Guinchard, published sixty-five proposals to reform the French justice system.[291] The report noted that 91% of press offenses tried in 2006 under the French Press Law were public insults or defamations and its Proposal number 12 recommended to decriminalize defamation to make it a civil tort, except for defamation with a discriminatory nature, such as racist or sexist discrimination. The report further noted that:

> "*complete decriminalization would allow France to bring its legislation in line with Resolution 1577 adopted on 4 October 2007 by the Parliamentary Assembly of the Council of Europe*," which called on the Member States "*to ban their legislation on defamation any enhanced protection of public figures, in accordance with the case law of the Court, and invite*[*d*] *France in particular to revise its* [*Press Law] in the light of caselaw of the* [*European Court of Human Right*]. Paragraph 13 of the Resolution

[289] Gertz, at 342.
[290] "*La loi la condamne, les moeurs la supportent*." Bernard Beignier, L'HONNEUR ET LE DROIT, LGDG, 1995, p.184.
[291] Serge Guinchard, *L'ambition raisonnée d'une justice apaisée*, LA DOCUMENTATION FRANÇAISE, (June 30 2008), https://www.vie-publique.fr/sites/default/files/rapport/pdf/084000392.pdf.

1577 exhorts "*states whose laws still provide for prison sentences – although prison sentences are not actually imposed – to abolish them without delay so as not to give any excuse, however unjustified, to those countries which continue to impose them, thus provoking a corrosion of fundamental freedom*."

So, at least for now, defamation is still a crime in France. Article 29 §1 of the French Law on the Freedom of the Press of July 29, 1881 (*Loi sur la liberté de la presse*, French Press Law)[292] defines public defamation as a statement alleging or attributing a fact detrimental to the honor and defines defamation as "*any allegation or imputation of an act that offends the honor or the consideration of the person or the body to which the act is imputed*." The action is open to individuals ("*personne physique*" and to entities ("*personne morale*.") French law distinguishes, however, defamation of a private person, article 32-1 of the French Press Law, from defamation of courts, armies, constituted bodies, public administration, punishable by a 45 000 Euros fine by article 30 of the French Press Law.

The crime of defamation has a short statute of limitations, only three months (article 65 of the French Press Law), which runs from the day of the commission, as the courts reasoned that it then published and thus made public.[293] However, this short statute of limitations can be renewed by a procedural act of investigation or prosecution and the statute of limitations for racial press crimes, such as provocation to racial discrimination, is one year. Both of these statutes of limitations are nevertheless much shorter than the statute of limitations for common misdemeanors, which, under article 8 of the French

[292] Loi du 29 juillet 1881 sur la liberté de la presse [Law of 29 July 1881 on freedom of the press], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF France], July 30, 1881.
[293] *See* for instance, Cass crim. Oct 13, 1987.

criminal procedure Code, is three years from the day of the commission of the offence, if, during that time, no investigation or other prosecution steps were taken.

To benefit from this favorable regime, speech has to be published, that is, to have be made public through the various means enumerated by article 23 of the French Press Law, by "*speeches, shouts or threats uttered in public places or public meetings or by written or printed matter, drawings, engravings, paintings, emblems, images or other support of writing, of speech or of image, which are sold or distributed, offered for sale or displayed in public places or meetings, or by posters or posters publicly displayed, or by any means of electronic public communication*." The *Cour de cassation*, France's highest civil and criminal court has held that "*electronic means*" includes the Internet,[294] and thus includes speech published on social media. However, not every social media post is public. To be considered non-public, abusive speech must have made in a private correspondence or among people sharing a "*community of interests*," which is defined by the *Cour de cassation* as " *a group of persons related by common ownership, aspirations and shared objectives, forming an entity sufficiently close as to not be regarded as being third parties in relation to the author of the speech at stake."*[295] We will see that, in an insult on social media case, the *Cour de Cassation* confirmed the ruling of a Court of appeals who had held that these insults had not been made publicly.[296] If the speech is published on the Internet, the three months period open to file a suit runs from the day of the first online publication. The *Tribunal de Grande*

[294] Cour de Cassation [Cass.] [supreme court for judicial matters] crim. May. 6, 2003, Bull.crim, No. 94 (Fr.).
[295] Cour de Cassation [Cass.] [supreme court for judicial matters] crim. April. 28, 2009, No.08-85.249, not published, but available at http://www.legifrance.gouv.fr/affichJuriJudi.do?oldAction=rechJuriJudi&idTexte=JURITEXT000020655951&fastReqId=1530926204&fastPos=1.
[296] Cour de Cassation [Cass.] [supreme court for judicial matters] 1e civ. April. 10, 2003, Bull.civ. I, No. 70 (Fr.).

Instance de Paris, the Paris trial court, held however on March 18, 2013, that creating a hypertext link allowing direct access to an article previously published is a new publication of the speech to which this hyperlink directs.[297] The *Cour de cassation* specified in 2017 that since "*any reproduction of a text already published constitutes a new publication of the said text which gives rise to a new limitation period… a new publication of a content previously posted on a website which owner has voluntarily reactivated said site after having disabled it constitutes such a reproduction*."[298] It is easy to retweet, and, by a push of a button, an entire tweet is republished on one's page, with the option to add commentaries.

There are three elements in defamation. Two of these elements are acts, that is (1) publicly alleging or attributing a fact which (2) is detrimental to the honor of a person or a group, while the third element is (3) the mental state, having known that the statement would harm the honor or the reputation of the person or the body. To be defamatory, speech must precisely articulate facts attributable to the plaintiff and must be able to be proven, without difficulty, following a contradictory debate. This is not the case, for instance, for the term "neo-Nazi", even if adding the adjective "notorious."[299] Such expression is considered under French law to be an insult, not a defamation. In a case where a promotional booklet published to promote the first album of a French rap group had harshly criticized the French police, the Plenary Assembly *(Assemblée Plénière)* of the *Cour de Cassation* held on June 25, 2010, that since the incriminated writings "*did not*

[297] Tribunal de grande instance [TGI] [ordinary court of original jurisdiction] Paris, March 18, 2013, available on LEGALIS.NET at http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=3664.
[298] Cour de cassation [Cass.] [supreme court for judicial matters] Crim, Feb. 7, 2017, n°15-83439, https://www.legifrance.gouv.fr/affichJuriJudi.do?oldAction=rechJuriJudi&idTexte=JURITEXT000034038323&fastReqId=1266105663&fastPos=2
[299] Cour de Cassation [Cass.] [supreme court for judicial matters] Crim., Feb. 14, 2006, n° 05-82.475.

alleged any specific fact, likely to be proven without difficulty by evidence or by adversarial hearing, the Court of Appeal had rightfully concluded that these writings, if they were indeed insulting, did not constitute the offense of defaming a public administration."[300] However, allegations, which do not go beyond "*the exercise of the right of free criticism*" and which are based on elements which are the result of a serious investigation, are not defamatory.[301] In that case, the investigation has been made by a journalist, but it does not appear that the *Cour de cassation* wanted to restrict this privilege to journalists. What counts is whether the allegations were made after a "*serious investigation*." Not many social media posts are being posted after a "serious investigation", unless they refer to an article or a book, and repeat the alleged defamatory statements.

Presenting the defamatory facts as being alleged is not a defense. Truth is, however, a defense, and so is good faith; these two defenses are distinctive.[302] Using "*interrogative, negative, conditional, doubtful or an euphemism*" is not a defense either as any expression containing contains an imputation of a specific and determinative fact of such a character as to bring to the honor or consideration of the person concerned constitutes defamation even if it is presented in disguised or dubitative form or by innuendo.[303] The regime of evidence in defamation cases differs from the usual evidence regime. The French criminal system is inquisitorial, not accusatory as in the U.S., where both parties play an active role gathering evidence. In France, it is the *juge d'instruction*, pursuant to article 81 of the

[300] Cour de Cassation [Cass.] [supreme court for judicial matters] Ass. Plen., June 25, 2010, Bull.crim. I, No. 1 (Fr.). "*To constitute defamation, the allegation or imputation ... must be in the form of a precise articulation of facts which, without difficulty, is the subject of proof and contradictory debate.*"
[301] Cour de cassation [Cass.] [supreme court for judicial matters] 1e civ. Nov. 29, 2005, n° 04-17.957.
[302] Cour de cassation [Cass.] [supreme court for judicial matters] crim., May 24, 2005, n° 03-86.460
[303] Cour de cassation [Cass.] [supreme court for judicial matters] crim., Nov. 23, 2010 n° 09-87.527.

criminal procedure Code, who is in charge of the pre-trial discovery process, "*in accordance with the law of all information acts he deems appropriate for finding the truth* " which adds, in a famous formula, that the *juge d'instruction* is doing so "*à charge et à décharge*," meaning that the *juge d'instruction* is examining the evidence regardless of whether they would tend to charge the defendant or to exonerate him from the crime with which he has been charged. However, proving the truth of a defamatory comment is the sole responsibility of the defendant, and he or she cannot rely on the *juge d'instruction*. The *Cour de Cassation* stated clearly in 1992 that:

> "*pursuant to articles 35 and 55 of the [French Press Law], the truth of the defamatory matter is only a defense to defamation inasmuch as the proof is given by the accused in accordance with the provisions laid down therein; that evidence can only result from an adversarial process before the trial court, and it is not the duty of the investigative judges to search, or receive that proof, as it would be an abuse of their power*."[304]

Truth is a defense under article 35 of the French Press Law, which lists several instances where the plaintiff cannot prove the truth of the defamatory statement. These exceptions had been to added article 35 by May 6, 1944 Ordinance, at a time when a provisional government, led by Charles de Gaulle, had just took over the Vichy government after World War II.[305] The Ordinance listed three exceptions to article 35: truth could not be a defense if the allegation was about the private life of the person defamed, if the allegation referred to facts older than ten years, or if the allegation "*referred to a fact*

[304] Cour de Cassation [Cass.] [supreme court for judicial matters] crim., May 26, 1992, Bull. crim, No. 212 (Fr.).
[305] Ordonnance du 6 mai 1944 relative à la répression des délits de presse [Ordinance of May 6, 1944 regarding the repression of publishing offences, JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF France], May 20, 1944, p. 402.

constituting an amnestied or prescribed offense, or which has resulted in a conviction expunged by rehabilitation or revision."[306] Of these three exceptions, however, only the exception about private life now remains. The exception for facts older than ten years is now abolished. The *Cour de Cassation* had noted in its 2010 Annual Report that the defense by article 35 to prove the veracity of an alleged defamatory statement as a defense if the facts referred in the statement were older than ten years put France "*in a delicate position vis-à-vis article 10 of the European Convention on Human Rights*."[307] The Annual Report referred to the November 7, 2006 ECtHR *Mamère v. France case,* where France had been found by the ECtHR to have violated Article 10 of the ECHR. The ECtHR explained in the paragraph 24 of the case that:

> "*it* [*did*] *see the logic behind a time bar of this nature, in so far as the older the events to which allegations refer, the more difficult it is to establish the truth of those allegations. However, where historical or scientific events are concerned, new facts may emerge over the years that enrich the debate and improve people's understanding of what actually happened*."[308]

[306] The purpose of the Ordinance was likely to prevent an outpour of defamation suits, at a time where France sought to rebuilt itself, including the harmony among its citizens, even though who had been a member of the *Résistance*, who had become rich through the black market, and who had denounced Jewish neighbors, was known by all in local communities.
[307] *Rapport Annuel de la Cour de Cassation, 2010,* Paragraph 2.2.2.2.1, COUR DE CASSATION, http://www.courdecassation.fr/publications_26/rapport_annuel_36/rapport_2010_3866/etude_droit_3872/e_droit_3876/droit_savoir_public_3878/droit_savoir_19408.html#2.2.2.2.1 (last visited Dec. 30, 2020).
[308] Mamère v. France, App. 12697/03, Eur. Ct. H.R. The French government had argued in Mamère v. France that the ten-year bar for the *exceptio veritatis* was justified because "*by the need for the law to ensure that the reality of past events could not be challenged without any limit in time*," but the Court was concerned that the law could be used to silence people even if they were commenting on a topic of general interest.

The *Conseil Constitutionnel* ruled in May 2011 that the ten years exception to the *exceptio veritatis* defense violated the French Constitution.[309] The constitutional Court had been seized by the *Cour de cassation* for a priority preliminary ruling on the issue of constitutionality of article 35. By enlarging the scope of the truth defense, the constitutional Council thus favored truth rather than protecting the interest of the individual who had been the topic of the message.[310] The constitutional Council explained that the objective of Article 35 was:

> "*to prevent freedom of expression from re-evoking allegations made in a distant past that harmed the honor and standing of the people affected by them; that the resulting limitation on freedom of expression pursues a goal in the general interest in the quest for social peace*."[311]

However, "*where they refer to events which occurred more than ten years before, this prohibition encompasses without distinction all statements or publications resulting from historical or scientific studies, including where the charges refer to events which were referred to or commented upon within a public discussion of general interest; that, due to its general and absolute nature, this prohibition violates freedom of expression in a manner that is not proportionate with the goal pursued; that, accordingly, it violates Article 11 of the 1789 Declaration of Human Rights*."

---

[309] Conseil constitutionnel [CC] [Constitutional Court] Decision No. 2011-131 QPC, May 20, 2011,
[310] See for exemple, Yves Mayaud, *L'exception de vérité désormais ouverte aux faits de plus de 10 ans,* Revue Sciences Criminelles 2011, 401, explaining how the Press Law originally sacrified honor and reputation to « *socially useful debate*. »
[311] Conseil Constitutionnel [CC] [Constitutitional Court], decision no. 2011-131 QPC, May 20, 2011, available in English: http://www.conseil-constitutionnel.fr/conseil-constitutionnel/english/priority-preliminary-rulings-on-the-issue-of-constitutionality/decisions-of-the-constitutional-council-qpc/decision-no-2011-131-qpc-of-20-may-2011.103662.html

In a defamation on Twitter case, a French journalist had invited women, a few days after the revelations by the *New York Times* that producer Harvey Weinstein had "*paid off sexual harassment accusers for decades*,"[312] to denounce their own sexual harassments at work and creating to that effect the hashtag #balancetonporc (denounce your pig). She posted a few hours later a tweet alleging that a French manager had sexually harassed her and stated in the tweet what he had allegedly told her.[313] The manager sued the journalist, claiming that the tweet alleged that he had sexually harassed her, which is incriminated by article 222-33 of the French criminal Code, defining sexual harassment as "*repeatedly imposing on a person words or behaviors having a sexual or sexist connotation which either undermine his or her dignity because of their degrading or humiliating character, or create an intimidating, hostile or offensive situation against this person.*" The Paris Court of first instance (*Tribunal de Grande Instance de Paris*) found the tweet to be defamatory, as it was not sexual harassment as de fined by article 2222-1 of the criminal Code. The manager had recognized, in an article published in French newspaper *Le Monde*, to have said some of these words, that he had complimented the journalist on her "big breasts" during a cocktail party with alcohol, and that he had added "ironically", after the journalist refused his advances, that he would have been able to sexually satisfy her all night. However, this is not sexual harassment under French law, which requires, as specified by the Paris Court, "*a*

[312] Jodi Kantor and Megan Twohey, *Harvey Weinstein Paid Off Sexual Harassment Accusers for Decades*, THE NEW YORK TIMES, (Oct. 5, 2017), https://www.nytimes.com/2017/10/05/us/harvey-weinstein-harassment-allegations.html.
[313] He allegedly told her "I like woman with big breasts, I will make you come all night long."

repetition or intense pression." As the journalist had "*lacked prudence in her tweet*" accusing the manger, the court found the tweet to be defamatory.[314]

There is no case law about retweets being found defamatory. An author, Christiane Féral-Schuhl, believes however that a retweet can be defamatory.[315] She reasoned that article 29 of the French Press Law provides that a direct publication or a reproduction of allegations and facts damaging the honor or the consideration of the person or body to which the act is attributed, can be defamatory, even if made in a non-declaratory way (dubitative) way. As such, a retweet "*is indeed a reproduction, [and] it does not matter whether one does not approve of its content*" since even speech made in dubious form can be defamatory. While the ECtHR has not (yet) had the opportunity to rule on whether reposting a defamatory social media post is defamatory, it held in December 2018, in *Magyar Jeti Zrt v. Hungary*, that the Hungarian courts could not rule that inserting a hypertext link was a "*dissemination of defamatory information, automatically entailing liability for the content itself*."[316] In this case, a Hungarian information website had published an article online about supporters of a soccer team allegedly insulted Rom children. The article linked to a *YouTube* video showing a local politician asserting that the supporters were members of Jobbik, an extreme right Hungarian political party. Jobbik had sued the web site for defamation. The European Court of Human Rights identified five

[314] *L'initiatrice de #BalanceTonPorc condamnée pour avoir diffamé sur Twitter l'ancien patron de la chaîne Equidia*, LEGIPRESSE, N°375 (Oct. 3, 2019).
[315] CHRISTIANE FÉRAL-SCHUHL, CYBERDROIT, 2020-2021, §713.115. *Propos diffamatoires sur un réseau social*, Dalloz, 2020.
[316] Magyar Jeti Zrt v. Hungary, Application 11257/16, (Dec. 4, 2018), §76.

relevant aspects when analyzing whether a particular company which had published a hyperlink is liable:

> *"(i) did the journalist endorse the impugned content; (ii) did the journalist repeat the impugned content (without endorsing it); (iii) did the journalist merely include a hyperlink to the impugned content (without endorsing or repeating it); (iv) did the journalist know or could he or she reasonably have known that the impugned content was defamatory or otherwise unlawful; (v) did the journalist act in good faith, respect the ethics of journalism and perform the due diligence expected in responsible journalism?"*[317]

These five aspects could likely be used by courts in the E.U. when assessing whether reporting defamatory social media post makes the social media user liable for defamation as well as the original author of the defamatory message.

B. Insults

It can be argued that determining what constitute an insult is a delicate undertaking and that making insults either a crime or a tort would chill speech.[318]

a. U.S. Law

Should insults be considered unprotected fighting words? Professor Heyman used[319] the *Gomez v. Hug*[320] case as an example of insults which "*amounts to almost physical*

[317] Ibid, §77.
[318] *See* for instance Jones v. Dirty World Entertainment Recordings LLC, 755 F. 3d 398, 407 6th Circ. 2014), while not discussing insults, but "porn revenge", arguing that "*the immunity provided by § 230 protects against the "heckler's veto" that would chill free speech*." I discuss Section 230 of the CDA later on.
[319] STEVEN J. HEYMAN, FREE SPEECH & HUMAN DIGNITY,143 (Yale University Press), 2008
[320] Gomez v. Hug, 645 P. 2d 916 (Kan. Court of Appeals 1982).

aggression."[321] In this case, a member of the Board of County Commissioners of Shawnee County told an employee entering an office: "*You are a fucking spic…. A fucking Mexican greaser like you, that is all you are. You are nothing but a fucking Mexican greaser, nothing but a pile of shit*." Professor Heyman argued that "[*a*] *speaker who uses language like that in Gomez indicates that he cannot be trusted to respect the target's safety*."

Online insults can spill into the real world and make the object of the insult fear for her or his physical safety.[322] "Gamergate" is a particularly egregious episode of online abuse against women. After blogger Eron Gjoni posted about his relationship with former girlfriend Zoe Quinn, implying that the reason of the success of the Depression Quest game she had created was due to her sexual promiscuity, social media posts tagged #Gamergate about the ethics of gaming journalism emerged on Twitter, using #gamergate as hashtag, mostly in an abusive and misogynistic way, using meme, doxing videos and documents to smear Ms. Quinn. Around 2 million messages using the hashtag were sent in September and October 2004.[323] Anita Sarkeesian, a feminist critic publishing the "Feminist Frequency" YouTube channel, was also the target of online threats, including rape and murder threats, most published on Twitter,[324] as were Brianna Wu and Leigh Alexander.[325] To show the

---

[321] Citing ALEXANDER M. BICKEL, THE MORALITY OF CONSENT 72 (1975).
[322] Professor Danielle Keats Citron gives as an example the fate of feminist author Jessica Valenti. Victim of attacks by a hate group, her home address was published on online forums, leading to her use of a fake name when travelling and keeping traveling plans private. See DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE, 7, (Harvard University Press), 2014.
[323] *How the world was trolled*, THE ECONOMIST, (Nov. 4, 2017), 21. The article notes that the mainstream media misinterpreted the issue as a "*serious debate, in which both sides deserved to be heard, rather than a right-wing bullying campaign*."
[324] Nick Wingfield, *Feminist Critics of Video Games Facing Threats in 'GamerGate' Campaign*, THE NEW YORK TIMES, Oct. 15, 2014, http://www.nytimes.com/2014/10/16/technology/gamergate-women-video-game-threats-anita-sarkeesian.html
[325] Sean T. Collin, *Anita Sarkeesian on GamerGate: 'We Have a Problem and We're Going to Fix This'*, ROLLING STONE, (Oct. 17, 2014 1:10PM ET), https://www.rollingstone.com/politics/politics-news/anita-sarkeesian-on-gamergate-we-have-a-problem-and-were-going-to-fix-this-241766.

extent of the threats she regularly receives on Twitter, Ms. Sarkeesian published on her *Tumblr* blog in January 2015 all the threats she had received between January 20, 2015 and January 26, 2015.[326] The post features screen shots of rape or death threats, most of them using crude and vulgar language, tweeting from anonymous accounts.[327] Professor Keats Citron argues that threats "*tell us nothing about victims. They do not constitute ideas that can be refuted unless responding that someone should not be raped amounts to a meaningful counterpoint*."[328] As such, the argument that such speech is protected as part of the marketplace of ideas becomes weaker, when considering that the marketplace is protected so that from unfettered debate emerges truth. Has social media made this a mere romantic idea, which naiveté is obvious when reading the deluges of threats and insults published every day on the platforms? Game developer Zoë Quinn, the first target of #Gamergate, co-founded with Alex Lifschitz the Crash Override site to provide support to online harassment victims. It describes itself as an "Online Anti-Harassment Task Force"[329] The *Games and Online Harassment Hotline* opened in August 2020, aiming at providing free support for women playing or making games.[330]

b. French Law

Article 29 §2 of the French Press Law defines insult as an "*offensive expression, a term of contempt or an invective which does not contain an allegation of fact."* The difference

---

[326] *One Week of Harassment on Twitter*: http://femfreq.tumblr.com/post/109319269825/one-week-of-harassment-on-twitter
[327] Nick Wingfield wrote in his New York Times' article that such "*malice directed recently at women… [is] "invigorated by the anonymity of social media and bulletin boards where groups go to cheer each other on and hatch plans for action*."
[328] DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE,198, (Harvard University Press), 2014.
[329] CRASH OVERRIDE, http://www.crashoverridenetwork.com/, (last visited Dec. 30, 2020).
[330] GAMES AND ONLINE HARASSMENT ONLINE, https://gameshotline.org. (last visited Dec. 30, 2020).

between defamation and insults is that the insults does not alleges any fact. The French Press Law incriminates both public insults, article 33 of the French Press Law, and private insults, article R. 621-2 of the criminal Code. Insulting someone on social media can be made publicly, for instance on a public Twitter account, or privately, if using a social media account open "by invitation only." Let's look first at a public insult case, which became a *cause célèbre* in France as the defendant was a Parisian attorney, who has been blogging for years under the pseudonym of 'Maître Eolas'. He has been active online for many years, first writing a blog, then becoming an active member of Twitter. He published two messages on Twitter, on November 8 and 9, 2011, to comment on an online petition circulated by a non-profit, l'*Institut pour la Justice*, about candidates to the upcoming French presidential elections. The tweets read: "*L'institut for Justice is reduced to using bots to spam on Twitter to promote its latest turd?*" and " *I would wipe myself off with the Institute for Justice if I weren't afraid of getting my poo dirty*." As such, Maître Eolas criticized the content of the petition and the way the petition votes were counted, casting doubt about the integrity of the process. The non-profit organization sued him for public insults and won, and a Court of appeal upheld the sentence. The *Cour de cassation*, however, did not approved the reasoning of the Court of appeal and did not order the remand of the case.[331]

The *Cour de cassation* reasoned:

"*… in thus determining, even though [the Court of appeal*] *had judged … the remarks of which it declared the defendant guilty to be part of … a public debate of general*

[331] Cour de cassation [Cass.] [supreme court for judicial matters] Crim., Jan. 8, 2019 n° 17-81.396, *Me Eolas (pseudonyme) et l'association Institut pour la justice*.

interest, the invective included [*in the two tweets*] *… responded spontaneously to a question of an Internet user about the ideas defended by the* [*non-profit organization*] *and this, on a social network imposing concise answers, and, whatever the coarseness and the virulence of the terms used, they did not tend to reach people in their dignity or their reputation, but expressed the opinion of their author in a satirical and schoolboy humor way, as part of an open controversy over the ideas advocated by an association defending a conception of justice opposed to the one which the accused, as a practicing attorney and public debater, himself intended to promote, so that despite their outrageousness, such remarks did not exceed the admissible limits of freedom of expression in a democratic country, the court of appeal disregarded the meaning and scope of* [*article 10 of the European Convention on Human Rights*] *and* [*freedom of speech*].

A few years earlier, the *Cour de cassation* approved on April 10, 2013 the Paris Court of appeals for having found that the posts made by a woman on her Facebook account and on the MSN instant messaging service were not public insults.[332] The employee had stated in no uncertain terms her dislike and contempt of her supervisor[333] in messages who could

[332] Cour de cassation [Cass.] [supreme court for judicial matters] 1e civ, April 10, 2013, Bull. civ. No. 344. In his comment about the case, Professor Cédric Manara wonders how the manager, who was not a "friend" of her employee, had become aware of the messages, and concludes that it was probably one of the friend (or foe?) of the employee who had alerted the supervisor, see Cédric Manara, *Des injures proférées sur des réseaux sociaux ne sont pas forcément publiques*, DALLOZ ACTUALITÉ, (April 24, 2013), https://www.dalloz-actualite.fr/breve/des-injures-proferees-sur-des-reseaux-sociaux-ne-sont-pas-forcement-publiques#.X6vx9e1OmUk (protected by a paywall, on file with the author).

[333] The employee wrote: ": "'Sarko [then French President Nicolas Sarkozy] should vote a law to exterminate pain in the ass managers like mine!!!' (MSN site); "Let's eliminate our bosses and especially our (badly fucked) bosses who are corrupting our lives !!! (Facebook); "Rose X motivated more than ever not to accept it. I am sick of cunts."

only be read by people invited to have access to them.[334] The Court of appeals, following settled case law,[335] held that the employee's accounts formed a "community of interest") (*communauté d'intérêt*) and therefore the insults had not been public. If the insult is published on a social media site, does this necessarily make it public? The employee had created on Facebook a private discussion group named "*extermination of pain in the ass managers*" which had only fourteen members. The manager and the agency filed a suit after the (by then former) employee, for public insults, but lost in the district Court of Meaux, which found that "*access to information posted online was limited to selected members, in very limited number, members who, considering the mode of selection, by friendly or social affinities, formed a community of interest*" which was private, not public.[336] On appeal, the Paris Court of appeals found that only the manager had been targeted, not the agency, and confirmed the first judgment, noting that the defendant had to authorize individuals to become members of her private Facebook group. The MSN group was even a secret one and did not appear when clicking on the user's profile of its members. Appellant had argued, unsuccessfully, that the discussion group was public, citing a 1935 *Cour de cassation* case, where the Court had found that a little interior courtyard was public enough for the made there to be considered public under the French Press Law, adding that Facebook Rules are like the administrative rules which regulated access to this

[334] For instance, she posted on Facebook: "*Let's eliminate our male bosses and especially our female bosses (sexually frustrated) who spoil our lives!!!".*
[335]TGI Paris, April 30, 1997, Gaz. Pal. 1997 1, somm. 257 and TGI Paris, July 10, 1997, Gaz.pal. 1998, 1.59, cited by Paris Court of Appeals, CA Paris, Pôle 2, Chamber 7, March 9, 2011, n° 09/21478. The Paris district Court (Tribunal de Grande Instance de Paris) held that disseminating defamatory remarks on the Internet to an indeterminate number of people not bound by a community of interest made these remarks public as soon as it had been made available to the site's potential users. *See* also Tribunal de grande instance [TGI] [ordinary court of original jurisdiction],Paris, Oct. 25, 1999, available at https://juriscom.net/wp-content/documents/tgiparis19991025.pdf.
[336] TGI MEAUX, Oct. 1, 2009, n° 09/01985.

courtyard.[337] The *Cour de cassation* approved the Court of appeals which had found that these messages were private, as only people approved by the former employee had access to these messages and thus formed a private community of interest, and insults published on a social media account which can only be accessed by a few people are not public. The petitioner unsuccessfully argued that the speech was public, because its recipients whatever their number, did not form between "*a community of interest*," which she defined "*as a group of people linked by common membership, aspirations and shared goals*."

If the insult was directed at a person or a group of persons because of their origin, their membership or non-membership in an ethnic group, nation, race or religion, or because of their gender, sexual orientation or disability, then the insult is punishable by six months' imprisonment and a 22,500 Euros fine, and the court may also order publication of the decision. The 17th criminal Chamber of the Paris court of first instance found on June 2, 2016, that a tweet published in July 2014 which likened Ms. Christiane Taubira, then France's Minister of Justice, to a bonobo ape, was racial insult.[338] The tweet expressed solidarity with Anne-Sophie Leclère, a former member of the French extreme-right party Front National, who had been sentenced to nine months in jail in July 2014 for having said on television that she would prefer to see Christiane Taubira "*in a tree swinging from the branches rather than in government*" and who had published on her Facebook account a photomontage representing Ms. Taubira as a chimpanzee.[339] In the 2016 Twitter case, the

[337] Cour de cassation [Cass.] [supreme court for judicial matters] crim. May 4, 1935, DH 1935, 349.
[338] *17e chambre correctionnelle du tribunal de grande instance de Paris* [17th criminal chamber of the ordinary court of original jurisdiction] Paris, June 2, 2016, N0. 14218000236 (on file with author).
[339] *Front National politician sentenced to jail for ape slur*, THE GUARDIAN, (July 16, 2014, 13:20 EDT), https://www.theguardian.com/world/2014/jul/16/french-national-front-politician-sentenced-to-jail-monkey-slur-christiane-taubira. Ms. Leclère had been excluded from her political party and sentenced to nine months in jail., but her sentence was however overturned by the Cayenne Court of appeals.

Paris court considered that the montage published on Twitter was meant to be degrading to Ms. Taubira, as "*it meant to put her on the same level of an ape of a primitive hominid*" and that there was a "*racist assumption of Blacks being behind compared to Whites in human evolution*."[340] The fact that the author of the tweet was not the author of the montage was not relevant, as he had appropriated its meaning by posting it along with his own message likening Ms. Taubira to a bonobo ape. The court found that the wording of the tweet "*considerably overpasses the limits granted by freedom of expression even… if expressing political ideas.*" The court, noting that it was not the first time the accused had been sentenced by a criminal court of law, sentenced him to two months in jail without suspending the sentence.[341]

Women are often the target of online hate speech, but the response of the law is not the same in all countries. In France, a man who had posted a call to rape journalist and anti-racism advocate Rokhaya Diallo on Twitter was found guilty in January 2014 of insults and incitement to crime, and given a 2,000 euro fine, of which 1,400 euros were suspended. In a sharp contrast, two people, were sentenced to jail in 2014 in the United Kingdom for having threatening to rape and insulted on Twitter Caroline Criado-Perez[342] and Member of Parliament Stella Creasy. One of the two people sentenced was a woman who had posted sixtenn insulting, harassing, and menacing tweets, while under the influence of alcohol. The

[340] Ms. Taubira is originally from the Guyenne department of France (French Guiana), located on the East Coast of South America.
[341] *See Prison ferme pour l'auteur d'un tweet comparant la ministre de la Justice à un singe Tribunal de grande instance de Paris, 2 juin 2016, n° 14-21.800236*, LEGIPRESSE, 393 (2016).
[342] Ms. Criado-Perez reported to the police that she had received "*about 50 abusive tweets an hour for about 12 hours*", see *Caroline Criado-Perez Twitter abuse case leads to arrest*, BBC (July 29, 2013), https://www.bbc.com/news/uk-23485610. The judgment is available at https://www.judiciary.uk/wp-content/uploads/JCO/Documents/Judgments/r-v-nimmo-and-sorley.pdf (last visited Dec. 30, 2020).

other one, a man, had send twenty tweets, menacing Ms. Criado Perez of rape.[343] The presiding judge, Judge Howard Riddle was quoted as saying that it was "*hard to imagine more extreme threats.*"[344] The authors of the threats had reacted to Ms. Criado-Perez campaign to put a woman on new British banknote. The Bank of England announced that Jane Austen would be featured on the ten-pond banknote,[345] which led to MP Stella Creasy congratulating Ms. Criado Perez. The two individuals had merely pleaded guilty to improper use of a public electronic communications network but were sentenced for their threats to eight weeks in jail, for the man, and to twelve weeks in jail, for the woman. The effect of these heinous actions is, unfortunately, more lasting on the victims. Ms. Criado Perez described to the presiding Judge the psychological and material effects this ordeal had had on her. Ms. Criado Perez was placed in fear by these threats, fearing that the menaces would be carried on. She felt terror when her doorbell ring and had to spend time and money to make herself untrackable. MP Creasy had a panic button installed at home. One person emphasized to the Judge, during the trial, that there was a "*disconnect between the pleasant and articulate person he has seen, and the person* [*the accused woman*] *record demonstrates*." This case led to Tony Wang, then general manager of Twitter UK, to post a personal apology "*to the women who have experience abuse on Twitter and for have they*

[343] "*I will find you and you don't want to know what I will do when I do… kill yourself before I do; rape is the last of your worries*."
[344] *Two jailed for Twitter abuse of feminist campaigner*, THE GUARDIAN, (Jan. 24 2014 11.04 EST), https://www.theguardian.com/uk-news/2014/jan/24/two-jailed-twitter-abuse-feminist-campaigner.
[345] Katie Allen and Heather Stewart, *Jane Austen to appear on £10 note*, THE GUARDIAN, (July 24, 2013 10.30 EDT), https://www.theguardian.com/business/2013/jul/24/jane-austen-appear-10-note.

have gone through,"[346] and to Twitter's decision to provide an abuse report button to its users.[347]

A Pew Research Center survey from Spring 2015 found that while 67% of Americans believed that governments should not be able to prevent speech offensive to minorities, only 46% of people living in the European Union (EU) believed it.[348] 51% of the French believed it, which is not surprising as, under French law, hate speech is often a crime, as we will now see.

C. Hate Speech

We will now examine how the U.S. (a), Germany (b) and France (c) are responding to hate speech online.

a. Does the First Amendment Protect Hate Speech?

Hate speech does not have a legal definition in the U.S., but it can be defined as harmful or offensive speech directed at people because of their ethnicity, religion, gender, or sexual orientations. An expression of bigotry and prejudice, both the scourge and the fuel of social media, hate speech is protected by the First Amendment. In *R.A.V. v. St. Paul*, the Supreme Court stroke down in 1992 the St. Paul Bias-Motivated Ordinance, an "anti-hate speech" law, which forbade to place a burning cross or a Nazi swastika on a public or

[346] @TonyW, Twitter (Aug. 3, 2013, 6:09 AM), https://twitter.com/TonyW/status/363602538022436864.
[347] Alexander Abad-Santos, *Twitter's 'Report Abuse' Button Is a Good, But Small, First Step*, THE ATLANTIC, (July 31, 2013), https://www.theatlantic.com/technology/archive/2013/07/why-twitters-report-abuse-button-good-tiny-first-step/312689.
[348] *Europe More Supportive Than U.S. of Censoring Statements Offensive to Minorities*, THE PEW RESEARCH CENTER, (Nov. 20, 2015), http://www.pewresearch.org/fact-tank/2015/11/20/40-of-millennials-ok-with-limiting-speech-offensive-to-minorities/ft_15-11-19_speecheurope/, (last visited Dec. 30, 2020).

private place if one knew that it would cause "*anger, alarm or resentment in others on the basis of race, color, creed, religion or gender*."

Revenge porn can also be considered hate speech. It can be summarily defined as the online posting of naked or sexual pictures of a person for personal revenge. It is often the deed of jilted men (sometimes women) who seek revenge by humiliating their former spouse, girlfriend, or boyfriend. The images may have been originally obtained with the consent of the victim who may have send them during the relationship, with the promise, explicit or implied, that they would remain private, or may have been obtained by surprise[349] or threat. Regardless of the origin of the images, it may be argued, as does a victim of the crime, that "[*r*]*evenge porn is cyberrape*."[350] As a preliminary remark, revenge porn sites are protected by Section 230 of the CDA, which will be discussed further below. It prevents victims of "revenge porn" from having postings removed.[351] For instance, the Connecticut District Court granted in 2019 Tumblr's motion to dismiss in a case where the social media platform had been sued for invasion of privacy and negligent infliction of emotional distress for having unlawfully displayed photographs of plaintiff and failing to

[349] By using a hidden camera, for example.
[350] Rebekah Wells, *The Trauma of Revenge Porn*, THE NEW YORK TIMES, (APRIL 4, 2019), https://www.nytimes.com/2019/08/04/opinion/revenge-porn-privacy.html.
[351] See Doe v. SexSearch. com, 502 F. Supp. 2d 719 (ND Ohio 2007), finding defendants immune from liability for having failed to remove Plaintiff's profile from the online adult dating site and failing to prevent him Defendant from communicating with a 14-year-old girl whose profile he had found on the site, purporting to be 18-year old. Plaintiff had sexual relations with the minor and was subsequently arrested and tried for engaging in unlawful sexual conduct with a minor. While not a revenge porn case, the case is notable as Defendants successfully argued that they were immune from most of Plaintiffs claims under Section 230 of the CDA. Plaintiff had argued that "*the dating website was "an information content provider because [it] "reserves the right, and does in fact, modify the content of profiles when they do not meet the profile guidelines and as such they are responsible in whole or part for the creation or development of the information*." The court noted that "[*t*]*he CDA clearly does not immunize a defendant from allegations that it created tortious content by itself, as the statute only grants immunity when the information that forms the basis for the state law claim has been provided by "another information content provider*." However, as the website "*may have reserved the right to modify the content of profiles in general, Plaintiff does not allege SexSearch specifically modified Jane Roe's profile and is thus not an information content provider in this case*."

remove them from their website.[352] In this case, Plaintiff had dated a man and they had "*exchanged intimate naked photographs of each other*." After the couple separated, Plaintiff deleted the photographs she had of her former boyfriend and blocked his telephone number, Facebook and other ways of communications. However, the former boyfriend posted five nude photos of Plaintiff to myex.com and Facebook. Plaintiff complained to the local police department and the photos were removed. The photos were then uploaded to Tumblr using Plaintiff's name, linked to Facebook and LinkedIn, along with their personal information. Some Tumblr's users re-blogged the photos to other sites with links to her social media accounts. Plaintiff received threatening messages, while other informed her that her nude photos were published on Tumblr. She asked the platform to take the photos down, to no avail. The Court found[353] that Tumblr had satisfied the three statutory requirements for immunity under section 230(c)(1), that is, a finding that defendant is a provider or user of an interactive computer service, that the claim is based on information provided by another information content provider and that the claim would treat the defendant as the publisher or speaker of that information. Plaintiff had argued further that Tumblr had acted in bad faith, but "*failed to provide authority for the proposition that section 230(c)(1) includes a requirement of good faith*."[354]

Most of the U.S. States have now passed 'revenge porn' statutes.[355] For instance, Illinois criminal law[356] incriminates "*non-consensual dissemination of private sexual*

[352] Poole v. Tumblr, Inc., 404 F. Supp. 3d 637 (Connecticut 2019).
[353] Citing FTC v. LeadClick Media, LLC, 838 F. 3d 158, 173 (2nd Circ. 2016).
[354] Poole, at 643.
[355] A list of "revenge porn' state laws can be found at *46 States + DC + One Territory NOW have Revenge Porn Laws*, CYBER CIVIL RIGHTS INITIATIVE, https://www.cybercivilrights.org/revenge-porn-laws (last visited Dec. 30, 2020).
[356] Section 11-23.5(b) of the Criminal Code of 2012 (720 ILCS 5/11-23.5(b).

images," which is defined as "*intentionally disseminat*[*ing*] *an image of another person*" who is at least 18 years old and identifiable from the image, or which information is displayed in connection with the image. The person must be either "*engaged in a sexual act or [having his or her] intimate parts … exposed, in whole or in part*." The image must have been obtained "*under circumstances in which a reasonable person would know or understand that the image was to remain private; and … knows or should have known that the person in the image has not consented to the dissemination*." In 2019, the Illinois Supreme Court found it not to violate the First Amendment.[357] In this case, a woman had been charged under charged of nonconsensual dissemination of private sexual images, as incriminated by 720 ILCS 5/11-23.5(b) (section 11-23.5(b)). She claimed that section 11-23.5(b)restricted speech based on its content and was not narrowly tailored to serve a compelling government interest, in violation of both the U.S. and the Illinois Constitution. The circuit Court had agreed, reasoning that "*when a girlfriend texts a nude selfie to a third party—her boyfriend—she gives up all expectations of privacy in the images. And if she cannot reasonably expect that the image remains private, then didn't the act of sharing it in the first place demonstrate she never intended the image to remain private?"*[358]

The Illinois Supreme Court was, however, not convinced by the argument. It pointed out that the boyfriend was not a third party to the communication, but a second party, and that as such the girlfriend does not relinquish all expectation of privacy in the image. The Illinois Supreme Court first found that there was no categorical exception to protection for

[357] Illinois v. Bethany Austin (Ill. Sup. Ct., Oct. 18, 2019), available at https://courts.illinois.gov/Opinions/SupremeCourt/2019/123910.pdf.
[358] Quoted by the Illinois Supreme Court, at 20.

nonconsensual dissemination of private sexual images, as speech of "*slight social value*"[359] and noting further that "*the Supreme Court has permitted content-based restrictions where confined to the few historic, traditional, and long-familiar categories of expression.*"[360] After determining that such speech is protected, the Supreme Court then determined that these restrictions were subjected to the intermediate level of scrutiny because the Illinois law is a content-neutral time, place, and manner restriction, and because it "*regulates a purely private matter*." The Supreme Court of Illinois found the *City of Renton* case "*instructive*," as the Renton ordinance, a time, place, and manner zoning regulation of adult movie theaters, did not aim at the content of the films projected, but that its "*predominate concerns*" were the "*secondary effects of such theaters on the surrounding community*."[361] Section 11-23.5(b) was justified because it protects privacy, and "*distinguishes the dissemination of a sexual image not based on the content of the image*," but instead on whether its recipient "*obtained* [*it*] *under circumstances in which a reasonable person would know that the image was to remain private and knows or should have known that the person in the image has not consented to the dissemination*." If the image is distributed with consent, there is not criminal liability, and it is the "*manner of the image's acquisition and publication, … not its content*, [*which*] *is… crucial to the illegality of its dissemination*." Therefore, the Illinois statute is a content-neutral law and thus subject an intermediate level of scrutiny. The Illinois Supreme Court then cited *Snyder v. Phelps*[362] to posit that purely private speech is

[359] Quoting R.A.V., 505 U.S. at 383 (quoting Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942)), about speech "*of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality*."
[360] Citing United States v. Alvarez, 567 U.S. 709, 717 (2012) and U.S. v. Stevens, 559 U.S. 468, 470 (2010).
[361] City of Renton v. Playtime Theatres, Inc., 475 U.S. 41 at 47-48 (1986).
[362] Snyder v. Phelps, 562 U.S. 443, 451-52 (2011): "*That is because restricting speech on purely private matters does not implicate the same constitutional concerns as limiting speech on matters of public interest: '[T]here is no threat to the free and robust debate of public issues; there is no potential interference with a meaningful*

less rigorously protected by the First Amendment than public speech. Applying intermediate scrutiny, the Illinois Supreme Court concluded that the Illinois statute serves a substantial government interest of protecting citizens' safety. Victims of "revenge porn" are often "*harassed, solicited for sex, and even threatened with sexual assault" and may "suffer profound psychological harm…[,] feelings of low self-esteem or worthlessness, anger, paranoia, depression, isolation, and thoughts of suicide*."[363] As such the Illinois Supreme Court "*ha*[*d*] *no difficulty in concluding that section 11-23.5 serves a substantial government interest unrelated to the suppression of speech*." The statute is narrowly tailored to promote the government interest of protecting Illinois residents from nonconsensual dissemination of private sexual images, which would be achieved less effectively without the law. The court rejected the argument that copyright can be used to fight "revenge porn," citing an article written by Erica Souza rejecting the contention that copyright is an optimal way for victims to fight 'revenge porn', as it would have the ill-wished consequences of forcing them to make the image public to obtain the copyright.[364] The statute did not burden substantially more speech than necessary either. Its scope is restricted to private, sexual images of a person who must be at least 18 years and recognizable from the image. The image must have been obtained in such way that a reasonable person would know or understand that it was to remain private, a legal requirement which the Illinois Supreme

dialogue of ideas'; and the 'threat of liability' does not pose the risk of 'a reaction of self-censorship' on matters of public import."

[363] At 67.

[364] Erica Souza, For His Eyes Only: Why Federal Legislation Is Needed to Combat Revenge Porn, 23 UCLA WOMEN's L.J. 101, 115-16 (2016): "*So, ironically, to copyright an image and stop strangers from seeing their nude pictures, victims have to send more pictures of their naked body to more strangers (the individuals at the U.S. Copyright Office). Though a successful registration can effectuate a takedown from the identified website, the registered images are sent to the copyright office and appear in the Library of Congress' public catalog alongside copyright owners' names and image descriptions*."

Court interpreted "*as requiring a reasonable awareness that privacy is intended by the person depicted*."[365] Further, the person disseminating the image is aware of his or her transgression, as he or she "*must have known or should have known that the person portrayed in the image has not consented to the dissemination*, " a legal requirement that the Illinois Supreme Court construed " *to incorporate a reasonable awareness of the lack of consent to dissemination*."[366] As the third party must have the intent to disseminate the image, "*the probability that a person will inadvertently violate section 11-23.5(b) while engaging in otherwise protected speech is minimal.*"[367] The Illinois Supreme Court concluded that the law was narrowly tailored to further the important governmental interest identified by the legislature and does not burden substantially more speech than necessary.[368]

U.S. laws do not have comprehensive laws aiming at fighting hate speech. Two European Union countries, Germany, and France, both recently passed laws aiming at fighting online hate speech, with, however, differing level of success.

---

[365] At 81.
[366] At 82.
[367] At 83.
[368] The victim of the revenge porn then unsuccessfully petitioned the U. S. Supreme Court for a writ of certiorari to respond to these two questions: "1. *Whether strict First Amendment scrutiny applies to a criminal law that prohibits nonconsensual dissemination of non-obscene nude or sexually- oriented visual material? 2.Whether the First Amendment requires a law that prohibits nonconsensual dissemination of non-obscene nude or sexually-oriented visual material to impose a requirement of specific intent to harm or harass the individual(s) depicted?*" The Supreme Court denied the petition on October 5, 2020. Several amici curiae briefs were field, asking the Court to grant cert. Amici curiae American Booksellers Association et al argued that content-based restrictions on speech are subject to strict scrutiny, *see* https://www.supremecourt.gov/DocketPDF/19/19-1029/138690/20200319202049168_200307a%20Austin%20%20v%20%20Illinois%20Amicus%20for%20efiling.pdf.

b. The German Hate Speech Law

Then-Justice Minister of Germany Heiko Maas wrote a letter to Facebook's managers in July 2016, stating that the social media network was not keeping its promise made as part of a joint working group to police speech:

> "*So far, however, the result of your efforts has lagged behind what we agreed on in the task force….Too little, too slowly and too often the wrong thing is deleted.*"

A year later, the Network Enforcement Act (*NetzDG*) was adopted by the German Parliament, on June 30, 2017, and by the Bundesrat on July 7, 2017. It entered into force on October 1, 2017, but, as the law gave the platforms three months to put in place the complaint-management system, it started to be enforced only on January 1, 2018. The NetzDG law applies to social networks platforms having two or more million registered users, unless these platforms have less than two million registered users in Germany, or offer "*journalistic or editorial content, the responsibility for which lies with the service provider itself.*" Users must be provided "*an effective and transparent procedure for handling complaints about unlawful content*", which must be "*easily recognizable, directly accessible and permanently available*". Section 3(2)-2 of the law provides that platforms must remove or block access to "*manifestly unlawful*" content within 24 hours of receiving the complaint, or at least within seven days after receiving a complaint about content which is not "*obviously unlawful*." Such "*obviously unlawful*" content is catalogued in the German Criminal Code (StGB), and includes public incitement to crime, violation of intimate privacy by taking photographs, defamation, treasonous forgery, forming criminal or terrorist

organizations, and dissemination of depictions of violence. Failure to take down such speech may lead to a fifty million euros fine.

The decision to delete unlawful content rests with the social network, without any judicial oversight. However, a post which has been removed as illegal under the NetzDG may also lead to prosecution. Such was the case for a tweet posted by Beatrix von Torch, the parliamentary group deputy leader of Germany's far right party *Alternative for Germany,* where she criticized the Cologne police office for having posted a New Year tweet in Arabic, describing the tweet as a way to communicate with "*barbaric, gang-raping Muslim hordes of men*." The delegate's Twitter account was suspended for 12 hours and the Cologne police filed a criminal complaint against von Torch for hate speech. Tweeting again after her account was reinstalled, she posted: "*Facebook has now also censored me. This is the end of the constitutional state*." This example of one of the first messages removed under the NetzDG shows that the societal good for deleting a social media message may not be verified. Instead, it may prevent informing the public about the opinion of a particular group, and deleting a message allows them to later present themselves, on social media, as martyrs of the free speech clause. Indeed, the deputy and the leader of her party were quick to post on Twitter a picture of themselves with their month tapped shut by red tape, their amused glance staring at the viewers, claiming that the Cologne police was investigating them "*because of alleged sedition in a tweet,*" and adding the # freedomofexpression and #censorship hashtags to their post, which claimed "*Genau unser Humor! Meinnungsfreiheit im Iran fordern-and in Deutschland unterhinden*" ("Exactly our humor! Demand freedom of expression in Iran and prevent it in Germany").

The NetzDG law was recently updated by a bill "to fight right-wing extremism and hate crime"[369] which aimed at obliging social media platforms to proactively report serious cases of hate speech to law enforcements. The bill also added to the list of crimes which must be reported to the social media platforms the denigration of the memory of the dead, a decision made following the murder of the Kassel District President Walter Lübcke in 2019 by a far-right individual, and which yielded disrespectful comments online, leading to criminal charges.[370] The bill was approved on June 18, 2020 by the Bundesrat, Germany's lower house.[371]

c. France and Hate Speech

When writing in *The Atlantic* about the debate over the decision of Pen America to award Charlie Hebdo the 2015 PEN/Toni and James C. Goodale Freedom of Expression Courage Award, cartoonist Garry Trudeau stated that "*Charlie wandered into the realm of hate speech, which in France is only illegal if it directly incites violence*."[372] However, stating that hate speech "*is only illegal* [*in France*] *if it directly incites violence*" is no true.

Article 24, §7 of the French Press Law incriminates public incitement to hatred, violence or racial discrimination, but it is far from being the only law making hate speech a

---

[369] *Gesetzentwurf der Fraktionen der CDU/CSU und SPD Entwurf eines Gesetzes zur Bekämpfung des Rechtsextremismus und der Hasskriminalität*, ["A bill to fight right-wing extremism and hate crime of the parliamentary groups of the CDU / CSU and SPD number 19/ 17741], available at https://dip21.bundestag.de/dip21/btd/19/177/1917741.pdf.
[370] *German police raid hate-speech suspects in politician's murder case*, DEUTSCHE WELLE, (June 4, 2020), https://p.dw.com/p/3dGQW.
[371] Janosch Delcker, *German parliament moves to toughen online hate speech rules*, POLITICO, (June 18, 2020, 5:46 PM CET, updated June 18, 2020) 6:12 PM CET), https://www.politico.eu/article/german-parliament-moves-to-toughen-hate-speech-rules.
[372] Gary Trudeau, *The Abuse of Satire, Garry Trudeau on Charlie Hebdo, free-speech fanaticism, and the problem with "punching downward"*, THE ATLANTIC, (Ap. 11, 2015), https://www.theatlantic.com/international/archive/2015/04/the-abuse-of-satire/390312/.

crime. The French Press Law was modified by the July 1, 1972 law on the fight against racism,[373] which added article 32 §2 on *diffamation raciale*, a crime defined as public defamation of an individual because he or she belongs, or do not belong to an ethnic group, nation, race, or religion, whether the plaintiff is correct in his or her assumption or not. Article 33 § 3 incriminates public insult made for the same motives than the one states by article 32 §2. Incitement to hatred, violence or racial discrimination, or public defamation carry each a sentence of one year in jail and a 45,000 Euros fine, while public insults carry a six-month jail sentence and a 22,500 Euros fine. The statute of limitations for these crimes is one year from the date of the publication. If these three crimes are not committed in public, they are incriminated as mere offenses, with a statute of limitations of three months from the date of the speech. While the crime is called *diffamation raciale* (racial defamation), truth is not a defense, of course, as it would lead to acquittal if the defendant could prove that the person target is indeed, say, homosexual, or disabled, and thus render the law completely ineffective. While "hate speech" is not a legal term, "*diffamation raciale*" is one and is punishable even if it does not incite violence.

Is this law still adapted to the generalized use of social media? Paris District Attorney Catherine Champrenault wrote in a June 2019 tribune that the law is no longer adapted to the repression of hate speech.[374] She argued further that the French Press Law had been enacted in 1881 to protect freedom of expression, but that nowadays, "[*i*]*n the*

[373] Loi n°72-546 du 1er juillet 1972 relative à la lutte contre le racisme [Law 72-546 of July 1, 1972 on the fight against racism], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], July 2, 1972, p. 6803.
[374] Catherine Champrenault, *Contre les discours de haine, la loi n'est plus adaptée,* LIBÉRATION, (June 6, 2019, 17:56), https://www.liberation.fr/debats/2019/06/06/contre-les-discours-de-haine-la-loi-n-est-plus-adaptee_1732143.

age of the Internet and social networks, freedom of expression is not lacking in means of dissemination of opinions, but in rules effectively punishing its abuse." Inciting to hate on social media is punishable. However, while it can be observed every day that such speech flourishes on social media, not every illegal post led to an incrimination, a trial and a sentence, and when it does, the sentences are relatively low fines.[375] However, a man who was tried for having posted on Twitter "*Not complicated: as long as you will not accuse the Jews of their numerous crimes, it will be them who will accuse you of theirs*" and also "*The Jews are primary responsible of the massacre of thirty millions of Christians between 1917 and 1947. "* He had been found guilty of both incitation to hatred under article 24, §7 and of *diffamation raciale* under article 32 §2 and was sentenced to 1,000 euros fine, payable to each of the plaintiffs, three non-profit organizations the mission of which is to fight racism and antisemitism, and to two months in jail.[376] The Seventeenth Chamber of the Paris criminal court, the Chamber dedicated to press crimes, noted these the posts, read in the order of their publication, "*tend to arouse a feeling of hatred or rejection vis-à-vis the Jewish community, considered as a whole and without any distinction, having regard only to the ethnic, racial or religious affiliation of its members*" and were incitement to hatred under the French Press Law. Only the second message, alleging responsibility of the Jewish community in a "massacre" was found to be racial defamation. The Court reasoned that "*the use in this sentence of the word "massacre" presupposes the commission of a fundamentally*

[375] Even though a 1,000 Euros, for instance, may appear overwhelming to an individual of low social economic status.

[376] *Deux mois d'emprisonnement ferme pour la diffusion de messages à caractère antisémite sur Twitter,* LÉGIPRESSE, N°339, June 2016), commenting on Tribunal correctionnel de Paris (17ème chambre) (March. 9 2016), N°15023000639. The court sentenced the man to jail because he had been sentenced several times before for similar acts.

inhuman act, which is part of a generalized attack and likely, as such, to fall into a category of crimes punished by international law."

Article 24 § 8 of the French Press Law incriminates public provocation to hate or violence towards a person or a group because of their origins, or because they belong, or do not belong, to a particular ethnic group, nation, race, or religion,[377] and article 24 § 9 of the law incriminates provocation to hate or violence towards a person or a group because of their sex, sexual orientation or handicap. This paragraph was added to the French Press Law by a December 30, 2004 law, which also created the *Haute Autorité de Lutte contre les Discriminations et pour l'Égalité* (Halde), since replaced by the *Défenseur des Droits*. Article 24 § 9 of the French Press Law was the legal basis for three persons being sentenced to a fine[378] in 2015 for having created two Twitter hashtags provoking hate against gays, "gays must disappear because" (#lesgaysdoiventdisparaîtrecar) and "Let's burn gays on…"(#brûlonslesgayssur).[379] All three of them had used these hashtags and finished the sentences in a way which the criminal court found was inciting hate and violence because of sexual orientation.[380] One of them declared to the police that he had published these messages on his Twitter account "*to have a laugh*."

Such provocation to hate, as explained by the Lyon district court (*Tribunal de Grande Instance de Lyon*) in 2009 is a "*positive act of manifest incitement, exhortation or*

[377] This paragraph was also introduced by Law 72-546 of July 1, 1972 on the fight against racism.
[378] One individual was sentenced to a 300 euro suspended sentence fine, as he was present at the trial and apologized for his acts, while two others, who had not done so, were sentenced to a 500 Euros fine.
[379] *Mise en ligne de tweets comportant des hashtags homophobes*, LÉGIPRESSE, N°326, April 2015), commenting on Tribunal correctionnel de Paris (17ème chambre) (Jan. 20 2015), N°13225000272.
[380] They read like an anthology of bigotry: "Let's burn gays on the shit they are", "Gays must disappear because Satan resides in them", etc.

excitement to these actions, attitudes or feelings"[381] targeting a person because of their origins or religion. The case was about an article published in Charlie Hebdo, written by veteran cartoonist and journalist Siné, who mocked the son of then-President Nicolas Sarkozy, alleging that he would convert to Judaism to marry his well-to-do fiancée, adding: "*he will make his way through life this little one*." Siné was sued by the LICRA, the International League Against Racism and Anti-Semitism, under article 24 § 8, but was acquitted. The court cited a comment published in a legal magazine about a similar case, in the same court, which had explained that the "provocation" of article 24 § 8 "*must obviously create, by the explosive charge and the meaning of the words composing it, an incitement to discrimination, hatred or racial violence in the mind of its reader. "*[382] Such was not the case here for the Lyon court, which also cited the ECtHR *Jersild v. Denmark case*, where the Strasbourg Court emphasized how important press is for democracy in its role of "*impart*[*ing*] *information and ideas of public interest*."[383]

[381] Tribunal de grande instance [TGI], [ordinary court of original jurisdiction] Lyon, Feb.24 2009, LICRA v. M. Sinet and Les Editions Rotative, LEGIPRESSE, n°260, April 1, 2009.
[382] TGI Lyon, Feb. 2009, citing Gaz. Pal. 1985, jurisprudence, p. 704.
[383] Jersild v. Danemark, req. N° 15890/89 (Sept. 23, 1994). In this case, a Danish journalist had interviewed members of a racist Danish group for a radio show, and they had expressed racist views during the broadcast. They were found guilty of racial insults as incriminated by the Article 266 (b) of the Penal Code, which provides that ""*Any person who, publicly or with the intention of disseminating it to a wide circle ("videre kreds") of people, makes a statement, or other communication, threatening, insulting or degrading a group of persons on account of their race, colour, national or ethnic origin or belief shall be liable to a fine or to simple detention or to imprisonment for a term not exceeding two year*s." The journalist was found to have aided and abetted the members of the racist group and was fined. The ECtHR found that Denmark had violated Article 10 of the ECHR and "*reiterate*[*d*] *that freedom of expression constitutes one of the essential foundations of a democratic society and that the safeguards to be afforded to the press are of particular importance…. Whilst the press must not overstep the bounds set, inter alia, in the interest of "the protection of the reputation or rights of others", it is nevertheless incumbent on it to impart information and ideas of public interest. Not only does the press have the task of imparting such information and ideas: the public also has a right to receive them. Were it otherwise, the press would be unable to play its vital role of "public watchdog"* (paragraph 31).

Articles 24 § 8 and 24 § 9 do not incriminate having a racist or anti-Semitic opinion opinion,[384] but incriminates inciting racial or religious discrimination. As such, it is similar to the U.S. fighting words doctrine.[385] However, the scope of article 24 § 8 is larger than the fighting words doctrine, as it incriminates provocation to acts of violence, but also provocation to feelings of hate. The Paris district Court explained in 2016, in a case where a man was found guilty to having provoked racial hate on social media, that the article 24 § 8's provocation "*is not necessarily an exhortation, but a positive act of manifest incitement to discrimination, hatred or violence, which does not require an explicit call to the commission of a specific fact.*"[386] As such, it is different from fighting words, which "*very utterance… tend to incite an immediate breach of the peace*."[387] The Paris district Court went on to explain that the words incriminated by article 24 § 8 "*both by its meaning and by its scope… tend to arouse a feeling of hostility or rejection towards a person or a group of people determined by reason of their origin or their membership or of their non-belonging to an ethnic group, a nation, a race or a religion*," but that "*the offense is constituted as soon as the content or scope of the subject is in direct connection with the origin, ethnicity, nation, race or religion*

[384] The French non-profit organization *Ligue Internationale Contre le Racisme et l'Antisémitisme* (LICRA) (International League against Racism and Anti-Semitism) created the formula "Racism is a crime, not an opinion", which is often used when debating the necessity of incriminating racial insults. *See* Jean-Yves Monfort, *Le racisme, le sexisme et l'homophobie ne sont pas des « opinions »*, LEGICOM, No 54, (2015), p.77, writing that this "slogan" has « *the merit of proclaiming, in a collected form, that freedom of expression is not absolute, that it has limits, and that the expression of racism - and today sexism and homophobia - constitutes an unlawful abuse of freedom of expression. »*

[385] *See* for instance Tribunal de grande instance [TGI], [ordinary court of original jurisdiction] Paris, 17th Chamber, March 24, 2002, where the Defendant was found guilty of such provocation. He had presented members of the Jewish community as responsible for the Palestinian situation, adding "*If you see Jewish people on the street, in any continent, beat them up, kill them*."

[386] Tribunal de grande instance [TGI] [ordinary court of original juridiction], Paris, Sept, 7, 2016. Avocats sans frontières, Licra, SOS Racisme v. X., available at LEGALIS, https://www.legalis.net/jurisprudences/tribunal-de-grande-instance-de-paris-17e-ch-correctionnelle-jugement-du-7-septembre-2016

[387] *Chaplinsky v. New Hampshire*, at 572.

[and] reflects on the entire community thus defined. " It is the group of persons of the same religion or ethnicity who is protected, not single members of the group. Such was the case in the 2016 case, where a man had posted on Twitter that there were "*too many blacks"* on the French soccer team, that Jewish people should wear a yellow star, and posted on Facebook that Jewish people should wear a police light on their heads and a rattle, so that even "*the most naïve of goys would know would be warned well in advance of the thing approaching him*." The Paris district Court found that these messages provoked hate, as the Twitter messages:

> "[*were*] *obviously likely to arouse hatred against the Jews, considered as a whole, since they urge readers to stigmatize them (referring to the wearing of the yellow star), that they exploit one of the most worn-out anti-Semitic themes*" and the Facebook message "*pushe*[*d] even further the exhortation to stigmatize Jews by advocating in degrading terms… that they should wear ridiculous and humiliating signs."*

Such crimes are, however, not easy to prove. The provocation must have been public, using one of the means enumerated by article 23 of the French Press Law, that is, "*speeches, shouts or threats uttered in public places or public meetings or by written or printed matter, drawings, engravings, paintings, emblems, images or other support of writing, of speech or of image, which are sold or distributed, offered for sale or displayed in public places or meetings, or by posters or posters publicly displayed, or by any means of electronic public communicatio*n." This last part about public electronic communication, was added to

article 23 by the June 21, 2004, on confidence in the digital economy. As such social media communications are clearly within the scope of the law.

Even if it established that the defendant has indeed publicly uttered hate speech, the *Cour de cassation* checks if his or her intentions were indeed to provoke others to discriminate. In one case, defendant had told plaintiff, during a security check at the Toulouse airport in the Southwest of France: *"If I would have known you sixty years ago in Vichy, I would have burned you up*." The lower court had acquitted defendant from the count of racial insults but had found him guilty of incitement to racial discrimination. The Toulouse Court of appeals confirmed, but the *Cour de cassation* 'broke' ("*casser*") the Court of appeals' decision, as it had not verified if the purpose of defendant's speech was to induce others to discrimination, hatred or violence.

The *Cour de cassation* recently upheld a Court of appeals' decision which had found a man guilty of inciting hate speech, holding that:

> "t*he Court of appeals… has rightly inferred from extraneous elements which it has sovereignly analyzed, that [speech at stake] referred to all immigrants of Muslim religion, and that it had rightly noted that the speech, under the pretext of a legitimate debate on the consequences of immigration and the place of Islam in France, presented all the members of the group thus referred to as "banditry" and "organized crime",…*

[and this speech] aimed, both in its meaning and in their scope, at provoking discrimination, hatred or violence."[388]

In another case, the *Cour de cassation* confirmed on November 8, 2011 a judgment of the Paris Court of appeals which had dismissed a case brought up, under article 24, § 8, by *SOS Racisme*, a non-profit organization dedicated to the fight against racism, against the author and publisher of a book about the Tutsis. The *Cour de cassation* quoted the incriminated part of the book, which presented the Tutsis as a group where a "*culture of lies and concealment*" prevails. However, the *Cour de cassation* held, that, while this speech "*can legitimately shock those for whom it is intended, it nevertheless contain no appeals or exhortation to discrimination, hatred or violence against the Tutsis*."

This reasoning differs from the position of the ECtHR, which considers that:

"*the incitement to hatred does not necessarily require the call to a particular act of violence or other criminal act. Crimes against persons committed by insulting, ridiculing or defaming parts of the population and specific groups…or incitement to discrimination… are sufficient for the authorities to prioritize the fight against hate speech against a freedom of expression irresponsible and detrimental to the dignity or safety of those parties or groups of the population. Political speeches that incite hatred based on religious prejudice, ethnic or cultural represent a threat to social peace and political stability in democratic states.*"[389]

[388] Cass. Crim, September 20, 2016, n. 15-83070, https://www.legifrance.gouv.fr/affichJuriJudi.do?oldAction=rechJuriJudi&idTexte=JURITEXT000033143735&fastReqId=562655858&fastPos=1.
[389] Féret v. Belgium App. N° 15615/07, paragraph 73 (my translation, not an official translation).

French President François Hollande had made the fight against racism and antisemitism a great national cause for 2015.[390] He was succeeded in 2018 by Emmanuel Macron whose government launched a national plan to fight racism and antisemitism.[391] One of the purposes of this plan was to fight hate speech online. It called for a European legislative initiative to require operators to quickly delete illicit content and noted that the "*binary distinction between the publisher's legal system, in which the publisher is civilly and criminally liable because of the content that it publishes, and the host's legal system, in which the host is only liable for illicit content under very limited conditions, is no longer adapted to the huge problem of hate on the Internet."*[392] Following this report, then-Prime Minister Edouard Philippe asked for a report on how to fight racism and antisemitism online. The report was presented in September 2018 by its three authors, among them Representative Laetitia Avia.[393] It proposed, inter alia, to update the *Loi pour la Confiance dans l'Économie Numérique* (LCEN)[394] to make social media companies responsible for hate speech published on their platforms. Such speech would have to be deleted within 24 hours. The report also proposed that companies failing to delete hate speech would face up to 37.5 million Euros in penalties. Inspired by the German initiative of the NetzDG, the report also

[390] *Grande cause nationale pour 2015.*
[391] *Plan national de lutte contre le racisme et l'antisémitisme 2018-2020.* https://www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2018/06/national_plan_against_racism_and_anti-semitism_2018-2020.pdf.
[392] *Plan National de lutte contre le racisme et l'antisémitisme (2018-2020),* p.5, https://www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2018/05/plan_national_de_lutte_contre_la_racisme_et_lantisemitisme_2018-2020.pdf.
[393] Karim Amella, Laetitia Avia, Dr Gil Taïeb, Rapport visant à renforcer la lutte contre le racisme et l'antisémitisme sur Internet (Sep. 20, 2018), https://www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2018/09/rapport_visant_a_renforcer_la_lutte_contre_le_racisme_et_lantisemitisme_sur_internet_-_20.09.18.pdf .
[394] Loi n°2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique [Law 2004-575 of June 21, 2004 for Confidence in the Digital Economy], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF France], June 22, 2004, p. 11168.

proposed to create a special status for the most popular of the social networks and search engines companies, named "content accelerators," which would have enhanced obligations, and further proposed to block access in France to sites which main purpose is to disseminate hateful speech. French law authorizes already such a system, albeit to regulate illegal gambling:

> "*These sites having organized their legal impunity, it is recommended to prohibit access from French territory by intervening at the "network" level, drawing inspiration from existing provisions for blocking illegal online gaming sites at ARJEL initiative."*[395]

Representative Avia then presented on March 20, 2019 a bill aiming a fighting hate on the Internet (the Avia Bill),[396] which incorporated the main recommendations of the September 2018 report to the Prime Minister, including updating the LCEN. Representative Avia argued in her report on the bill to the Parliament (the Avia Report) that a law fighting hate speech online was necessary because "*cyber-hate*" was prevalent in France, admitting however, that she "[*had*] *come up against difficulties knowing the real extent of the phenomenon, in particular because of the weakness and scattering of statistics on the subject*."[397] Indeed, the European Commission Against Racism and Intolerance (ECRI) in its

---

[395] The *Autorité de régulation des jeux en ligne* (ARJEL), the Regulatory Authority for Online Gambling, was created by the LOI n° 2010-476 du 12 mai 2010 relative à l'ouverture à la concurrence et à la régulation du secteur des jeux d'argent et de hasard en ligne [Law 2010-476 of May 12, 2010 on opening to competition and regulation of the online gambling sector], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], May 13, 2010, p. 8881. The ARJEL President of the has the power, under article 61 the May 12, 2010 law, to send a formal notice to the operator of an illegal gambling site reminding them that such activities is a crime punishable by a 100.000 Euros fee.

[396] Proposition de loi nº 1785 visant à lutter contre la haine sur internet [Draft law n ° 1785 aimed at combating hatred on the internet] 9march 20, 2019), available at http://www.assemblee-nationale.fr/dyn/15/textes/l15b1785_proposition-loi#.

[397] Laetitia Avia, *Rapport Fait au nom de la Commission des lois constitutionnelles, de la législation et de l'administration générale de la République sur la proposition de loi après engagement de la procédure accélérée, sur la loi visant à lutter contre la haine sur internet* (n° 1785) (June 19 2019), http://www.assemblee-nationale.fr/dyn/15/rapports/cion_lois/l15b2062_rapport-fond.

Report on France on March 1, 2016, noted that "[*h*]*ate speech has also increased on the Internet and social networks, despite the efforts of the authorities to curb the phenomenon*."[398] The Avia report noted that most of the complaint made on the PHAROS platform were about online hate speech, for instance, 14,000 of such complaints were made in 2018. The French government created the PHAROS platform[399] by an administrative order in June 2009.[400] It is placed under the responsibility of the *Office central de lutte contre la criminalité liée aux technologies de l'information et de la communication*[401] (central Office for the fight against crime related to information and communication technologies) and allows individuals to report illegal online content, such as pedophilia, child pornography, racism, anti-Semitism, xenophobia, incitement to violence, if publicly available and incriminated by French laws. The reports are then processed by police and military police (*gendarmerie*) officers and a criminal investigation may be opened. The Avia Report noted that the platforms themselves take more and more hate speech down following a PHAROS notice and that YouTube and Facebook remove almost 80% of hate

[398] European Commission Against Racism and Intolerance (ECRI) Report on France (fifth monitoring cycle). Adopted on March 8, 2015. Published on March 1st, 2016, p. 9, available at https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/france. https://rm.coe.int/fifth-report-on-france/16808b572d.

[399] *See* https://www.internet-signalement.gouv.fr/PortailWeb/planets/Accueil!input.action (last visited Dec. 30, 2020).

[400] Arrêté du 16 juin 2009 portant création d'un système dénommé « PHAROS » (plate-forme d'harmonisation, d'analyse, de recoupement et d'orientation des signalements [Executive Order of June 16, 2009 creating a system called "PHAROS" (platform for harmonization, analysis, cross-checking and orientation of reports]. JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF France], June 20, 2009, available at https://www.legifrance.gouv.fr/loda/id/JORFTEXT000020763903/2020-11-10.

[401] *Sous-direction de lutte contre la cybercriminalité* , POLICE NATIONALE, MINISTERE DE L'INTERIEUR, (Dec. 13, 2011), https://www.police-nationale.interieur.gouv.fr/Organisation/Direction-Centrale-de-la-Police-Judiciaire/Lutte-contre-la-criminalite-organisee/Sous-direction-de-lutte-contre-la-cybercriminalite.

content reported to them within 24 hours by PHAROS, while Twitter only removes 50% of such content.

The Avia Report argued that the e-Commerce Directive and the French LCEN, which implemented the e-commerce Directive into French law, are not enough to fight online hate speech, as the platforms are thus given "*an editorial role*,"[402] citing the *Conseil d'État*, the Council of State, which, in its capacity as advisor to the government, had written in its *avis* (opinion) on the Avia bill that:

> "*…the emergence of new players (social networks and search engines) who, as active intermediaries allowing content-sharing and by accelerating access through their algorithmic processes of prioritization and optimization, do not limit themselves to a purely technical role, nevertheless without being able to be qualified as content editors, makes the current regime, based on the neutrality of providers of online communication services to the public with regard to content, partly outdated*."[403]

The Avia Bill's statement of purpose stated that "[*s*]*ocial media platforms too often play on the ambivalence of their legal status as hosts to justify their inaction*," thus announcing that the bill was about intermediary's liability and how they should moderate hate speech. The statement of purpose went on:

> "*However, large platforms have a responsibility: that of being able to generate virality around their content, and thereby further expose victims of hate speech. Given the*

[402] Avia Report, I C (b).
[403] Conseil d'État [highest administative court]. Avis n° 397368 du 16 mai 2019 sur la proposition de loi visant à lutter contre la haine sur Internet [Opinion n ° 397368 of May 16, 2019 on the proposed law to combat hatred on the Internet], § 8, available at https://www.conseil-etat.fr/ressources/avis-aux-pouvoirs-publics/derniers-avis-publies/avis-sur-la-proposition-de-loi-visant-a-lutter-contre-la-haine-sur-internet.

importance they have in our daily digital uses, these platforms must better ensure the protection and security of their users. ***This involves restoring the rule of law on the internet and remembering that legislative provisions take precedence over the general conditions of use of each operator. What is not tolerated on the streets or in public space should not be tolerated on the Internet.***" (my emphasis).

These two highlighted phrases indicate that the bill, while well-intentioned and sincere in its goal to fight hate speech, was somewhat disingenuous when presenting the French legal field at the time of the introduction of the bill as devoid of any power to fight hate speech online. We saw earlier that several laws criminalizing hate speech were applicable online.[404]

During the debates. Representative Avia explained that the French laws regulating online speech are based:

"*on an obsolete dichotomy between publishers, with strong responsibility, and hosts, a regime under which all the platforms we know today are placed and which are never worried. While the objectives of the e-commerce directive remain relevant, it never aimed to set up a system allowing the free flow of hate via online communication services. It is therefore our duty to no longer leave this sector in the grip of faltering self-regulation and to fully accept that it is our mission to protect our fellow citizens*

[404] The bill's statement stated indeed that "*the applicable provisions result mainly from the law of June 21, 2004 for confidence in the digital economy,... promulgated almost 15 years ago, at a time when the social networks that we know today were not accessible in France*."

and to decide on the legacy we leave. For my part, I want this heritage to be a virtuous internet."[405]

The main disposition of the bill was its article 1, which would have had online platform operators, as defined by article L. 111-7 I of the Consumer Code,[406] and which activity would have exceeded a threshold of connections from the French territory, to be later determined by decree, would have been required "*to remove or make inaccessible within 24 hours after notification any content manifestly contravening the fifth and sixth paragraphs of Article 24, as well as the third and fourth paragraphs of the article 33 of the law of July 29, 1881 on freedom of the press*." We saw that these articles of the French Press Law incriminate incitement to hatred or an insult that is discriminatory on the basis of race, religion, sex, sexual orientation or disability.[407] The *Conseil Supérieur de l'Audiovisuel* (Superior Council of Audiovisual communications) would have had the power, after issuing a formal notice, to "*impose a financial penalty, the amount of which may take into account the seriousness of the breaches committed and their repeated nature, without being able to exceed 4% of the total worldwide annual turnover for the previous financial year*." Article 4 of

[405] Compte rendu, Commission des lois constitutionnelles, de la législation et de l'administration générale de la République. Examen de la proposition de loi visant à lutter contre la haine sur internet (n° 1785) (Mme Laetitia Avia, rapporteure). Compte rendu n° 88, session ordinaire de 2018-2019 [Report, Commission of the constitutional laws, of legislation, and of the general administration of the Republic. Examination of the bill aimed at combating hatred on the internet (n ° 1785) (Ms. Laetitia Avia, rapporteur) Report n ° 88, 2018-2019 regular session] (June 19, 2019), available at http://www.assemblee-nationale.fr/dyn/15/comptes-rendus/cion_lois/l15cion_lois1819088_compte-rendu#.

[406] The article defines an online platform operator as "*any natural or legal person offering, in a professional capacity, whether paid or unpaid, an online communication service to the public based on …*[c]*lassification or referencing, by means of computer algorithms, content, goods or services offered or put online by third parties;… Or the bringing together of several parties with a view to selling goods, supplying services or exchanging or sharing content, goods or services."*

[407] Article 1 of the bill as modified after the debates, and as sent to the Senate, had a broader scope as it specifically referred to holocaust denial and apology of terrorism. See Senate Bill n°645, available at http://www.senat.fr/leg/ppl18-645.html.

the bill directed the Superior Council of Audiovisual communications to give recommendations to the platforms, but only if necessary ("*en cas de nécessité*"), without defining what would trigger such recommendations. The platforms would also have had, under article 1 of the bill, the obligation to put in place systems providing users a way to contest the removal of the content they had posted,[408] and should also have provided a way for users who had reported content to contest the platform's decision not to remove it. Representative Avia explained during the debates that the purpose of this requirement was to implement a unique reporting button (*bouton de signalement unique*) for all the platforms, and to oblige platforms "*to have adequate human or technological resources*." During the debates, it was added to the bill an article creating a public prosecutor's office and a court specialized in the fight against online hate speech. The bill, as amended during the debates, would also have made it a crime punishable by one year in jail and 250 000 Euros fine, to fail to remove hateful content within 24 hours.

The amended bill was then sent to the Senate, and the Senate rapporteur expressed his reservation about the bill in his report,[409] noting that requiring platform operators to assess whether a particular post is unlawful while having such a short time to make this

[408] We will see further on that Facebook established in 2020 its Oversight Board, which has the power to review content-moderation decisions.

[409] Rapport n° 197 (2019-2020) de M. Christophe-André FRASSA, fait au nom de la commission des lois, déposé le 11 décembre 2019 [*Report n ° 197 (2019-2020) by Mr. Christophe-André FRASSA, made on behalf of the Law Commission, filed on December 11, 2019*], (thereafter Frassa Report), 26-27, available at http://www.senat.fr/rap/l19-197/l19-1971.pdf. It appears that it would have been the platform's legal representative in France which would have had to bear the responsibility, and would have risked being sentenced to jail, *see* Frassa Report p. 28, noting however that the bill was not clear about that point. The *rapporteur* wondered further about the crime's mens rea: "*In the case of the absence of removal within 24 hours of clearly illegal content with regard to the "hate" offenses covered by the proposed law, your rapporteur wonders whether the mere lack of removal will be sufficient to lead to the conviction of the natural or legal person ("obligation of result") or if, more likely, it will be necessary for the prosecuting authority to characterize an absence of normal diligence on the part of the operator in his capacity to qualify the manifest illegality of a content.*" (Frassa Report, p. 29).

decision, when making such decision is difficult even for judges, would likely lead platforms to have a heavy hand when making such decisions, especially as they would be criminally liable if making an error. The *rapporteur* added that:

> "[*o*]*ther perverse effects are also to be feared*," such as "*the increasing use of automated filters; - the instrumentalization of reports by organized pressure or influence groups* ("*digital raids*" *against lawful but controversial content*); *the impossibility of prioritizing, within a uniform cut-off time of 24 hours, which content is obviously the most harmful and must be removed even more quickly - terrorism, child pornography* …[*and*] "*the circumvention of the judge and the abandonment of policing online freedom of expression to major foreign platforms*."

The rapporteur also noted that the European Commission had criticized the bill in no uncertain terms.[410] Indeed, the Commission had informed France that the draft bill would likely breach articles 3, 14 and 15(1) of the e-Commerce Directive. For the Commission, "*the obligations set out in the bill could constitute a restriction to the cross-border provision of information society services, in violation of Article3(2) of the e-Commerce Directive, inasmuch as they would apply to those online platforms established in other Member States*." Such restrictions included having to appoint a legal representative in France or providing a notification mechanism in the language of the user. The French

[410] Notification 2019/412/F, Law aimed at combating hate content on the internet, Delivery of comments pursuant to Article 5(2) of Directive (EU)2015/1535 of 9 September 2015, C(2019) 8585 (Nov. 22, 2019), available at https://ec.europa.eu/transparency/regdoc/rep/3/2019/EN/C-2019-8585-F1-EN-MAIN-PART-1.PDF. The French authorities notified to the Commission in August 2019 about the draft bill, pursuant to Article 5(2) of Directive (EU)2015/1535 of September 9, 2015, which states that "[*t*]*he Commission and the Member States may make comments to the Member State which has forwarded a draft technical regulation; that Member State shall take such comments into account as far as possible in the subsequent preparation of the technical regulation.*"

government had argued that such "*restrictions to the freedom to provide information society services from other Member States would be justified by the objective pursued by the notified draft which is the protection of fundamental rights and, in particular, of human dignity*." This virtuous argument, however, did not convince the Commission, noting that, while article 3(4) of the e-Commerce Directive authorizes member States to derogate to the strict prohibition of its article 3(2) not to "*restrict the freedom to provide information society services from another Member State*," this exception is strictly limited and the measures must be proportionate to the objective pursued, which the Commission deemed not to be the case here.

Articles 12 to 14 of the e-Commerce Directive lay down the principles that intermediary services providers, which include social media platforms, are not liable for information they transmit, store or host for their users.[411] Under article 14 of the e-commerce Directive, services providers are liable for hosting illegal activity or information only if they are aware of it, while article 15(1) of the e-Commerce Directive specifies that they do not have the obligation to actively monitor the information transmitted or stored. For the Commission, the French hate speech bill did not require that the notices had to identify the exact location of the content reported, and the platforms would have therefore search for it, which would be a great burden for them. Also, the notices would not have had to identify which legal requirement has been breached by the content being reported. The Commission found that "*the minimum conditions for notification* [*of the French bill*] *would arguably not be sufficiently precise nor adequately substantiated as to lead to actual*

[411] We will discuss further the e-Commerce Directive later.

knowledge or awareness by online platforms in the sense of Article 14 of the e-Commerce Directive."

During the July 2019 debates at the Senate, the article 1*ter*, which had been introduced by the government by an amendment during the debates at the National Assembly, and which simplified the procedure for notifying illegal content by reducing the information notifiers had to provide,[412] was suppressed. The final vote of the bill took place at the Parliament on May 13, 2020. Article 1-I of the law would have modified article 6.1. of the LCEN to direct platforms to take down, within 24 hours, speech promoting certain crimes, causing discrimination, hatred or violence or denying crimes against humanity, speech which is sexual harassment, incite terrorism or its apology. The platforms have however only one hour to remove terrorist and child pornography speech.[413] During the debates,[414] opponents of the law[415] repeatedly expressed their concerns over the dismissal of the judge in the process in favor of the platforms' accountability mechanism.[416] The

[412] The notifier would had had to only indicate his or her email address, not his or her profession, domicile, nationality, place, and date of birth of natural persons, and the notification would no longer have to indicate the precise location of the content reported, not which legal rules it violates.
[413] This special one-hour requirement was introduced on January 21, 2020 in the bill by an amendment presented by the government, as it is possible for the government to do so under French law. *See* http://www.assemblee-nationale.fr/dyn/15/amendements/2583/AN/161.
[414] For instance, Representative François Ruffin, from the *La France Insoumise* party, had voiced his concern on July 3, 2019, during the first debates at the *Assemblée Nationale*, that the law would lead an « *automatic, algorithmic, robotic censorship, without any human being behind to weigh, decide, without any human being having scruples to silence another human being*." Assemblée nationale, XVe legislature, Extraordinary session, 2018-2019, July 3, 2019, http://www.assemblee-nationale.fr/15/cri/2018-2019-extra/20191003.asp#P1794566.
[415] Both the *Les Républicains* political party, the main right wing party , and the *La France Insoumise* party, on the very left of the spectrum, were against the bill, which was presented and supported by the center-right party created by President Emmanuel Macron*, La République en Marche.* The historical left-wing party, the *Parti Socialiste*, which has lost his influence during the 2017 Presidential elections, curiously abstained from voting, but participated in the debates and criticized the bill.
[416] *See* also Christophe Bigot, *Loi « contre la haine » sur internet : objectif louable mais danger pour la liberté d'expression!* LE FIGARO, (June 18 2019, 13:02, update June 18, 2019 13:09), https://www.lefigaro.fr/vox/politique/loi-contre-la-haine-sur-internet-objectif-louable-mais-danger-pour-la-liberte-d-expression-20190618. The author, a prominent press law attorney, criticized the bill as it would give the main platforms, Facebook, Twitter and Google, the responsibility of "cleaning up" the web under the

platforms failing to delete such speech were facing fines as high as 4% of their worldwide revenue, which would have been given by the French Superior Audiovisual Council (*Conseil Supérieur de l'Audiovisuel*), not by a judge. Such high fines do not nurture subtle decisions, and this would have likely led to over-blocking and consequently chilling of speech.[417] In its final version, article 1-II of the law would also have required platforms to take down, at the request of the administrative authority, terrorist and pedo-pornographic content within one hour, not twenty-four hours as currently required by article 6-1 of the LCEN. Failure to do so would have carried a one-year jail sentence and 250,000 euros fine. Article 1-II of the bill would have directed platforms to take down some speech within twenty-four hours of receiving a notice.[418]

On June 18, 2020, the *Conseil constitutionnel* struck down most of the law.[419] It declared article 1-I unconstitutional in its entirety. It reasoned that the illegality of the content removed would have been exclusively determined by the administrative authority, and the one-hour delay to remove terrorist and pedo-pornographic content would have

---

threat of hefty fines, up to 4% of their worldwide revenue. Me. Bigot further noted that " *the scope of the offense of incitement to racial hatred has changed four times in about fifteen years*, " as sometimes expression giving rise to a feeling of rejection leads to sanction, but sometimes only "exhortation" to commit an act of violence, hatred or discrimination is sanctioned. He wondered how the "*moderators [would] approach these subtle developments, which must be followed day by day…They obviously will not be able to do so and... will apply the precautionary principle,"* thus deleting speech with a heavy hand.

[417] For instance, Representative Ruffin said during the July 3, 2019 debates: "*I have an additional concern: the possibility that it will come to a pre-censorship. I fear that Facebook and company, to avoid trouble, will decide to eliminate, marginalize, refrain from referencing polemical or political content from the start or relegate it. I fear that this will lead to depoliticizing social networks, which will then be limited to kittens and merchandise*."

[418] This included: apology for the commission of certain crimes, incitement to discrimination, hatred or violence with a discriminatory motive, contestation of a crime against humanity, insult committed with a discriminatory motive, sexual harassment, transmission of an image or child pornographic representation, direct provocation to acts of terrorism or apologia for these acts, dissemination of a pornographic message likely to be seen or perceived by a minor.

[419] Conseil constitutionnel [CC] [Constitutional Court] Décision n° 2020-801 DC du 18 juin 2020 Loi visant à lutter contre les contenus haineux sur internet [Decision n ° 2020-801 DC of June 18, 2020 Law to combat hateful content on the internet], June 18, 2020, https://www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm.

been too short, especially since appealing to the judge would not have suspended the delay and thus the platforms would have been able to ask a judge to determine whether the content at stake is illegal or not. As such, "*the legislature has infringed freedom of expression and communication in a way which is inappropriate, unnecessary and disproportionate to the aim pursued*."[420] The *Conseil constitutionnel* also found that article 1-II of the bill violated the Constitution. It noted that the platforms would have to examine whether the content at stake is illegal or not, "*even though the constituent elements of some of* [*the crimes*] *may present a legal technicality or, particularly in the case of press offenses, call for an assessment with regard to the context of the enunciation or dissemination of the content at stake*."[421] The Council noted further that the bill did not provide a safe harbor for the platforms and that the 24-hour delay was "*particularly brief*" in regard to the difficulty in appreciating the illegality of the content.

For the *Conseil constitutionnel:*

> "*… given the difficulties in assessing the manifestly unlawful nature of the content reported within the time limit, the penalty incurred from the first breach and the absence of a specific cause for exemption from liability, the provisions [of the bill] contested can only encourage online platform operators to remove the content that is reported to them, whether or not they are clearly illegal. They therefore undermine the exercise of freedom* of expression *and communication in a way which is not necessary,*

[420] Paragraph 8 of the Constitutional Court Decision.
[421] Paragraph 15 of the Constitutional Court Decision.

appropriate and proportionate. Therefore … paragraph II of Article 1 is unconstitutional."

Only a few articles of the hate speech law were spared by the Constitutional Council decision, such as article 16 creating the *Observatoire de la haine en ligne* (online hate Observatory), which mission is to follow and analyze the evolution of online hate speech.[422] Even though Article 1 of the law has been declared unconstitutional, the Council recognized the value of wishing to monitor this type of content online. It was the method for doing so which has been found unconstitutional. The law, as finally enacted, put the Superior Council of Audiovisual communications in charge to overview the Observatory. In July 2020, it nominated representatives of several platforms to become the first members of the Observatory for two years.[423] Also, it has been announced that there will be a pole dedicated to online hate speech at the office of the Paris District Attorney.[424] French Minister of Justice Éric Dupond-Moretti announced in an interview in November 2020 that he had submitted a new bill aiming at fighting hate speech only to the Conseil d'État. The bill would not modify the French Press Law but would instead add an article in the Code of criminal procedure to allow an individual accused of having published hate speech to be

---

[422] "*An online hate observatory monitors and analyzes the evolution of the content mentioned in article 1 of this law. It brings together operators, associations, administrations and researchers concerned with the fight and prevention against these offenses and takes into account the diversity of audiences, particularly minors. It is placed with the Superior Council of audiovisual, which ensures the secretariat. Its missions and its composition are fixed by the Superior Council of audiovisual*."

[423] Décision n° 2020-435 du 8 juillet 2020 relative à la composition et aux missions de l'observatoire de la haine en ligne [*Decision n ° 2020-435 of July 8, 2020 relating to the composition and missions of the online hate observatory*] [J.O.] [OFFICIAL GAZETTE OF FRANCE], July 24, 2020, p. 59 (available at https://www.legifrance.gouv.fr/jo_pdf.do?id=JORFTEXT000042143892). A representative from Dailymotion, Facebook, Google, Linkedin, Microsoft, Qwant, Snapchat, TikTok, Twitch,Twitter, Wikipedia, and Yubo have been named.

[424] *Haine en ligne. Le parquet de Paris va créer un pôle spécialisé*, OUEST FRANCE, (Nov. 23, 2020, 08h44), https://www.ouest-france.fr/high-tech/internet/haine-en-ligne-le-parquet-de-paris-va-creer-un-pole-specialise-7060291.

immediately presented to a judge (*comparution immédiate*), but only for the more serious cases.[425] Article 20 of the bill, presented at the French lower Chamber on December 9, 2020, proposes to allow individuals suspected of having committed one of the offenses incriminated by article 24 of the French Press Law, that is, to have posted "hate speech" online, to be immediately presented to a judge, thus following the criminal procedure for common crimes, not the special criminal procedure for press law crimes. The bill explained that the purpose for this change was "*to provide a rapid response to behavior which, in the context of major changes in communication tools, seriously undermines our ability to live together*."[426]

d. Hate Speech and Anonymity

It is easy to hide who we are on social media, at least on most platforms. While *Facebook* and *LinkedIn* require their users to use their real names when creating an account, Twitter Instagram, or TikTok allows their users to tweet anonymously. Some accounts are obviously using a fake name, such as the numerous accounts created over the years in the name of an escaped animal.[427] However, some accounts are created to make people believe that they are the legitimate account of a real individual. Some of these accounts are impersonating accounts, created by a user pretending to be someone else,

---

[425] Jean-Baptiste Jacquin, *Une procédure de comparution immédiate pour les propos haineux en ligne*, LE MONDE, (Nov. 19 2020 10h29), https://www.lemonde.fr/societe/article/2020/11/19/une-procedure-de-comparution-immediate-pour-les-propos-haineux-en-ligne_6060328_3224.html.

[426] Projet de loi nº 3649 confortant le respect des principes de la République [Bill n ° 3649 confirming respect for the principles of the Republic], Dec. 9, 2020, available at https://www.assemblee-nationale.fr/dyn/15/dossiers/respects_principes_republique.

[427] Nick Bilton, *A Snake Escapes the Bronx Zoo and Appears on Twitter*, THE NEW YORK TIMES, (March 29, 2011 2:28 pm), https://bits.blogs.nytimes.com/2011/03/29/a-snake-escapes-the-bronx-zoo-appears-on-twitter/; Anna Orso, *After coyote sightings, the internet meets @RadnorCoyote, the latest in a line of Twitter accounts by Philly wildlife*, THE PHILADELPHIA INQUIRER, (March 20, 2019). https://fusion.inquirer.com/news/radnor-coyote-twitter-west-philly-turkey-stormy-escaped-cow-bucks-county-bear-devin-nunes-cow-20190320.html.

usually a famous person. This practice led Twitter to create its "*verified account*" service, which places a blue checkmark next to the account name if Twitter has verified that the account truly belongs to the person presented as its owner.[428] Being verified allows the owner of the account to easily report impersonation. This practice is, however, on hold since November 2017, following a controversy around Twitter's verification of the account of Jason Kessler, the organizer of the Charlottesville "Unite the Right" alt-right rally in 2017. Twitter revoked Kessler's badge, but then suspended the verification program.[429] A verified account status can be lost if the owner break Twitter's rules,[430] among them "[*p*]*romoting hate and/or violence against, or directly attacking or threatening other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or disease*." These restrictions are like the ones imposed in Europe on unfettered free speech.

---

[428] *About verified accounts*, TWITTER, https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts (last visited Dec. 30, 2020).

[429] Mashable revealed in April 2019 that Twitter has however continued quietly to verify some accounts, among them Jack Dorsey's parents, see Karissa Bell, *Twitter secretly verified Jack Dorsey's mom and thousands of others despite 'pause'* , MASHABLE (April 16, 2019), https://mashable.com/article/twitter-verification-pause. Twitter announced it would resume verifying accounts in 2021, see *About verified accounts*, TWITTER HELP CENTER, https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts.

[430] *Verified account FAQs*, TWITTER, https://help.twitter.com/en/managing-your-account/twitter-verified-accounts, (last visited Dec. 30, 2020) :

"*Twitter reserves the right to remove verification at any time without notice. Reasons for removal may reflect behaviors on and off Twitter that include:*
Intentionally misleading people on Twitter by changing one's display name or bio.
Promoting hate and/or violence against, or directly attacking or threatening other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or disease. Supporting organizations or individuals that promote the above.
Inciting or engaging in harassment of others.
violence and dangerous behavior
Directly or indirectly threatening or encouraging any form of physical violence against an individual or any group of people, including threatening or promoting terrorism
Violent, gruesome, shocking, or disturbing imagery
Self-harm, suicide
Engaging in activity on Twitter that violates the Twitter Rules."

Other accounts are created pretending to be a real person but are in fact bots.[431] Such accounts represent a significant percentage of all Twitter accounts.[432] Bots may be created for a purpose, such as manipulating the opinion on a particular issue,[433] and the percentage of bots of all the accounts communicating on that issue may be even higher than the percentage of bots on Twitter.[434] Is a bot's speech protected by the First Amendment? The issue of "fake news" published on social media which aims at influencing an election is concerning as such publications are mostly the fruit of bots, not irate or misinformed citizens. The preliminary report on the French bill on disinformation,[435] written by Representative Bruno Studer, cited Representative Pieyre-Alexandre Anglade, which claimed that "*for 40,000 euros you can launch political propaganda operations on social networks, for 5,000 euros you can buy 20,000 hateful comments, and for 2,600 euros you can buy 300,000 followers on Twitter. At this price, entire sites, Facebook pages, Twitter threads*

---

[431] Bots are defined by Professors Madeline Lamo and Ryan Calo as "*automated agents that initiate communication online, by phone, or through other technologically mediated mea*n." Madeline Lamo & Ryan Calo, *Regulating Bot Speech*, 66 UCLA L. REV. 988, 993 (2019).

[432] Lara O'Reilly*, Twitter admits 8.5% of its users are bots* (Aug. 12, 2014), MARKETING WEEK, https://www.marketingweek.com/twitter-admits-8-5-of-its-users-are-bots; Zoey Chong, *Up to 48 million Twitter accounts are bots, study says*, CNET,)March 14, 2017 9:05 a.m. PT), https://www.cnet.com/news/new-study-says-almost-15-percent-of-twitter-accounts-are-bots/.

[433] Bots may also be used to advance worthy cause, such as helping people suffering from depression, see Clive Thompson, *May A.I. Help You? Intelligent chatbots could automate away nearly all of our commercial interactions — for better or for worse*, THE NEW YORK TIMES, (Nov. 14 2018), https://www.nytimes.com/interactive/2018/11/14/magazine/tech-design-ai-chatbot.html, writing about Woebot, a text-chatbot therapist built by Alison Darcy, a clinical research psychologist at Stanford University, to help people suffering from depression or anxiety by providing cognitive behavioral therapy.

[434] *See* Virginia Alvino Young, *Nearly Half of the Twitter Accounts Discussing 'Reopening America' May Be Bots*, CARNEGIE MELLON UNIVERSITY, SCHOOL OF COMPUTER SCIENCE, (May 20, 2020), https://www.scs.cmu.edu/news/nearly-half-twitter-accounts-discussing-reopening-america-may-be-bots. Carnegie Mellon University researchers analyzed bot activity during the 2020 COVID-19 pandemic, using a sample of more than 200 million tweets discussing coronavirus or COVID-19. They found that 82% of the top 50 influential retweeters are bots, and that 62% of the top 1,000 retweeters were bots.

[435] We will examine the bill further on.

are spreading false information and sowing trouble in the minds of our fellow citizens."[436] We are far away from the marketplace of ideas envisioned by the Founding Fathers in the U.S., or by the French revolutionaries in 1789 enacting the Declaration of Human Rights and of the Citizens, only a few days after the Bastille prison had been taken. There were however mercenaries in the armies at the time, and bots are now fighting (dis)information wars.[437] The author of the preliminary report noted further that the law did not aimed at preventing the publication of false information, but rather, at containing their spreading, taking care to assert that "*it is by no means desirable in a democracy to prevent citizens from sharing the information they want, whether it is true or false*."[438] As such, posting a false information online is still legal.

Professor Laurent Sacharoff argued that the bots'rights to speak anonymously may be protected by the First Amendment,[439] citing *National Institute of Family and Life Advocates v. Becerra*, where the Supreme Court, in an opinion written by Justice Thomas, held that the California Reproductive FACT Act which required unlicensed clinics primarily serving pregnant women[440] to notify women that California has not licensed the clinics to

---

[436] Rapport fait au nom de la commission des affaires culturelles et de l'éduction sur la proposition de loi relative à la lutte contre la manipulation de l'information [Report made on behalf of the Committee on Cultural Affairs and Education on the proposed law on the fight against the manipulation of information], N° 990, p. 36, cited by Emmanuel Dreyer, *Fausse bonne nouvelle : la loi du 22 décembre 2018 relative à la manipulation de l'information est parue*, LEGIPRESSE N ° 367, 19. The report in French is available online : http://www.assemblee-nationale.fr/dyn/15/rapports/cion-cedu/l15b0990_rapport-fond.pdf.

[437] One out of every five political messages generated during the 2016 U.S. presidential election campaign were generated by bots, see *How the world was trolled*, THE ECONOMIST, (Nov. 4, 2017), 22.

[438] Preliminary report, p. 25.

[439] Laurent Sacharoff, *Do Bots Have First Amendment Rights?,* POLITICO, (Nov. 27 2018), https://www.politico.com/magazine/story/2018/11/27/bots-first-amendment-rights-222689

[440] These clinics are defined as facilities not licensed by the State of California, not having a licensed medical provider on staff or under contract and having the primary of providing pregnancy-related services. They also had to satisfy at least two of four requirements: "(*1) The facility offers obstetric ultrasounds, obstetric sonograms, or prenatal care to pregnant women. "(2) The facility offers pregnancy testing or pregnancy diagnosis. "(3) The facility advertises or solicits patrons with offers to provide prenatal sonography, pregnancy tests, or pregnancy options counseling. "(4) The facility has staff or volunteers who collect health information*

provide medical services violated the First Amendment.[441] Justice Thomas wrote that such notices were content-based regulation of speech and, that, by compelling individuals to speak a particular message, they were altering the content of their speech.[442] Similarly, bots could be defined by a statute, and accounts corresponding to this definition would be required to post a conspicuous disclaimer "I am a bot." Would such statute be a content-based statute violating the First Amendment? It appears to be so under *National Institute of Family and Life Advocates v. Becerra.*

Professors Madeline Lamo and Ryan Calo discussed whether laws making mandatory for bots' operators to identify bots as such are constitutional, noting that "*the unintended consequence of bot disclosure laws for speech and privacy could be significant*."[443] The authors concluded that bot speech is protected by the First Amendment, and that the right to read or listen to bot speech is also protected, as the First Amendment protects the right to receive information.[444] Professors Lamo and Calo believe that requiring bot disclosure should be made with caution, citing *Brown v. Entertainment Merchants Ass'n* which held that "*new categories of unprotected speech may not be added to the list by a*

---

from clients." Such unlicensed facilities had to provide a notice, onsite and in all advertising materials, stating that "[*t*]*his facility is not licensed as a medical facility by the State of California and has no licensed medical provider who provides or directly supervises the provision of services*." Cal. Health & Safety Code Ann. §123472(b)(1).

[441] National Institute of Family and Life Advocates v. Becerra, (June 26, 2018), available at https://www.supremecourt.gov/opinions/17pdf/16-1140_5368.pdf.

[442] Citing Riley v. National Federation of Blind of N.C., Inc., 487 U.S. 781, 795 (1988): "Mandating *speech that a speaker would not otherwise make necessarily alters the content of the speech*." At issue in this case was the requirement of the North Carolina Charitable Solicitations Act that professional fundraisers disclose to potential donors, before an appeal for funds, the percentage of charitable contributions collected during the previous 12 months that were actually turned over to charity.

[443] Madeline Lamo & Ryan Calo, *Regulating Bot Speech*, 66 UCLA L. REV. 988, 991 (2019), giving as example a person having to disclose his or her true identity if accused of running an illicit bot.

[444] Madeline Lamo & Ryan Calo, *Regulating Bot Speech*, 66 UCLA L. REV. 1005-1006 (2019), citing Board of Education v. Pico, 457 US 853, 867 (1982).

legislature that concludes certain speech is too harmful to be tolerated,"[445] and arguing that "[*l*]*ike videogames, bots can be a vehicle for speech that society finds problematic*."[446] A bot's operator may, however, may have to disclose her of his identity and disclose that he or she is operating a bot if required to do so by a judge acting upon the complaint of an individual or a government, if speech published by the bot has been deemed illegal. Professor Oren Etzioni from the Allen Institute for Artificial Intelligence argued in 2017 that "*an A.I. system must be subject to the full gamut of laws that apply to its human operator… We don't want A.I. to engage in cyberbullying, stock manipulation or terrorist threats*,"[447] and that view is shared on many. But revealing its nature as a bot is not the same as disclosing the identity of the person or entity who created the bot and spoke through it.

Not all anonymous speech on social media is created by bots. We know that at the root of the U.S. Supreme Court position on allowing hate speech is the desire of more speech to foster a vibrant marketplace of ideas, and that the First Amendment protects the right to speak anonymously. The Supreme Court held in 1960, in Talley v. California that "[*t*]*here can be no doubt that … an identification requirement would tend to restrict freedom to distribute information and thereby freedom of expression*."[448]

Anonymous speech is protected as:

"[*u*]*nder our Constitution, anonymous pamphleteering is not a pernicious, fraudulent practice, but an honorable tradition of advocacy and of dissent. Anonymity is a shield*

---

[445] Brown v. Entertainment Merchants Ass'n, 564 US 786, 791(2011), citing Stevens, 559 U.S, 460, 470 (2010).
[446] Madeline Lamo & Ryan Calo, ibid, at 1026.
[447] Oren Etzioni, *How to Regulate Artificial Intelligence*, THE NEW YORK TIMES, (Sep. 1, 2017), https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html.
[448] Talley v. California, 362 US 60. 64(1960).

from the tyranny of the majority," explaining that "[*a*]*nonymous pamphlets, leaflets, brochures and even books have played an important role in the progress of mankind. Persecuted groups and sects from time to time throughout history have been able to criticize oppressive practices and laws either anonymously or not at all*."[449]

However, when anonymous speakers exchange ideas online with speakers using their own name, the marketplace of ideas is unbalanced. Trolling breeds trolling[450] and even people ordinarily not inclined to hate speech may be tempted to engage in antisocial behavior on social media.[451] Anonymity provides a false sense of impunity to so many people that they many become part of a "hate herd," which constant postings on a high scale makes it difficult to fight or to ignore.[452] *New York Times* editor Jonathan Weisman wrote in one of his lasts tweets before leaving Twitter in 2016: "*So I will be moving to Facebook where at least people need to use their real names and can't hide behind fakery to spread their hate*."[453] Another journalist, Paul Wells, left Twitter in September 2016 to protest against Twitter's perceived lack of courage when facing demands from Turkey to block the account in Turkey of journalist Mahir Zeynalov. Mr. Wells explained why in an

---

[449] McIntyre v. Ohio Elections Comm'n, 514 US 334, 357 (1995).
[450] *See* Mattathias Schwartz, *The Trolls Among Us*, THE NEW YORK TIMES (Aug. 3 2008), https://www.nytimes.com/2008/08/03/magazine/03trolls-t.html: "*In the late 1980s, Internet users adopted the word "troll" to denote someone who intentionally disrupts online communities*."
[451] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, Jure Leskovec, *Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions*, CSCW '17: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (Feb. 2017), 1217-1230, https://www.cs.cornell.edu/~cristian/Anyone_Can_Become_a_Troll_files/anyone_can_become_a_troll.pdf. The authors noted that "*negative mood increased the probability of a user subsequently trolling in an online news comment section, as did the presence of prior troll posts writ-ten by other users*."
[452] *See* Nick Wingfield, *Feminist Critics of Video Games Facing Threats in 'GamerGate' Campaign*, THE NEW YORK TIMES (Oct. 15, 2014), https://www.nytimes.com/2014/10/16/technology/gamergate-women-video-game-threats-anita-sarkeesian.html?_r=2, writing that "[*t*]*he malice directed recently at women, though, is more intense, invigorated by the anonymity of social media and bulletin boards where groups go to cheer each other on and hatch plans for action*."
[453] Tom Kludt, *New York Times editor quits Twitter over anti-Semitic tweets*, CNN BUSINESS, (June 9, 2016: 9:21 AM ET), https://money.cnn.com/2016/06/08/media/new-york-times-jon-weisman-twitter/.

online article, writing that "[*u*]*ntil Twitter makes it clear that it has the back of Zeynalov and other public truth-tellers, it cannot credibly protest that its users have to put up with brigades of anonymous liars*."[454]

We regularly hear calls to make anonymous speech on social media illegal, and these calls sometimes have surprising sources. French President Emmanuel Macron declared in 2019 being in favor of a "*gradual lifting*" of anonymity on social media, leading to French law professor Emmanuel Netter to react by stating that, while "*the feeling of impunity online is real... it is not about anonymity, which governments can lift at any time. It is due to the absence of a clear criminal policy, the lack of training and resources within the police and justice services. It is due to the lack of prosecution that nothing prevents, neither technically nor legally*," and by calling for a more generous funding of the public service of justice.[455] Indeed, it can be argued that is easier to be anonymous IRL ("In Real Life") than online, where our IP address, email address, geolocation and other meta data can be as many crumbs leading interested parties to the troll's burrow. One can however easily be pseudonymous, even though this veil is thinner than we believe, and many people may even know the real identity behind a popular account.[456] "Doxing" (or "doxxing") is a term describing the fact to reveal the identity of a person who had chosen to stay anonymous online. Doxing is a crime in Singapore since passage of the Protection from Harassment (Amendment) Act 2019, which defines doxing as the fact "*with intent to cause harassment,*

[454] Paul Wells, *Why I left Twitter,* THE STAR, (Sep. 27, 2016), https://www.thestar.com/news/canada/2016/09/27/why-i-left-twitter-paul-wells.html.
[455] Emmanuel Netter, *Opinion, Contre la levée de l'anonymat en ligne*, LES ÉCHOS, (Feb. 11, 2019, 19:29), https://www.lesechos.fr/idees-debats/cercle/opinion-contre-la-levee-de-lanonymat-en-ligne-963580.
[456] Such is the case for the most famous and active online French attorney "Maître Eolas", whose real identity is known by many, but whose wish to remain anonymous is respected, even by the French *Cour de cassation* when ruling about the online defamation case stemmed by one of his tweets., as we saw earlier.

alarm or distress to another person [the target person] … by any means [to]… publish any identity information[457] *of the target person or a related person of the target person, and as a result causing the target person or any other person… harassment, alarm or distress*."[458]

The French government seems inclined to view anonymity on social media as a threat to democracy. Prime Minister Jean Castex said in a 2020 interview that he found anonymity on social media shocking, even comparing it to the collaborationist Vichy regime of France during WWII,[459] alluding to a time where neighbors and friends were denounced anonymously to the police for being Resistants,[460] Jewish, black market profiteer, or simply immoral. This view is shared across the Channel by Labour Member of Parliament Dame Margaret Hodge, which is often attacked anonymously on social media.[461]

III. Protecting Morality

[457] "Identity information" is defined by the Act as "*any information that, whether on its own or with other information, identifies or purports to identify an individual, including (but not limited to) any of the following:(a) the individual's name, residential address, email address, telephone number, date of birth, national registration identity card number, passport number, signature (whether hand written or electronic) or password;(b) any photograph or video recording of the individual;(c) any information about the individual's family, employment or education*."

[458] An Act to amend the Protection from Harassment Act (Chapter 256A of the 2015 Revised Edition) and to make related amendments to certain other Acts, available online at https://perma.cc/VCG5-Y9ZJ. See also *Singapore: Amendment to Harassment Law Passed to Criminalize "Doxxing"*, GLOBAL LEGAL MONITOR, LIBRARY OF CONGRESS, (Oct. 25. 2019), https://www.loc.gov/law/foreign-news/article/singapore-amendment-to-harassment-law-passed-to-criminalize-doxxing/.

[459] Pierre Lepelletier, *Pour Castex, l'anonymat sur les réseaux sociaux rappelle le «régime de Vichy»*, LE FIGARO, (July 16, 2020, 08:34, Updated July 16 09:18), https://www.lefigaro.fr/politique/pour-castex-l-anonymat-sur-les-reseaux-sociaux-rappelle-le-regime-de-vichy-20200716.

[460] To read such a letter, see *Une lettre de dénonciation*, MUSÉE AQUITAINE BORDEAUX, http://www.musee-aquitaine-bordeaux.fr/fr/article/une-lettre-de-denonciation (last visited Dec. 30, 2020) : « I *have the honor to inform you that Mr xxxxxxxx residing at Mr xxxxxxxxx at Pinier C.ne de Laruscade…has by his own admission hidden explosives I believe in or under the feeders of his stable and states being able to get more.* "

[461] Jessica Elgot, *Margaret Hodge calls for ban on social media anonymity*, THE GUARDIAN, (Dec. 6 2020, 12.08 EST), https://www.theguardian.com/society/2020/dec/06/margaret-hodge-calls-for-ban-on-social-media-anonymity. The article cites an Amnesty report which had found that MP Abbott "*had received 45% of all abusive tweets sent to female MPs in the six weeks before election day in 2017*."

A. Blasphemy

Social media sites are accessible from all over the world, even in countries where blasphemy is still a crime. If an anonymous social media user posts a message which is considered blasphemous in a particular country, it can be seen there, unless the country is asking the social media site to block the account, at least in the country where the post is blasphemous. Such was the case in May 2014, when the Pakistan Telecommunications Authority asked Twitter to block several accounts[462] for having used the name the Prophet Muhammad in their handles in an injurious way. The requests cited the Pakistani penal Code as legal basis of the violation. However, Twitter restored access to the next month, explaining that it "*made an initial decision to withhold content in Pakistan based on information provided to us by the Pakistan Telecommunication Authority… We reexamined the requests and, in the absence of additional clarifying information from Pakistani authorities, have determined that restoration of the previously withheld content is warranted. The content is now available again in Pakistan*."[463] The right to freely worship must also include the right not to worship a God, or any God at all. When this opinion is expressed online some countries, it may lead to murder. For instance, a secular blogger was killed in August 2015, probably because his killers had taken offense of several of his posts critical of Islam.[464]

---

[462] Robert Mackey, Twitter Agrees to Block 'Blasphemous' Tweets in Pakistan, THE NEW YORK TIMES, (May 24, 2014), http://www.nytimes.com/2014/05/22/world/asia/twitter-agrees-to-block-blasphemous-tweets-in-pakistan.html.

[463] *Pakistan's PTA Requests Removal of Various "Blasphemous" Tweets,* LUMENDATABASE, https://www.lumendatabase.org/notices/1412999, https://www.lumendatabase.org/notices/1412998, https://www.lumendatabase.org/notices/1412994 , https://www.lumendatabase.org/notices/1412993, (Last visited Dec.30, 2020.) *See* also *Pakistan's PTA Requests Removal of "Un-Ethical" Tweets and Blasphemous URL*, LUMENDATABASE, https://www.lumendatabase.org/notices/1412996, (Last visited Dec. 30, 2020).

[464] Bangladesh: Secular Blogger is Killed, NEW YORK TIMES, (Aug 8, 2015), p. A5.

Justice Brennan, who delivered the opinion of the court in *Roth v. U.S.*,[465] noted that thirteen out of the fourteen states which had ratified the Constitution in 1792 had laws making blasphemy a crime.[466] Blasphemy is incriminated in Germany since 1871 and was originally defined as denying God. However, the definition of what is blasphemy was modified in 1969 as insults against any religion which could endanger the peace or public order.[467] Paragraph 166 of the German penal code states that:

> *"(1) Whosoever publicly or through dissemination of written materials…defames the religion or ideology of others in a manner that is capable of disturbing the public peace, shall be liable to imprisonment not exceeding three years or a fine.(2) Whosoever publicly or through dissemination of written materials… defames a church or other religious or ideological association within Germany, or their institutions or customs in a manner that is capable of disturbing the public peace, shall incur the same penalty*."[468]

However, blasphemy laws are usually adopted to protect the only, or at least the dominant religion of the country. This was the case in France, where the blasphemy law of the *Ancien Régime* protected the catholic religion. King Louis IX, Saint Louis, published an Ordinance in 1268, punished it by a fine, exposure to pillory, three days in jail, or whipping.

---

[465] Roth v. United States, 354 U.S. 476 (1957).
[466] Roth, at 482, noting that "Massachusetts made it criminal to publish "any filthy, obscene, or profane song, pamphlet, libel or mock sermon" in imitation or mimicking of religious services. Acts and Laws of the Province of Mass. Bay, c. CV, § 8 (1712), Mass. Bay Colony Charters & Laws 399 (1814). A note 12 in Roth lists all the statutes, particularly: Act Against Drunkenness, Blasphemy, §§ 4, 5 (1737), 1 Laws of Del. 173, 174 (1797; Act for the Punishment of Profane Cursing and Swearing (1791), N. H. Laws 1792, 258; 1715-1790); Act to Prevent the Grievous Sins of Cursing and Swearing (1700), II Statutes at Large of Pa. 49 (1700-1712; Act for the More Effectual Suppressing of Blasphemy and Prophaneness (1703), Laws of S. C. 4 (Grimke 1790).
[467] *Scrap German blasphemy law to promote tolerance, legal expert demands*, DEUTSCHE WELLE (Jan. 24, 2015), http://www.dw.de/scrap-german-blasphemy-law-to-promote-tolerance-legal-expert-demands/a-18213179.
[468] The English translation of the German Criminal Code is available at http://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html (Last visited Dec. 30, 2020).

Saint Louis also wanted to punish blasphemy by burning the lips of the offender but is seems that Pope Clement IV asked him not to do so.[469] King Philippe VI introduced cutting the blasphemous' lips as a penalty in 1330. Louis XII published an Ordinance on blasphemy in 1511, with lesser penalties: the pillory was a penalty for blasphemy only after the fifth offence, followed by cutting the upper lip to punish the sixth offence, the lower lip the seventh offence, with the tongue being finally cut at the eighth offence.[470] The Chevalier de la Barre was sentenced to death in 1766 after having been found guilty of blasphemy, after a crucifix planted on the side of a road had been found damaged in the town where he lived. He was known as having shown signs of impiety and thus made a readily available culprit. His fate is still well known as Voltaire wrote about him in his *Relation de la mort du chevalier de la Barre*, published a few days after the death sentence had been carried out in July 1766. In this book, Voltaire wrote that it was "*very doubtful that the crucifix had been mutilated on purpose*" but, rather, a chariot must have done the deed by accident.[471] The Chevalier was finally pardoned by Louis XVI in 1788, twenty-two years after the Chevalier's death, and one year before the French Revolution.

France is now, of course, a Republic, and article 1 of its Constitution states that the country is a secular Republic which respects all beliefs. Indeed, France does not recognize any religion, nor does it differentiate between them. An author explained that while some laws make it a crime to offend religion, they should not be considered exception to the principle of *laicité* (secularity) so dear to the French, but rather an expression of it, as "*from*

[469] JEAN-MARIE CARBASSE, HISTOIRE DU DROIT PÉNAL ET DE LA JUSTICE CRIMINELLE, 330, note 1 (PUF), 2nd ed. 2009.
[470] CARBASSE, p. 330.
[471] VOLTAIRE, L'AFFAIRE DU CHEVALIER DE LA BARRE, in L'AFFAIRE CALAS, 313, (Folio eds.), 2008.

the respect of secularity stems freedom of conscience."[472] Curiously, blasphemy was still incriminated in France recently, albeit not on the entire territory, but only in Alsace Moselle. This is because this area, located in the east of France, had been annexed to the German Empire in 1871 by the Treaty of Frankfurt, following France's defeat in the war against Germany. When Alsace and Moselle were reunited with France after World War I, they both kept the regime of the *Concordat*, which had been put in place by Napoléon Bonaparte in 1801 and had been replaced in France by the separation of Church and State law of December 9, 1905 which claimed, for the first time in France, that its Republic does not recognize any religion. Under article 166 of the local Penal Code, which was maintained in the Alsace Moselle legal framework by a November 25, 1919 decree:

> "*he who caused a scandal by publicly blaspheming against God by outrageous speech or publicly insults one of the Christian cults or religious communities [recognized as such by the law, that is, Catholicism, Protestantism and Judaism] (…), or who, in a church or other place devoted to religious assemblies, commits offensive and scandalous acts, shall be punished by imprisonment of up to three years*."

A Senator from the Moselle, Jean Louis Masson, asked in 2006 the Minister of Police, which is also in charge of religion affairs, whether this law was still in force. The minister answered that the law was still in force, but that the determination of its scope, "*particularly as regards to its extension to non-recognized religions, [was] within the exclusive competence of the judiciary*." However, representatives of the Catholic, Protestant,

[472][472] Bernard Beignier, ibid, p. 335.

Jewish, and Muslim religions asked on January 6, 2015 the suppression of article 166.[473] On May 12, 2015, France's *Observatoire de la laïcité* (*Secularism Observatory*), which mission is *"to assists the Government in its efforts to respect the principle of secularism in France*," published its opinion (*avis*) on the local Alsace Moselle church regime, which addressed the issue of blasphemy.[474]

For the *Observatoire de la laïcité*:

"[*i*]*t appears that [the survival of article 166 of the local penal Code incriminating blasphemy) does not imply it has any legal effect, since it is not part of the provisions which have been formally translated in order to be introduced into the French law by the Decree of August 27, 2013 concerning the publication of the translation of laws and regulations kept in force by local laws from 1 June 1924 in the departments of Bas-Rhin, Upper Rhine and Moselle*."

It seems that blasphemy is, finally, no longer a crime in France. The seventeenth chamber of the Paris criminal Court of first instance, which specializes in press law,[475] held on December 10, 2015, that Jean-Michel Ribes, director of the Parisian theater *Théâtre du Rond-Point*, and Pascale Vurpillot, editor, were not guilty of provoking racial discrimination and hate against Christians, for having shown in his theater, and for having published in

[473] Jean-Christophe Dupuis-Remond, *En Alsace-Moselle, les cultes préconisent l'abrogation du délit de blasphème*, FRANCE 3 RÉGIONS, January 12, 2015, 18 :25, http://france3-regions.francetvinfo.fr/lorraine/2015/01/12/en-alsace-moselle-les-cultes-preconisent-l-abrogation-du-delit-de-blaspheme-631294.html.
[474] Observatoire de la laïcité, *Avis sur le régime local des cultes en Alsace et en Moselle*, May 12, 2015, p. 14, http://www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2015/05/avis_alsace-moselle_definitif_0.pdf .
[475] 17eme chambre correctionnelle du tribunal de grande instance de Paris.

France, a play written by Rodrigo Garcia, *Golgotha Picnic*.[476] The play presented Jesus Christ as vain man incapable of enjoying life's simple pleasures, followed by only a few fools, and argued further that being subjected to Christian iconography, its tears and open wounds visited by inquisitive fingers, leads to pedophilia acts, murders, and voracious eating of Big Macs. The play was published in France in November 2011 and presented the same month at the *Théâtre du Rond-Point*, on a stage entirely covered by hamburgers buns. The French Bishops issued an official statement of protest, and a nonprofit organization, which name translate as the 'general alliance against racism and for the respect of French identify' filed a criminal complaint, arguing that the play was a provocation to discrimination, hate, or violence against Christians, and that freedom of speech, even artistic freedom of speech, cannot justify such "*excess*." Jean-Pierre Ribes argued in defense that he felt to be his duty to program "*disturbing*" writers and noted that the play was humorous and satirical, a further argued that plaintiff used the guise of complaining about the alleged provocation to hate towards Christians, which is a crime under French law, to instead punish what they view as blasphemous speech, which is not a crime in France. French criminal procedure calls for a *juge d'instruction*, an investigating magistrate, to examine both the incriminating and exonerating facts in an impartial manner, to decide whether to send or not to send the case to trial. The investigating magistrate had found in this case that the play:

> *"exceed[ed] the limits of freedom of expression and can be analyzed, by the discredit it imparted to the Christ and his actions, by its violence and stigmatization of the*

[476] Tribunal de grande instance de Paris [TGI] [ordinary courts of original jurisdiction] Dec. 10 2015, n°12-30.5023020, LEGIPRESSE 2016 p.15.

community of those who are the disciples of Christ, as inciting to hatred and rejection of Catholics."

The Court, however, held in favor of defendants. It noted that blaspheme is not a crime in France, except in Alsace-Moselle,[477] and that the play was "*a work of fiction with a purely artistic vocation*," not an historical or scientific work. Also, the authors of such work do not have to respect any morale or religion as this would be "*a censure inherently antithetical to the exchange of ideas and opinion essential to any democratic society.*" While France tolerates blasphemy, profane and indecent words may still be considered undesirable in the public place, including on social media.

B. Profane and Indecent Words

"[O]ne man's vulgarity is another's lyric"[478]

The French often use colorful language, including on social media, and, sometimes, even if posting from a corporate account. Such was the case when the Twitter account of Winamax Sport, posted, after the qualifications of two French soccer teams to the UEFA Champions League, a vulgar and indecent tweet, inspired by a rap song.[479] A French Representative, Olga Givernet, asked for the tweet to be taken down and even sent a letter to French Prime Minister Jean Castex, calling the words "*shocking*" and "abject," arguing that Winamax had used these *"homophobic words*" to create a buzz in order to raise

[477] At least at the time, as we saw earlier.
[478] Cohen v. California, 403 US 15, 25 (1971).
[479] « On prend l'Europe, on l'encule à deux » (One takes Europe, one both fucks her), a tweet inspired by a song, as explained in Marc Rees, *Sermonnée par l'ANJ, Winamax retire son tweet polémique dénoncé par une députée LREM*, NEXTINPACT, (Aug. 18, 2020, 09:21), https://www.nextinpact.com/article/43381/sermonnee-par-lanj-winamax-retire-son-tweet-polemique-denonce-par-deputee-lrem.

interest in sport betting.[480] Representative Givernet also asked to agency in charge of regulating betting, the *Autorité nationale des jeux* (National games authority) to suspend Winamax's rights to engage in sports. betting. The Authority, however, found that Winamax had not breached the agreement allowing him to engage in sports betting, that this decision would have to be made following a procedure respecting due process, but expressed, however, its surprise that it had:

> "*chosen to base its promotional campaign on a tweet which, directly inspired by a French Rap group with a very strong audience among young audiences and especially minors, may entail a risk of incitement to gambling for them, a risk to which operators are legally required to prevent them*."[481]

The Authority further noted that "*it would be particularly welcome for Winamax to take the initiative to immediately remove this communication from the accounts of its various social networks*." Winamax complied.

In the U.S., the Supreme Court considered the issue of obscene words uttered in public in *FCC v. Pacifica*. At issue was the twelve-minute "*Filthy Words*" monologue George Carlin had recorded before a live audience in a California theater, which was then broadcast by a New York radio at two o'clock in the afternoon. The monologue dealt mainly with sex and excretion. Answering the question of whether the government can restrict

[480] Representative Olga Givernet made her letter to the Prime Minister public by posting it on Twitter, @OlgaGivernet, Twitter, (2:38 AM Aug 18, 2020), https://twitter.com/OlgaGivernet/status/1295610922225065985. She d

[481] Tweet polémique de Winamax : réaction de l'Autorité nationale des jeux (ANJ), AUTORITÉ NATIONALE DES JEUX, (Aug. 18, 2020, 10:03), https://anj.fr/tweet-polemique-de-winamax-reaction-de-lautorite-nationale-des-jeux-anj.

public broadcasting of indecent language in any circumstances, without violating the First Amendment, the Supreme Court held narrowly that the Federal Communications Commission could sanction the radio station for having broadcast the monologue at a time when children could hear it.[482] Writing for the Court, Justice Stevens noted that "*these words ordinarily lack literary, political, or scientific value"* but that they "*are not entirely outside the protection of the First Amendment."*[483] The words were not obscene, yet could be censored, at least during a certain time of the day and if the medium used was the radio. Justice Stevens had noted that "*broadcasting is uniquely accessible to children, even those too young to read*."[484]

Social media platforms are similarly "uniquely accessible": Facebook's terms of use do not allow users to register if they are under the age of thirteen,[485] but Twitter does not have such rules. In *Pacifica*, Justice Stevens noted that the Court had long recognized "*that each medium of expression presents special First Amendment problems,"*[486] and noted further that *"the broadcast media have established a uniquely pervasive presence in the lives of all Americans*." These remarks can certainly be applied today to social media platforms, which ubiquity and generalized use made them a unique medium of expression which has created a worldwide marketplace of ideas.[487] Unlike a broadcast, social media platforms are available 24/7, around the globe, if one has a phone or a computer connected to the Internet. The terms used by Carlin in his monologue, piss, fuck, cunt, cocksucker,

---

[482] Floyd Abrams called this decision "*one of the Court's worst in the last quarter century*", see FLOYD ABRAMS, FRIEND OF THE COURT, ON THE FRONT LINES WITH THE FIRST AMENDMENT 42 (2013).
[483] Pacifica, at 746.
[484] Pacifica, at 749
[485] Facebook Statement of Rights and Responsabilities, article 4.5.
[486] Pacifica, at 748
[487] This is particularly the case with Twitter, which posts can be read even if one if not logged in.

motherfucker, and tits, are routinely used on Twitter, as a simple search can reveal, and are thus readily available to minors. The place where the speech takes place is relevant, as explained by the Supreme Court in *Cohen v. California*.[488] In this case, the offensive speech at stake was a jacket bearing "Fuck the Draft "on its back, worn inside the Los Angeles County Courthouse. The government had argued that speech was directed at unwilling or unsuspecting viewers, and that thus it could legitimately act to protect the sensitivity of the public, who had been forced to read the profane message. The court noted however that the mere fact that an unwilling audience may be present is not enough to justify curtailing the speech as the public is often captive outside their home and may encounter objectionable speech. To do so, the government would have to show "*that substantial privacy interests are being invaded in an essentially intolerable manner.*"[489] Any broader view of this authority would effectively empower a majority to silence dissidents simply as a matter of personal predilections. In the *Cohen* case, the public could simply avert their eyes to avoid "*further bombardment of their sensibilities.*"[490] The equivalent of averting one's eyes on social media is to block, to mute, or to unfollow. However, while it may be easy to avert our eyes from offensive social media messages which are not directed at us, it is more difficult to do so if the offensive message is directly addressed to us. A message can be sent on social media either directly, as a private message, or can be directed at a person by tagging him or her or using the "@" sign. Twitter allows its users to mute accounts[491]

[488] Cohen v. California, 403 US 15 (1971).
[489] Cohen, at 21.
[490] Cohen at 21.
[491] *How to mute accounts on Twitter,* TWITTER, https://help.twitter.com/en/using-twitter/twitter-mute. (Last visited Dec. 30, 2020).

and to block them.[492] However, blocking often occurred only after an account has been used to direct insults and abusive language at one's user. As noted by Professor John Deigh, *Cohen v. California* gave protection to "*emotive speech that causes personal and private harms …* [*such as*] [*e*]*xpletives, slurs, vulgarities, and other forms of speech that express contempt, disgust, animus" toward a person or a group*."[493] "Fuck" is still a transgressive term, although probably less so than in 1968. at least with the large public. It had lost however its obscene character by 1968 and was used as a vulgar way to emphasize one's statement. For instance, President Lyndon B. Johnson said to the Greek Ambassador to the U.S. in 1965: "*Fuck your parliament and your constitution. America is an elephant. Cyprus is a flea. Greece is a flea. If these two fellows continue itching the elephant they may just get whacked by the elephant's trunk, whacked good*."[494] Would "Fuck" have been still obscene, Cohen's speech may not have been protected by the First Amendment.

C. Pornography and Obscenity

a. USA

The First Amendment protects sexual expression which is indecent but not obscene,[495] as obscenity has no "*redeeming social importance*" and is not protected by the First Amendment.[496] Child pornography is not protected speech either, even it is not obscene.[497] Obscenity was defined in the late 19th century by Charles Porterfield in the

[492] *How to block accounts on Twitter*, TWITTER, https://help.twitter.com/en/using-twitter/blocking-and-unblocking-accounts, (Last visited Dec. 30, 2020).
[493] THE OFFENSIVE INTERNET, 196 (JOHN DEIGH, FOUL LANGUAGE: SOME RUMINATIONS ON COHEN V. CALIFORNIA), Saul Levmore and Martha C. Nussbaum, eds., 2010).
[494] Harold Pinter, VARIOUS VOICES, PROSE, POETRY, POLITICS, 1948-1998 (1999).
[495] Sable Communications of Cal. Inc. v. FCC, 492 U.S. 115, 126 (1989).
[496] Roth v. United States, 354 U.S. 476, 484-485 (1957).
[497] New York v. Ferber, 458 US 747 (1982) In this case, the Court found that N. Y. Penal Law, Art. 263-15, which made the use of a child in a sexual performance a class C felony, was not unconstitutional. Under this

American and English Encyclopædia of Law as being "*generally as that which is offensive to decency or chastity, which is immodest, indelicate, or impure, exciting lewd thoughts of an immoral tendency* ."[498] Some sixty years later, the Supreme Court placed obscenity outside of the scope of protection of the First Amendment in *Roth v. United States*.[499] However, what was obscene at the time Charles Porterfield wrote his encyclopedia was no longer obscene in 1957, particularly speech "*offensive to decency or chastity*." This was necessary as obscenity does not create a "clear and present danger" and thus this exception could not apply to obscenity.[500] Instead, the Court found that it had no redeeming social importance.

What is considered to be obscene by U.S. law? Sex is not necessarily obscene, even though, as the Supreme Court explained in *Cohen*, obscene expression "*must be, in some significant way, erotic,*" a speech which "*conjure*[*s*] *up* [*a*] *psychic stimulation*."[501] The Supreme Court defined "obscenity" in *Roth* as "*deal[ing*] *with sex in a manner appealing to prurient interest*"[502] and explained that obscene material is one which dominant theme taken as a whole appeals to the prurient interest of the average person applying contemporary community standards.[503] Defining what is "prurient" is not an easy task, as

---

law, a person was guilty of using a child in a sexual performance " *if knowing the character and content thereof he employs, authorizes or induces a child less than sixteen years of age to engage in a sexual performance or being a parent, legal guardian or custodian of such child, he consents to the participation by such child in a sexual performance*." Such performance was defined as any performance including sexual conduct by a child younger than 16 years. The Statute defined `sexual conduct' as "actual or simulated sexual intercourse, deviate sexual intercourse, sexual bestiality, masturbation, sado-masochistic abuse, or lewd exhibition of the genitals." The New York Court of Appeals, applying the Miller test, had found that §263.15 violated the First Amendment, 52 N. Y. 2d 674, 422 N. E. 2d 523 (1981). The Supreme Court found that the State had a compelling interest in "safeguarding the physical and psychological well-being of a minor", *Ferber*, at 757.

[498] AMERICAN AND ENGLISH ENCYCLOPEDIA (Volume 21, 759), David S. Garland, et al. Editors, American and English (1896).

[499] Roth v. United States, 354 US 476 (1957).

[500] Donna A. Demac, LIBERTY DENIED, p. 41, Rutgers University Press, 1990.

[501] Cohen, at 20.

[502] Roth, at 487.

[503] Roth, at 489.

acknowledged by Justice Stewart in his concurrence in *Ginzburg v. United States* : *"I know it when I see it*."[504] The Supreme Court recognized however in *Ginsberg v. New York*[505] that the "community standards" which must be used to define obscenity may vary. If one applies this principle to social media, the content found on accounts open to anyone must be judged on the broadest community standard possible. Could it be a universal community standard? And if it could, which universal community standard would apply to measure whether a content is obscene? In *Miller v. California*,[506] the Supreme Court proposed a test to be used by the courts having to decide whether speech is obscene or not: they must examine (a) whether the average person, applying contemporary community standards would find that the work, taken as a whole, appeals to the prurient interest, (b) whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law, and (c) whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value.

Even if using a test, what is obscene is a subjective concept. The book *Ulysses*, by James Joyce, now a regular entry on "100 best books" lists,[507] was seized for obscenity upon entry in the United States, under section 305 (a) of the Tariff Act of 1930.[508] The Second Circuit noted in 1934 that:

---

[504] Ginzburg v. United States, 383 U.S. 463
[505] Ginsberg v. New York, 390 U.S. 629 (1968).
[506] Miller v. California, 413 US 15, 24 (1973)
[507] *See* for example, at number 46, the list compiled in 2015 by Robert McCrum for The Guardian, *The 100 best novels written in English: the full list*, THE GUARDIAN, (Aug.17 2015 05.11 EDT)
Last modified on Tue 10 Sep 2019 19.56 EDT
https://www.theguardian.com/books/2015/aug/17/the-100-best-novels-written-in-english-the-full-list
[508] 19 USC § 1305 (a), which provides that "*all persons are prohibited from importing into the United States from any foreign country any obscene book, pamphlet, paper, writing, advertisement, circular, print, picture, drawing, or other representation, figure, or image on or of paper or other material*."

> "*numerous long passages in Ulysses contain matter that is obscene under any fair definition of the word cannot be gainsaid; yet they are relevant to the purpose of depicting the thoughts of the characters and are introduced to give meaning to the whole, rather than to promote lust or portray filth for its own sake*."[509]

As such, the book has the serious literary and artistic value required now by the Miller test. As sexually explicit speech is protected, and often published on social media, deciding what is obscene and what is not is a task undertaken by the corporations and their algorithms, which are designed to rake wide, thus englobing even artistic images.[510] Tumblr prohibits posting "adult content" on its site, which it defines as "*primarily includ*[*ing*] *photos, videos, or GIFs that show real-life human genitals or female-presenting nipples, and any content—including photos, videos, GIFs and illustrations—that depicts sex acts*."[511] This definition thus prevents posting on Tumblr a reproduction of the *Venus of Urbino* painted in 1538 by Titian, *Lunch on the Grass* painted by Edouard Manet in 1863, or *Nude on a Blue Cushion* painted by Modigliani in 1917, which are all in public view in renowned national museums. As such, Tumblr censors speech, which is clearly protected by the First Amendment, as "*portrayal of sex, e. g., in art, literature and scientific works, is*

---

[509] United States v. One Book Entitled Ulysses, 72 F. 2d 705 (2nd Circ. 1934)
[510] Jennifer Evans, *Why social media sites shouldn't censor erotic images*, THE WASHINGTON POST, (Dec. 14, 2018 at 6:00 a.m. EST), https://www.washingtonpost.com/outlook/2018/12/14/why-social-media-sites-shouldnt-censor-erotic-images/. The author provides several examples of photographs which were considered obscene at the time but are now featured in American museum collections and recognized as artistic achievements.
[511] *Adult content*, TUMBLR, https://tumblr.zendesk.com/hc/en-us/articles/231885248-Adult-content (Last visited Dec. 30, 2020).

not itself sufficient reason to deny material the constitutional protection of freedom of speech and press."[512] We will examine further how social media are censuring such speech later on.

b. The U.K.

UK Digital Economy Act 2017 defines "pornographic material" as "*material produced solely or principally for the purposes of sexual arousal*."[513] It defines "*extreme pornographic material*" as "*material whose nature is such that it is reasonable to assume that it was produced solely or principally for the purposes of sexual arousal, and which is extreme*." Such material is "extreme" under the Act "*if its content is as described in section 63(7) or (7A) of the Criminal Justice and Immigration Act 2008*,[514] *and it is grossly offensive, disgusting or otherwise of an obscene character*." What does "grossly disgusting" mean? In the *Handyside* case, Great Britain had considered that *The Little Red School Book* to be obscene, even though most of the countries part of the Council of Europe had not found it to be. The Obscene Publications Act defined then an article (a speech) as "obscene" if "*its effect or… the effect of one of its items is, if taken as a whole, such as to tend to deprave and corrupt persons who are likely, having regard to all relevant circumstances, to read, see or hear the matter contained or embodied in it*." This definition would surely be found to violate the First Amendment if used in a US law as it is vague and overbroad, and the Supreme Court explained in 1926 that "*a statute which either forbids or requires the doing of an act in terms so vague that men of common intelligence must necessarily guess at its meaning and differ as*

[512] Roth, at 487. Justice Brennan wrote further that "*Sex, a great and mysterious motive force in human life, has indisputably been a subject of absorbing interest to mankind through the ages; it is one of the vital problems of human interest and public concern*."
[513] Part 15(1), http://www.legislation.gov.uk/ukpga/2017/30/section/15/enacted.
[514] This law defines an extreme image as one which "*is grossly offensive, disgusting or otherwise of an obscene character*," http://www.legislation.gov.uk/ukpga/2008/4/section/63.

to its application, violates the first essential of due process of law. "[515] Part 3, Paragraph 1 of the Digital Economy Act 2017 forbids to make "*pornographic material available on the internet to persons in the United Kingdom on a commercial basis other than in a way that secures that, at any given time, the material is not normally accessible by persons under the age of 18.*"[516] How can a provider of online commercial pornography make sure that its users are all and only people who have already celebrated their eighteenth birthday? The UK Secretary of State published in January 2018 a "*Guidance from the Secretary of State for Digital Culture Media and Sport to the Age-verification Regulator for Online Pornography*."[517] The UK government boasted in its press release that "[*t*]*he UK will become the first country in the world to bring in age-verification for online pornography when the measures come into force on 15 July 2019*."[518] Is it possible to bar minors from accessing porn online? Probably not.[519] The British Board of Film Classification (BBFC) has been giving the responsibility to ensure compliance with the new laws.[520] Porn users are invited to provide their official, government-issued I.D. or their credit card information in order to access commercial pornographic material online. They can also buy a "porn pass" at stores, after proving to the clerk that they are indeed at least 18 years old. The law has been criticized by many,[521]

---

[515] Connally v. General Constr. Co., 269 US 385, 391 (1926).
[516] Available at http://www.legislation.gov.uk/ukpga/2017/30/section/14/enacted.
[517] *Guidance from the Secretary of State for Digital, Culture, Media and Sport to the Age-Verification Regulator for Online Pornography* : https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673425/Guidance_from_the_Secretary_of_State_for_Digital__Culture__Media_and_Sport_to_the_Age-Verification_Regulator_for_Online_Pornography_-_January_2018.pdf.
[518] *Age-verification for online pornography to begin in July*, GOV.UK (April 17, 2019), https://www.gov.uk/government/news/age-verification-for-online-pornography-to-begin-in-july.
[519] In one word: VPN…
[520] *Age-verification for online pornography to begin in July*, BBFC (April 17, 2019), https://bbfc.co.uk/about-bbfc/media-centre/age-verification-online-pornography-begin-july.
[521] Rowland Manthorpe, *Why the UK's porn block is one of the worst ideas ever*, WIRED UK, (April 17, 2019), https://www.wired.co.uk/article/porn-block-uk-wired-explains; John Herrman, *How the U.K. Won't Keep*

including by sex workers who fear that the law may prevent their ability to make a living.[522] Another main concern is privacy,[523] as the companies placed in charge of implementing the mandatory age-verification system will gather data on people watching porn. One may remember that, in July 2015, the *Ashley Madison* site, which offering extramarital dating services, was hacked, and threatened to make the data stolen public, unless the site would shut down.[524] One can imagine that the data set of U.K. porn users is of great interest to hackers.

Social media sites are, however, out of the scope of the law, as the Online Pornography (Commercial Basis) Regulations:

> "*does not apply in a case where it is reasonable for the age-verification regulator to assume that pornographic material makes up less than one-third of the content of the material made available on or via the internet site or other means… of accessing the internet by means of which the pornographic material is made available*."[525]

As the task of verifying the age of users proved to be impossible, the U.K. government announced in October 2019 it would abandon the policy.[526] The U.K. law seems however to

---

Porn Away From Teens, THE NEW YORK TIMES, (May 3, 2019), https://www.nytimes.com/2019/05/03/style/britain-age-porn-law.html.

[522] Rachel Thompson,*How The New UK Porn Block Could Put Independent Sex Workers At Risk*, ELLE.COM.UK, (May 15, 2019), https://www.elle.com/uk/life-and-culture/a27416553/uk-block-sex-workers-risk/ .

[523] Jim Waterson, *UK online pornography age block triggers privacy fears*, THE GUARDIAN, (March 16, 2019 05.00 EDT), https://www.theguardian.com/culture/2019/mar/16/uk-online-porn-age-verification-launch.

[524] Lisa Weintraub Schifferle , *Ashley Madison settles with FTC over data security*, THE FTC BLOG, (Dec 14, 2016 12:05PM), https://www.ftc.gov/news-events/blogs/business-blog/2016/12/ashley-madison-settles-ftc-over-data-security.

[525] The Online Pornography (Commercial Basis) Regulations 2019, Art. 2(4), UK Statutory Instruments2019 No. 23 (available at https://www.legislation.gov.uk/uksi/2019/23/regulation/2/made).

[526] Jim Waterson, *UK drops plans for online pornography age verification system*, THE GUARDIAN, Oct. 16, 2019, 19 10.20 EDT), https://www.theguardian.com/culture/2019/oct/16/uk-drops-plans-for-online-pornography-age-verification-system.

have inspired Australia's Parliament's Standing Committee on Social Policy and Legal Affairs to recommend mandatory age verification to access online pornography.[527] It left open the possibility of placing social media sites within the scope of a future age verification law, noting that the UNSW Law Society had argued:

> "*that the 'narrow focus' of the UK scheme meant that children would be able to access pornographic material from free sites, through sharing on mobile phones, or from non-pornographic sites such as Twitter, Reddit, and Imgur*."[528]

France passed on July 30, 2020 a law aiming at protecting woman from domestic abuse, which also modified article 227-24 of the French criminal Code which incriminated "*manufacturing, transporting, disseminating by any means whatsoever and whatever the medium a pornographic message…. when this message is likely to be seen or perceived by a minor*,"[529] which applies to platforms.[530]Article 227-14, as modified, now incriminates such actions even of the minor has declared to be 18 years old of age, the age of majority in France. The offense is punishable by three years' imprisonment and a 75,000 euro fine.

---

[527] Parliament of the Commonwealth of Australia, *Protecting the age of innocence Report of the inquiry into age verification for online wagering and online pornography House of Representatives Standing Committee on Social Policy and Legal Affairs*, available at https://parlinfo.aph.gov.au/parlInfo/download/committees/reportrep/024436/toc_pdf/Protectingtheageofinnocence.pdf;fileType=application%2Fpdf.

[528] Article 3.98.

[529] Representative Bérengère Couillard, *Rapport fait au Nom de la Commission des Lois Constitutionnelles, de la Législation et de l'Administration Générale de la République sur la Proposition de Loi visant à protéger les victimes de violences conjugales (n° 2478*) [Report made on Behalf of the Committee on Constitutional Laws, Legislation Made on Behalf of the Committee on Constitutional Law, Legislation and of the General Administration of the Republic on the Bill to Protect Victims of Domestic Violence, (n ° 2478)], (Jan. 15, 2020), available at https://www.assemblee-nationale.fr/dyn/15/rapports/cion_lois/l15b2587_rapport-fond#_Toc256000023. See Timothy B. Lee, *French parliament passes porn age-verification legislation*, ARS TECHNICA, (July 10, 2020, 11:33A), https://arstechnica.com/tech-policy/2020/07/french-parliament-passes-porn-age-verification-legislation.

[530] Article 227-14 § 2 of the French criminal Code: "*When the offenses provided for in this article are made by … communication to the public online, the specific provisions of the laws governing these matters are applicable with regard to the determination of the persons responsible*."

However, the law is not an age verification scheme, but rather directs an online communications service provider, when receiving a formal notice of the Superior Audiovisual Council informing them that its service allows minors to access pornographic content, to take any measure likely to prevent the access of minors to the incriminated content. It has fifteen days to submit observations. After that, if the content has not been taken down and remains accessible to minors, the President of the Superior Audiovisual Council can seize the President of the Paris court of original jurisdiction to issue an order requiring access to the service to be terminated.[531] The Senator who proposed the amendment which led to article 23 to be added to the July 30, 2020, law argued that the amendment was necessary, even though article 227-24 of the Penal Code is in force, as "*in practice, this article is not applied in the digital world, [as] the justice system fails to reach the editors of these sites, often based in tax havens countries which do cooperate with France*."[532] While article 23 applies to social media platforms as it applies to "*a person whose activity is to publish an online communication service to the public allowing minors to have access to pornographic content*" and social media platform may be used by third parties to publish pornographic material, platforms do not allow their users to publish such material, and it it thus likely that pornographic content would have been identified and deleted, and the account of the user suspended, either temporarily of definitely, before even receiving a formal notice from the French Superior Audiovisual Council.

[531] Loi n° 2020-936 du 30 juillet 2020 visant à protéger les victimes de violences conjugales [July 30, 2020 Law to protect victims of domestic violence] [J.O.] [OFFICIAL GAZETTE OF France] July 31, 2020, article 23.
[532] *Proposition de loi, Protéger les victimes de violences conjugales, (1ère lecture), Amendement présenté par Mme Marie MERCIER au nom de la commission des lois, article additionnel après article 11*, https://www.senat.fr/amendements/2019-2020/483/Amdt_92.html.

While the nature of social media platforms does not make them am easy tool to spread pornographic material, we know now that they may become the unwilling vehicles of information. How can the law respond to this threat to democracy?

IV. Advancing Knowledge and Truth

In a world where everyone can publish, it is very hard to know what to believe.[533]

We are living in times when "post-truth" is a 2016 Oxford Dictionaries Word of the Year.[534] Yet, truth is, according to John Milton, "*a streaming fountain; if her waters flow not in perpetual progression, they sicken into a muddy pool of conformity and tradition*."[535] This Seventeenth century definition is well adapted to our social media word, where posts may indeed be flowing in perpetual progression, 24/7, and where many of them are of the post-truth types, making the platforms muddy post-truth pools.

Does the law require speech to be truthful? Justice Powell wrote in *Gertz v. Robert Welch* that "*there is no constitutional value in false statements of fact*."[536] However, Justice Black argued in his concurring opinion in Rosenbloom (a case which was disapproved by *Gertz*) that in his view, "*the First Amendment does not permit the recovery of libel judgments against the news media even when the statements are broadcast with knowledge they are false,*"[537] because "*the First Amendment was intended to leave the press free from the harassment of libel judgments*."[538] As stated by the Supreme Court in *New York Times v.*

---

[533] LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE, 171, (Basic Books 2000).
[534] Word of the Year 2016, OXFORD LANGUAGES, https://languages.oup.com/word-of-the-year/2016/.
[535] JOHN MILTON, AREOPAGITICA, in COMPLETE POEMS AND MAJOR PROSE 716 (Merritt Hughes ed., 1957), cited by ETERNALLY VIGILANT FREE SPEECH IN THE MODERN ERA, 64 (Lee C. Bollinger & Geoffrey R. Stone eds., The University of Chicago Press (2002).
[536] Gertz v. Robert Welch, Inc, at 340
[537] Rosenbloom, at 57.
[538] Justice Black concurring and dissention opinion in *Curtis Publishing Co. v. Butts*, at 172.

Sullivan,[539] "*erroneous statement[s*] *are inevitable in free debate, and … must be protected if the freedoms of expression are to have the "breathing space" that they "need . . . to survive*…""[540] Professor Alan Chen recognized in an article published in 2020 that the First Amendment "*likely represents a significant barrier to… efforts* [*to address the issue of fake news*]"[541] and posits that "*laws regulating fake news maybe but failed attempts to force people to behave rationally in a universe that is full of irrationality*."[542] Professor Chen believes, however, that "*lies… have some social value, not just in a utilitarian sense by promoting other public good but also by enhancing individuals' ability and freedom to internally experience self-realization*."[543]

Protecting "fake news" may also be a way to affirm a country's dedication to democracy and the plurality of speech. In a March 3, 2017 Joint Declaration, the United Nations Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe Representative on Freedom of the Media, the Organization of American States Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights Special Rapporteur on Freedom of Expression and Access to Information expressed their alarm at instances of denigration, intimidation and threats to the media by public authorities,

> *"including by stating that the media is "the opposition" or is "lying" and has a hidden political agenda, which increases the risk of threats and violence against journalists,*

[539] New York Times v. Sullivan, 376 U.S. 254, 271-272, (1964).
[540] Quoting N. A. A. C. P. v. Button, 371, U. S. 415, 433 (1963).
[541] Alan K. Chen, *Fake News, Rational Deliberation, and Some Truths About Lies*, 61 WM. & MARY L. REV. 357, 361 (2020).
[542] Ibid., at 399.
[543] Ibid., at 404.

undermines public trust and confidence in journalism as a public watchdog, and may mislead the public by blurring the lines between disinformation and media products containing independently verifiable facts."

It noted, however, that "[*g*]*eneral prohibitions on the dissemination of information based on vague and ambiguous ideas, including "false news" or "non-objective information", are incompatible with international standards for restrictions of freedom of expression, and should be abolished*."[544] This is in line with U.S. jurisprudence, as Justice O'Connor wrote in Virginia v. Black that "*the First Amendment "ordinarily" denies a State "the power to prohibit dissemination of social, economic and political doctrine which a vast majority of its citizens believes to be false and fraught with evil consequence*," citing the 1927 concurring opinion of Justice Brandeis in Whitney v. California. However, the Supreme Court explained in United States v. Alvarez[545] that "*false speech* [*is not*] *a general category that is presumptively unprotected*." The European Court on Human Rights regularly states that article 10 of the Convention "*protects journalists' right to divulge information on issues of general interest provided 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*."[546] Social media users do not have to follow ethic rules, and, as a result, social media platforms have however allowed the proliferation of false information being published online, to great danger to democracy. A response must be made to the spread of "fake news." But which response? Wendell Phillips wrote that "[*t*]*he community that will not protect its most*

[544] Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda, FOM/3/17, March 3, 2017, https://www.osce.org/fom/302796?download=true.
[545] United States v. Alvarez, 567 U.S. 709, 722 (2012).
[546] *See* for instance, see Fressoz and Roire v. France [GC], no. 29183/95, § 54, ECHR 1999-I.

ignorant and unpopular member in the free utterance of his opinions, ***no matter how false or hateful****, is only a gang of slaves"* (our emphasis).[547] However, the Supreme Court noted in a dictum in 1915 that "[*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 not on the conscience of judges and juries but on the competition of others ideas."*[548] The Court noted however that false statements have no constitutional values.

While U.S. adults in their sixties or even fifties are still relying on the papers or television to be informed, young ones are relying more and more on social media.[549] However, people relying on social media to be informed "*tends to pay less attention to news than those who rely on most other pathways*."[550] Is fake news a threat to the electoral process, thus a threat to democracy, and must thus be regulated? (A) Are some truths self-evident, beyond those recognized by the Declaration of Independence, that they must not be denied? This is the case in Europe where some countries have criminalized Holocaust denial[551] (B).

---

[547] Wendell Phillips, THE SCHOLAR OF THE REPUBLIC, cited by James Brewer Stewart, THE CONSTITUTION, THE LAW, AND FREEDOM OF EXPRESSION 1787-1987, p. 1, Southern Illinois University Press, 1987.
[548] Mutual Film Corp. v. Industrial Commission of Ohio, 236 U.S. 230 (1913).
[549] Source: Pew Research Center Survey of US adults, October 29-November 11, 2019, *see* Amy Mitchell, Mark Jurkowitz, J. Baxter Oliphant and Elisa Shearer, *Americans Who Mainly Get Their News on Social Media Are Less Engaged, Less Knowledgeable*, PEW RESEARCH CENTER, (July 30, 2020), https://www.journalism.org/2020/07/30/americans-who-mainly-get-their-news-on-social-media-are-less-engaged-less-knowledgeable. For instance, in early June of 2020, just a few months before the Presidential elections, considered by many as crucial for the future of democracy, only 8% of U.S. adults getting most of their political news from social media say they are following news about the 2020 election "very closely," while the figure for those who getting their news from cable TV (37%) and print (33%) were much higher.
[550] Ibid.
[551] We will see later on that Facebook recently forbade posting speech denying the Holocaust on its platform.

A. Fake News and the Safety of the Electoral Process

What is fake news? A report published in 2018 by the European Commission and written by the *independent High-level Group on fake news and online disinformation* (HLEG) noted that the definition many include:

> "*relatively low-risk forms such as honest mistakes made by reporters, partisan political discourse, and the use of click bait headlines, to high-risk forms such as for instance foreign states or domestic groups that would try to undermine the political process… through the use of various forms of malicious fabrications, infiltration of grassroots groups, and automated amplification techniques.*"[552]

a. In the U.S.

"*The Fake News Media, the true Enemy of the People, must stop the open & obvious hostility & report the news accurately & fairly***."** Donald J. Trump (@realDonaldTrump). October 29, 2018.**[553]**

---

[552] *Communications Networks, Content and Technology, A multi-dimensional approach to disinformation Report of the independent High level Group on fake news and online disinformation*, EUROPEAN COMMISSION (2018), p. 10, available to download at *Shaping Europe's digital future, Final report of the High Level Expert Group on Fake News and Online Disinformatio*n, EUROPEAN COMMISSION (MARCH 12, 2018), https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation. The HLEG noted further that it had avoided to use the term "fake news" in its report for two reasons: (1) because it "*is inadequate to capture the complex problem of disinformation, which involves content that is not actually or complete-y "fake" but fabricated information blended with facts…, and practices …. includ[ing] some forms of automated accounts used for astroturfing, networks of fake followers, fabricated or manipulated videos, targeted advertising, organized trolling, visual memes, and much more" and (2) because the term "also involve a whole array of digital behaviour that is more about circulation of disinformation than about production of disinformation, spanning from posting, commenting, sharing, tweeting and re-tweet-ing etc.*" (HLEG Report, p. 10).

[553] Donald J. Trump (@realDonaldTrump), Twitter (Oct. 29, 2018), https://twitter.com/realDonaldTrump/status/1056879122348195841?ref_src=twsrc%5Etfw.

President Trump posted two tweets from his @realDonaldTrump account on May 26, 2020, which warrant being cited in their entirety for the sake of our discussion:

> "*There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forged & even illegally printed out & fraudulently signed. The Governor of California is sending Ballots to millions of people, anyone.....*" then "*.... living in the state, no matter who they are or how they got there, will get one. That will be followed up with professionals telling all of these people, many of whom have never even thought of voting before, how, and for whom, to vote. This will be a Rigged Election. No way*!"[554]

Twitter added to the bottom of each of these tweets the phrase "*! Get the facts about mail-in ballots*" which formed a blue link towards a page informing user on its top about:

> "*What you need to know. Trump falsely claimed that mail-in ballots would lead to "a Rigged Election." However, fact-checkers say there is no evidence that mail-in ballots are linked to voter fraud. Trump falsely claimed that California will send mail-in ballots to "anyone living in the state, no matter who they are or how they got there." In fact, only registered voters will receive ballots. Five states already vote entirely by mail and all states offer some form of mail-in absentee voting, according to NBC News.*"[555]

---

[554] Donald J. Trump, @realDonaldTrump, Twitter (May 26, 2020, 8 :17AM), https://twitter.com/realDonaldTrump/status/1265255835124539392, and Donald J. Trump, @realDonaldTrump, Twitter (May 26, 2020, 8 :17AM), https://twitter.com/realDonaldTrump/status/1265255845358645254.
[555] TWITTER, https://twitter.com/i/events/1265330601034256384, (last visited Dec. 30, 2020).

The page also presented timeline of posts refuting the claims of the President, often linking to pages allowing users to gather information at the source, such as a page from the office of the Governor of California website publishing the press release informing California voters that Governor Newsom had issued an executive order allowing every registered California voters to vote by mail in the November 2020 general election. This was the first time that Twitter had taken such action, and the President reacted by posting the same day:

> *"....Twitter is completely stifling FREE SPEECH, and I, as President, will not allow it to happen!"*[556] and *".@Twitter is now interfering in the 2020 Presidential Election. They are saying my statement on Mail-In Ballots, which will lead to massive corruption and fraud, is incorrect, based on fact-checking by Fake News CNN and the Amazon Washington Post...."*

The President's statement is inaccurate, as the First Amendment prohibits Congress from making laws abridging freedom of speech or of the press but does not prohibits private speakers from abridging freedom of speech. Also, merely adding a link to the tweets warning users that the facts stated by the President were not accurate is not abridging speech but is instead participating in the "marketplace of ideas." Judge Learned Hand explained in 1943 that the First Amendment "*presupposes that right conclusions are more*

[556] Donald J. Trump, @realDonaldTrump, Twitter (May 26, 2020, 7:40 PM), https://twitter.com/realDonaldTrump/status/1265427539008380928 and Donald J. Trump, @realDonaldTrump, Twitter (May 26, 2020, 7 :40 PM), https://twitter.com/realDonaldTrump/status/1265427538140188676.

likely to be gathered out of a multitude of tongues, than through any kind of authoritative selection. To many this is, and always will be, folly; but we have staked upon it our all."[557]

The President posted the next day from his Twitter account:

"*Twitter has now shown that everything we have been saying about them (and their other compatriots) is correct. Big action to follow!"*, adding *".@Twitter is now interfering in the 2020 Presidential Election. They are saying my statement on Mail-In Ballots, which will lead to massive corruption and fraud, is incorrect, based on fact-checking by Fake News CNN and the Amazon Washington Post..."*"[558]

Then, on May 28, 2020, President Trump issued an Executive Order on Preventing Online Censorship (EO)[559] which directed the Secretary of Commerce, in consultation with the Attorney General, to request the Federal Communications Commission to "*expeditiously propose"* regulations clarifying when providers of interactive computer services, such as social media platforms, could benefit or not from the immunity provided Section 230 (c)(2)(A) of the Computer Decency Act (CDA) when screening offensive content. The EO called this "*practice*… *fundamentally un-American and anti-democratic*," arguing that, when "*powerful social media companies censor opinions with which they disagree, they exercise a dangerous power*" and "*are engaging in selective censorship that is harming our national discourse*." Section 230 (c)(2)(A) of the CDA provides that:

---

[557] United States v. Associated Press, 52 F. Supp. 362, 372, (D. C. S. D. N. Y. 1943).
[558] Donald J. Trump, @realDonaldTrump, Twitter (May 27, 2020, 10 :22 AM), https://twitter.com/realDonaldTrump/status/1265649545410744321.
[559] *Executive Order on Preventing Online Censorship*, THE WHITE HOUSE, https://www.whitehouse.gov/presidential-actions/executive-order-preventing-online-censorship/.

"provider or user of an interactive computer service" are not civilly liable for "*any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected*."

The EO argued that "[*i*]*t is the policy of the United States to ensure that, to the maximum extent permissible under the law, this provision is not distorted to provide liability protection for online platforms that — far from acting in "good faith" to remove objectionable content — instead engage in deceptive or pretextual actions (often contrary to their stated terms of service) to stifle viewpoints with which they disagree."* We will discuss later in detail Section 230 and the bills aiming at amending it.

Facebook announced on September 3, 2020, that it would add a label to candidate or campaigns' posts declaring victory before the final results of an election and would "*direc*[*t*] *people to the official results from Reuters and the National Election Pool*."[560] It also vowed not to accept new political ads in the week before the election, and to "*attach an informational label to content that seeks to delegitimize the outcome of the election or discuss the legitimacy of voting methods, for example, by claiming that lawful methods of voting will lead to fraud*." We will go back to this issue in the part of the article about the law of the platforms.

[560] *New Steps to Protect the US Elections*, FACEBOOK, (Sep. 3, 2020), https://about.fb.com/news/2020/09/additional-steps-to-protect-the-us-elections, (last visited Dec. 30, 2020).

Copyright laws can also be used to fight disinformation. On June 18, 2020, Twitter added a *"! Manipulated Media*" warning to a tweet posted by Donald Trump which showed a video of a white toddler running after a black toddler with a chyron made to look like a CNN chyron stating "*Breaking News, Terrified Todler* (sic) R*uns From Racist Baby*", then "*Breaking News, Racist Baby Probably a Trump Voter*."[561] The warning linked to a dedicated page explaining that CNN had actually reported in September 2019 a video about the friendship between two toddlers, one white, the other black. The version of the report shared by the President had been edited and doctored, adding a fake CNN chyron, which misspelled "toddler."[562] The video was taken down the next day after one of the parents sent a DMCA take down claim.[563]

[561] @realDonaldTrump, Twitter, (June 18, 2020, 8:12 PM), https://twitter.com/realDonaldTrump/status/1273770669214490626.
[562] https://twitter.com/i/events/1273790055513903104
[563] Donie O'Sullivan, *Parent of toddler in 'manipulated' Trump video forces Facebook and Twitter to remove it*, CNN, (June 19, 2020, 6:44PM ET), https://www.cnn.com/2020/06/18/business/trump-video-twitter-manipulated-media/index.html. The parents of both toddlers later filed a suit in the Supreme Court of the State of New York against the man who had created the video and against Donald Trump, claiming, *inter alia*, that the video and tweeting the video had violated the right of publicity of their children, Michael Cisneros, et al. v. Logan Cook, et al., Index No. 157550-2020 (Sep. 17, 2020). The complaint alleged that Donald Trump had published the image of the two toddlers to advertise himself and his brand "*…and advance his economical and political goals for his own business purposes rather than in connection with a broadcast or social media communication known as a "tweet" that was (a) informative and (b) concerned a matter of public interest or current news interest…*" (Complaint, at 29). Donald Trump moved to dismiss on December 1, 2020, arguing that the New York right of publicity law, New York Civil Right Law, Section 50 & 51, does not apply to the case as the use of the toddlers' image had not been commercial: "*the usage here was clearly intended to convey a political and artistic image*." Donald Trump claimed further that, even if the use of the image is commercial, it would still not be actionable because New York right of publicity law carves out an exception for newsworthy, artistic, or incidental use. Donald Trump claims that the lawsuit "*is political in nature* "as *"Plaintiffs and their attorney of record "are active Democrats and supporter of candidate Joe Biden*." Donald Trump also argued that the suit was *"improperly directed at Defendants' exercise of their right to publicly petition and participate*" and thus violated New York's anti-SLAPP law, N.Y. Civ. Rights Law § 70-A. See Eriq Gardner, *Donald Trump Stands Up for Legal Right to Retweet a Meme*, (December 03, 2020 11:06am PT), THE HOLLYWOOD REPORTER, https://www.hollywoodreporter.com/thr-esq/donald-trump-stands-up-for-legal-right-to-retweet-a-meme.

b. French "Fake News" Laws

One of the tasks of the Commission for the Enrichment of the French Language (*Commission d'enrichissement de la langue française*), a group of French personalities placed under the authority of the French Prime Minister, is to create French words or neologisms to replace words, most of them English, created and used in the global community. One of such words is "fake news." In October 2018, it offered two alternatives for "fake news", "*information fallacieuse*" or "*infox*." The Commission specified that these two expressions were to be used when *"designating misleading or deliberately biased information, for example to favor one political party over another, to tarnish the reputation of a person or a company, or to contradict a scientifically established truth*." The Commission also noted that "[*i*]*t will also be possible, in particular in a legal framework, to use the terms used in the [French Press Law] as well as in the electoral code, the criminal code or the monetary and financial code: "false news," or " false informatio*n."[564]

That said, false information is protected speech in France. The *Cour de cassation* held in 2013 that "*freedom of expression is a right the exercise of which is only abusive in cases specially determined by law, and comments… even if they are untrue, do not come within the scope of none of these cases*."[565] In this case, a woman had falsely alleged on a website to have founded a museum in Normandy commemorating the legacy of British airmen during

[564] Recommandation sur les équivalents français à donner à l'expression fake news [Recommendation on the French equivalents to be given to the expression fake news], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIEL GAZETTE OF France], Oct. 4, 2018, available at https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037460897.

[565] Cour de cassation [Cass.] [supreme court for judicial matters] 1e civ. Ap. 10, 2013, Bull. I, n° 67 (Fr.). Available at https://www.legifrance.gouv.fr/affichJuriJudi.do?oldAction=rechJuriJudi&idTexte=JURITEXT000027303673&fastReqId=656168342&fastPos=1.

WWII. The Court held that the woman had the right to publish this false information, referring to article 10 of the ECHR which does not take out of its scope false information. The European law thus acknowledges, as the U.S. law does, that the "marketplace of ideas" will eventually regulates itself.

There are, however, several French laws prohibiting the publication of fake news. Article 322-14 of the French penal Code incriminates "*communicating or disclosing false information with the aim of making believe that destruction, degradation or a deterioration dangerous for people will be or has been committed.*" It is punishable by two years' imprisonment and 30,000 euros fine. Article 224-8 of the French penal Code makes it a crime punishable by five years' imprisonment and a fine of 75,000 euros to communicate or attempt to communicate false information to "*knowingly compromise the safety of an aircraft in flight or of a ship*." Article 224-8 can apply to social media posts, but, fortunately, courts have not yet had the opportunity to confirm it.

Article 27 of the French Press Law incriminates publishing, distributing, or reproducing "*by any means whatsoever false, fabricated, falsified news or news deceptively attributed to third parties when, done in bad faith, it will have disturbed the public peace, or will have been likely to disturb it.*" This crime is punishable by 45,000 euros fine. If such act "*is likely to shake the discipline or the morale of the armies or to hamper the nation's war effort*," the fine is increased to 135,000 euros. The law is however rarely the basis of criminal proceedings.[566] The law applies to social media platforms and thus a tweet or post

[566] Emmanuel Dreyer, Fausse bonne nouvelle : la loi du 22 décembre 2018 relative à la manipulation de l'information est parue, LEGIPRESSE N ° 367, 19. According to Professor Dreyer, there has been no criminal proceedings based on Article 27 for some fifty years.

could be incriminated if it disturbs the public peace or is likely to do so, and if the information posted is "news," that is, "*announcing a precise and detailed fact, current or past, but not yet disclosed*."[567] As such, "*biased and misleading statements or comments about a fact which had already been revealed*"[568] are not withing the scope of article 27. In one case, a man had been charged for spreading false news under article 27 for having distributed pamphlets stating that the drowning of a young man, while the police was trying to arrest him, was not an accident, but a racist crime. The Court of appeals remanded, reasoning that the article 27 crime "*requires the existence of a "news," this term being taken in the sense of "announcement of an event that has happened recently, made to someone who is not yet aware of it*." The Court of appeals considered, however, the pamphlets to be defamatory for the police: commenting news may be a defamation, if outrageous, but not "fake news."

Article 27 also requires that the publication has been made in bad faith, that is, with the knowledge that the news is false.[569] The publication of the "fake news" must also have disturbed "*public peace*," and, as such, must "*contain by its object a ferment of public*

---

[567] Cour d'appel de Paris (CA) [regional court of appeals], May 18, 1988, D. 1990. 35.
[568] Cour de cassation [Cass.] [supreme court for judicial matters], Crim. April 13, 1999, Bull. crim. N° 78 (Fr.). Available at https://www.legifrance.gouv.fr/affichJuriJudi.do?oldAction=rechJuriJudi&idTexte=JURITEXT000007070114&fastReqId=1251072116&fastPos=3. "…*it follows that comments, however shocking they may be, relating to previously revealed facts cannot fall within the scope of this text; … in this case, … the incriminated leaflet concerns the death of Sada X …, which occurred on July 8, 1996, an event which immediately gave rise to multiple information and comments, and retain that, if the leaflet presents these facts in a tendentious way and includes unfounded assertions and imputations undermining the honor and the consideration of the military policemen, constituting the crime of defamation, the distribution of such a document does not come within the scope of application of the text referred to in the prosecution*." The *Cour de cassation* approved the reasoning of the Court of appeals.
[569] Cour de cassation [Cass.] [supreme court for judicial matters], Crim. March 16, 1950, Bull. crim. N° 100 (Fr.).

disturbance, that is, disorder, panic, collective emotion and disarray."[570] This definition by the Paris Court of appeals reminds us that the public may or may not react to a particular news item in such a way that it will create panic. A "fake news" tweet, if meet by derision, doubt, or cynicism, instead to panic and disarray, would likely not be incriminated under article 27. The cynicism of the public, or a healthy dose of skepticism, can thus be a defense… In March 2020, after the world finally understood the threat represented by the COVID-19 virus, shoppers rushed around the world to stock toilet paper,[571] a move which quickly led to shortages. News of shortages were relayed on social media, and numerous posts featured videos of shopping carts piled high with the newly precious commodity.[572] Did panic lead to shortages, or did shortages lead to panic? We do not know it and it would be a difficult task to trace panic and disarray to a particular message, at least on social media, where virality may quickly sows confusion while creating an illusion of truth, especially if a false news is reposted in good faith by a person of confidence.

Another French law protects the integrity of the electoral process. Article L 52-1 of the French Electoral Code, which is part of Chapter V of the French Electoral Code on "*propaganda*" provides that:

> "[*d*]*uring the six months preceding the first day of the month of an election and until the date of the ballot on which it is won, the use for electoral propaganda purposes of any*

[570] Cour d'appel de Paris (CA) [regional court of appeals], Jan. 7, 1998, Dr. penal 1998, 63.
[571] Frances Mao, *Coronavirus panic: Why are people stockpiling toilet paper?* BBC NEWS, (March 4, 2020), https://www.bbc.com/news/world-australia-51731422.
[572] Adam Westbrook, *People Around the World Are Panic-Buying … Toilet Paper?* THE NEW YORK TIMES, (March 12, 2020), https://www.nytimes.com/2020/03/12/opinion/toilet-paper-coronavirus.html.

process of commercial advertising by means of the press or by any means of audiovisual communication is prohibited."

Social media platforms are thus within the scope of the law. As such, the law prohibits using social media platforms as part of an electoral strategy campaign. Article L 97 of the same Code makes it as crime punishable by one year in jail and 15,000 euros fine to have deflected votes or to have influenced one or several voters to abstain from voting, if done so "*with the help of false news, slanderous rumors or other fraudulent maneuvers.*"

This law can be used to protect consumers-as-voters by making mandatory to inform them that a political message has been sponsored and thus protects the transparency and integrity of the electoral campaigns. Social media platforms have also a general obligation to inform this information to users, as article L.111-7- II of the French Consumer Code requires them "*to provide consumers with fair, clear and transparent information on:*

1 ° The general conditions of use of the intermediation service that it offers and the terms of referencing, classification and de-referencing of the content, goods or services to which this service provides access;

2 ° The existence of a contractual relationship, of a capital link or of remuneration for its benefit, since they influence the classification or referencing of the content, goods or services offered or posted online;

3 ° The quality of the advertiser and the rights and obligations of the parties in civil and fiscal matter, when consumers are put in contact with professionals or non-professionals."

It is also illegal, under a July 20, 1977 French law, to publish or disseminate any opinion poll having a direct or indirect connection with a referendum, a presidential election or any other election regulated by the French election Code, as well as the election of representatives to the European Parliament.[573] Article 11 of the law expressively forbids, the day before an election, and the day of the election to publish to disseminate or to comment, by any means whatsoever, on such polls. Such offense is punishable by 75,000 euros fine, under article L 90-1 of the electoral Code.

However, this law applies only to publications in France, and thus polls can be leaked by social media accounts located abroad or purporting to be. Even if posting from France, social media platforms can be used to allude to the poll results, even if forbidden by law. Ahead of the 2012 Presidential election, Paris District Attorney François Molins stated that "[*i*]*n cooperation with the Paris judicial police, a device was adopted allowing the Paris prosecutor's office, in the event of violation of this ban, to immediately seize for investigation the brigade for the repression of delinquency… of the regional direction judicial police*,"[574] thus indicating how serious the offense of publishing poll results ahead of the 8:00 PM official election result announcement was considered to be. At the time, article L. 52-2 of

[573] Loi n° 77-808 du 19 juillet 1977 relative à la publication et à la diffusion de certains sondages d'opinion [Law No. 77-808 of July,19 1977 on the publication and dissemination of some opinion polls] JOURNAL OFFICIEL DE LA RÉ'PUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], July 20, 1977 p.3837.
[574] M.S., *«Fuites» des résultats avant 20 heures : Sarkozy et Hollande s'emparent du débat*, LE PARISIEN, (April 19, 2012, 15h08), https://www.leparisien.fr/elections/presidentielle/fuites-des-resultats-avant-20-heures-sarkozy-et-hollande-s-emparent-du-debat-19-04-2012-1962219.php.

the electoral Code forbade to publish election results, whether partial or final *"in metropolitan France, before the closing of the last office. of voting in the metropolitan territory*" and, in the overseas departments, "*before the closing of the last polling station in each of the departments concerned*."[575] The ban was circumvented by using code words to communicate the results in "code," such as "*The sun is shining in the Netherlands*" to announce that François Hollande was ahead in the polls.[576]

"Fake news" published to influence the results of an election are particularly pervasive. Online rumors claiming falsely that presidential candidate Emmanuel Macron had several offshore accounts were published on social media during the 2017 presidential election campaign,[577] leading to Macron filing lawsuits claiming that these publications were "*linked to Russian interests*."[578] Once elected President, he supported passing a law incriminating attempts to sway a Presidential election by publishing false information. Two bills aiming at addressing the issue of fake news were introduced in March 2018 by all the Representatives members of the *La République en Marche* political movement, created by the new President, one ordinary law bill, and one organic law bill addressing the issue of fake new during the Presidential election. The organic law bill was introduced in the French Chamber of Representatives March 16, 2018[579] and aimed at making applicable to the presidential election the provisions introduced by the bill "*aiming at fighting false*

---

[575] Article L 52-2 has been modified twice since 2012, but the general prohibition remains.
[576] The French often refer to The Netherlands as "La Hollande", not "Les Pays-Bas", its correct name.
[577] Sputnik, the Russian government-controlled news agency, reported falsely during the French Presidential campaign that candidate François Fillon was leading the race. See *How the world was trolled*, THE ECONOMIST, (Nov. 4, 2017), 22.
[578] *French election: Macron takes action over offshore claims*, BBC NEWS, (May 4, 2017), https://www.bbc.com/news/world-europe-39802776.
[579] Proposition de loi organique relative à la lutte contre les fausses informations, n° 772 [Organic law bill relating to the fight against false information, n ° 772], available at http://www.assemblee-nationale.fr/dyn/15/dossiers/lutte_fausses_informations?etape=15-AN1-DEPOT.

information" introduced in the French Chamber of Representatives a few day later, March 21, 2018.[580]

The ordinary law bill explained that such law was necessary as:

> "[*r*]*ecent electoral news has demonstrated the existence of massive campaigns to disseminate false information intended to alter the normal course of the electoral process through online communication services. By its importance in the democratic life of the Nation and the special place occupied by the President of the Republic in our institutions, the campaign for the presidential election is particularly threatened by the massive dissemination of false information. It is therefore necessary to make applicable to the presidential campaign the common law system put in place by the ordinary law on the fight against false information*."

The Council of State (*Conseil d'État)*, France's highest administrative Court, stated, in its advisory opinion about the two bills, its concern about the use of fake news on social media to influence the outcome of elections, including by foreign actors, noting that:

> "*the echo given to this false information has been amplified by digital platforms, in particular social networks, whose economic logic leads to promoting, in particular, the content for the promotion of which they were paid and those causing the most controversy*."[581]

---

[580] Proposition de loi relative à la lutte contre les fausses informations [Bill n ° 799 on the fight against false information], available at http://www.assemblee-nationale.fr/dyn/15/textes/l15b0799_proposition-loi#.
[581] Conseil d'État, *Avis consultatif du Conseil d'État sur la lutte contre les fausses informations* (May 4, 2018) paragraph 7, https://www.conseil-etat.fr/ressources/avis-aux-pouvoirs-publics/derniers-avis-publies/lutte-contre-les-fausses-informations.

While noting that French laws already protect against false information, as we saw, there was nevertheless a need for a new law, because "*recent news have revealed that the dissemination of false information is now carried out according to new logics and vectors*." As such, the French law was passed because the way social media platforms have been used to influence an election. The motives of the bill explained that platforms, including social media "*are used in a massive and sophisticated manner by those who wish to spread false information*."

The "law on the fight against the manipulation of information"[582] was finally enacted, with difficulties however, as the French Senate first rejected it refusing to even examine it. During the debates at the lower Chamber, Representative Alexis Corbière wondered how "*a single judge, who is not necessarily specialized in the field at stake, sometimes very technical*" would not be able to *" disentangle the truth from the false*" and argued that "[*o*]*nly the rigorous verification work of a pluralist and independent press*," along with referring the case to an ordinary judge, not to an emergency judge, would guarantee freedom of expression.[583] His party, *La France Insoumise*,[584] had proposed an amendment to the bill which would have required the emergency order to be issued by several judges, not just one, arguing that it was "*necessary, given the consequences of the ruling on the elections and*

[582] Loi no 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information, [Law 2018-1202 of 22 December 2018 relating to the fight against the manipulation of information] JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O] [OFFICIAL GAZETTE OF FRANCE], Dec. 23, 2018, text n° 2
[583] Compte rendu N° 75 Commission des lois constitutionnelles, de la législation et de l'administration générale de la République, [Report N. 75 of the Commission of Constitutional Laws, Legislation and General Administration of the Republic] May 23, 2018, http://www.assemblee-nationale.fr/dyn/15/comptes-rendus/cion_lois/l15cion_lois1718075_compte-rendu#.
[584] *La France Insoumise is* a left-wing party.

the "Streisand effect" it entails … as a court decision can have a magnifying effect on information that would otherwise have gone more or less unnoticed."

The constitutional Council found the law constitutional on December 20, 2018,[585] noting that the law sought:

> "*to ensure the clarity of electoral debate and compliance with the principle of the honesty of elections*"[586] and that the law "*strictly defined the information which may be subject to interlocutory proceedings…. [which] are only intended for incorrect or misleading allegations or accusations which have the effect of altering the honesty of the upcoming elections. These allegations or accusations do not relate to opinions, parodies, partial inaccuracies or simple exaggerations. They are those for which it is possible to objectively demonstrate falseness.* "[587]

The law created an emergency procedure, article L163-2 of the electoral Code, which can be used to stop the dissemination of "*inaccurate or misleading allegations or imputations of a fact likely to alter the sincerity of the upcoming election"* if done "*in a deliberate, artificial or automated and massive manner by means of an online communication service to the public."* The scope of the law is limited to the "*three months preceding the first day of the month of general elections and until the date of the ballot in which they are acquired."* The district attorney, a candidate, a political party of or any person having standing (*intérêt à agir*) can seize the emergency judge (*juge des référés*), acting alone, to

[585] Conseil constitutionel [CC] [Constitutional Court] decision No.2018-773 DC, December 20, 2018, available in English at https://www.conseil-constitutionnel.fr/en/decision/2018/2018773DC.htm.
[586] Decision No.2018-773 DC, December 20, 2018, paragraph 18.
[587] Decision No.2018-773 DC, December 20, 2018, paragraph 21.

issue an emergency order within forty-eight hours requiring hosting providers, or failing that, the Internet service providers, to take "*all proportionate and necessary measures to stop this distribution*."

Interestingly, article L163-2 was first used to ask Twitter to delete a tweet posted by Emmanuel Macron's then-Minister of police, Christophe Castaner on May 1st, 2019, International Labor Day, which had seen protests in Paris. The post read: "*Here at La Pitié-Salpêtrière,*[588] *one has attacked a hospital. Its caregivers have been assaulted. And one has injured a police officer mobilized to protect it. Unfailing support for our law enforcement agencies: they are the pride of the Republic.*" Two elected officials seized the emergency judge, asking for an order to take down the tweet as having disseminated an inaccurate information. The Paris Court of original jurisdiction, the *Tribunal de Grande Instance de Paris*, in its capacity as emergency court, refused to do so.[589]

The plaintiffs had argued that the tweet had been posted to make electors believe that there was:

> "*a climate of violence to bring into play the spring of fear and chaos, which can only disrupt the campaign for the European elections*." The Court was not convinced, noting that, while "*the tweet may have used exaggerated terms… it had not obscured the debate, since it had been immediately disputed, that many articles in the press or the web had indicated that the facts did not unfold this way [and]… different versions have emerged,*

[588] A Parisian hospital.
[589] Tribunal de Grande Instance de Paris [TGI] [ordinary court of original jurisdiction], May 10, 2019, RG 19/53935, available at https://www.dalloz-actualite.fr/sites/dalloz-actualite.fr/files/resources/2019/05/19_53935.pdf.

thus allowing each voter to form an enlightened opinion, without obvious risk of manipulation."

The plaintiffs had produced as evidence an article published by right-wing newspaper *Le Figaro* about the incident, which related that the protesters had indeed entered the hospital without permission, but that the scene was "*very brief and non-violent."* The Court noted that the Paris District Attorney had opened an enquiry about the incident, that thirty-two persons had been arrested, and that the director of the hospital had described some of the protesters as menacing. Center-left newspaper *Le Monde* had related that, when protesters had gathered at the steps of a stair, several hospital employees had quickly tried to block their entry. Considering these two news reports, as well as another one, the Court concluded that:

"*the intensive care unit was not the object of an attack by the demonstrators who remained outside the building and the nursing staff was not injured. [While] the message written by Mr. Christophe Castaner appears to be exaggerated in its evocation of an attack and injuries, this exaggeration relates to facts which are real, the intrusion of demonstrators in the enclosure of the hospital of the Pitié-Salpêtrière on May 1, 2019*. *As the information is not devoid of any link with real facts, the condition according to which the allegation must be manifestly inaccurate and misleading is not met*."

The Court also noted that that article L.163-2 of the Electoral Code requires that the information at stake has been distributed by artificial or automated means, and there were no element tending to prove that the Minister's tweet had been posted using such means. The Court thus made clear that the scope of the law is limited to messages posted by

bots,[590] and would probably have reached a different conclusion if the tweet had been posted repeatedly and relayed by bots. However, a simple tweet, while inaccurate, reflects personal bias, not propaganda.[591]

The law also aims at providing transparency in political sponsorship, a concern likely to have been triggered by the *Cambridge Analytica* scandal in 2018,[592] Russia' s interference in the 2016 U.S. Presidential elections[593] and in the 2018 French Presidential elections.[594] Indeed, the law addresses the issue of political-ads transparency, by creating

---

[590] French Representative Naïma Moutchou, rapporteur of the bill, had noted during the debate that "[*t*]*he objective [of the bill] is to fight against the artificial and massive dissemination of false information - sponsored content, use of bots - based on bad faith, and therefore on the desire to do harm.*" See Compte rendu N° 75 Commission des lois constitutionnelles, de la législation et de l'administration générale de la République, [Report N. 75 of the Commission of Constitutional Laws, Legislation and General Administration of the Republic] May 23, 2018, http://www.assemblee-nationale.fr/dyn/15/comptes-rendus/cion_lois/l15cion_lois1718075_compte-rendu#.

[591] If there is a bias on social media, it may be created by the users, not the content moderation practices. The Pew Research Center reported in October 2020 that the small percentage of very active users on Twitter, that is, the 10% of users posting 92% of all tweets from U.S. adult, are mostly Democrats, *See Differences in How Democrats and Republicans Behave on Twitter,* THE PEW RESEARCH CENTER, (Oct.15, 2020), https://www.pewresearch.org/politics/2020/10/15/differences-in-how-democrats-and-republicans-behave-on-twitter.

[592] Nicholas Confessore, *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*, THE NEW YORK TIMES, (APRIL 4, 2018), https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html. See also Mark Zuckenberg, FACEBOOK, (March 21, 2018), https://www.facebook.com/zuck/posts/10104712037900071. The data analytics company harvested tens of millions of Facebook users 'profiles, unbeknownst to them, or, apparently, to Facebook. A data scientist had built a personality quiz app which was able to harvest data from Facebook's users who had used the app. The data thus collected was later acquired by Cambridge Analytica which used it to build a proprietary algorithm, used by the Trump's Presidential campaign. Steve Bannon was at the time one of Cambridge Analytica's Vice Presidents and had been at the origin of the creation of the company, funded by Republican donor Robert Mercer. Bannon later became Donald Trump's senior adviser.

[593] Mark Mazzetti, *G.O.P.-Led Senate Panel Details Ties Between 2016 Trump Campaign and Russia*, THE NEW YORK TIMES, (Aug. 18, 2020), https://www.nytimes.com/2020/08/18/us/politics/senate-intelligence-russian-interference-report.html. See also Erik Brattberg and Tim Maurer, *Russian Election Interference: Europe's Counter to Fake News and Cyber Attacks*, CARNEGIE ENDOWNMENT FOR INTERNATIONAL PEACE, (May 23, 2018), https://carnegieendowment.org/2018/05/23/russian-election-interference-europe-s-counter-to-fake-news-and-cyber-attacks-pub-76435. The authors of the report note that *"it is important to differentiate between "fake news" and hacking operations*."

[594] Andy Greenberg, *The NSA Confirms It: Russia Hacked French Election 'Infrastructure'*, WIRED, May 9, 2017 (12:36 PM), https://www.wired.com/2017/05/nsa-director-confirms-russia-hacked-french-election-infrastructure. Admiral Michael Rogers, then Director of the National Security Agency (NSA) testified before the Senate Armed Services Committee on May 9, 2017 that the NSA had been made aware of the hacking of Emmanuel Macron's email by Russia and had warned its French counterpart before the hacking was made public, *see* the video at *Admiral Rogers Says Intel Community Warned of Russian Hacking Ahead of Macron*

article L163-1 of the electoral Code, which provides that "*the three months preceding the first day of the month of general elections and until the date of the ballot in which they are acquired*," online platform operators, as defined by Article L. 111-7 of the consumers Code,[595] a definition which include social media platforms, if exceeding a determined number of connections on the French territory, defined by a decree as 5 million unique visitors a month within the French territory:[596]

> "*are required, because of the general interest attached to the enlightened information of citizens during an election period and to the fairness of the poll:*
>
> *1 ° To provide the user with fair, clear and transparent information on the identity of the natural person or on the company name, registered office and corporate purpose of the legal person and of the person on whose behalf, where applicable, it has declared that it is acting, which pays the platform remuneration in return for the promotion of information content relating to a debate of general interest*;
>
> 2 ° *To provide the user with fair, clear and transparent information on the use of his personal data in the context of the promotion of information content relating to a debate of general interest;*

Leak, C-SPAN, (MAY 9, 2017), https://www.c-span.org/video/?c4668917/admiral-rogers-intel-community-warned-russian-hacking-ahead-macron-leak.

[595] It defines "online platform operator" as "*any natural or legal person offering, in a professional capacity, paid or unpaid, an online communication service to the public based on: 1 ° The classification or referencing, by means of computer algorithms, of content, goods or services offered or put online by third parties; 2 ° Or the bringing together of several parties with a view to the sale of a good, the supply of a service or the exchange or sharing of content, a good or a service*."

[596] Décret n° 2019-297 du 10 avril 2019 relatif aux obligations d'information des opérateurs de plateforme en ligne assurant la promotion de contenus d'information se rattachant à un débat d'intérêt général [Decree No. 2019-297 of April 10, 2019 relating to the information obligations of online platform operators promoting information content related to a debate of general interest] [J.O.] [OFFICIAL GAZETTE OF France], April 11, 2019, n°0086.

3 ° *To make public the amount of remuneration received in return for the promotion of such information content when their amount exceeds a determined threshold*.

This information is aggregated in a register made available to the public electronically, in an open format, and regularly updated during the period mentioned in the first paragraph of this article."

Online platforms operators, including social media platforms, already have a general obligation under, Article L.111-7- II of the French Consumer Code, "*to provide consumers with fair, clear and transparent information on:*

1 ° The general conditions of use of the intermediation service that it offers and the terms of referencing, classification and de-referencing of the content, goods or services to which this service provides access;

2 ° The existence of a contractual relationship, of a capital link or of remuneration for its benefit, since they influence the classification or referencing of the content, goods or services offered or posted online;

3 ° The quality of the advertiser and the rights and obligations of the parties in civil and fiscal matter, when consumers are put in contact with professionals or non-professionals."

The constitutional Council had noted in its decision on the fake news law that:

"*the obligations imposed on online platform operators [by article L163-1 of the electoral Code] is limited to the period of the electoral campaign and only relates to those whose activity exceeds a certain threshold. It is limited to imposing on them the*

duty to provide true, clear and transparent information to the individuals for whom they have promoted, in exchange for payment, certain information content related to the electoral campaign. It seeks to provide citizens with the means to assess the value or the scope of the information thus promoted and therefore contributes to the honesty of the electoral debate. Given the objective of public interest sought as well as the limited nature of the obligation imposed on online platform operators, the contested provisions do not disproportionately infringe on the freedom of enterprise."[597]

Under the new law, platforms must provide "*clear and transparent information*" about the person or entity having sponsored the political advertising message, and to publicize online how much the platform has been paid. As only platforms which have received a "*remuneration in return for the promotion of information content relating to a debate of general interest*" are within the scope of the law, Google found a way around the law by regularly updating its *Advertising Policies* on *Political Content* to inform users that "*France Informational content ads*" will be prohibited on Google platforms during a certain period of time, which is regularly updated, following the French electoral calendar.[598]

Article 12 law entrusted the French broadcasting authority, *Conseil Supérieur de l'Audiovisuel* (CSA), to contribute to the fight against misinformation by:

"[*i*]*f necessary, … address[ing]… to the online platform operators mentioned in the first paragraph of article L. 163-1 of the electoral code, recommendations aimed at improving the fight against the dissemination of such information*." The CSA was also

[597] Decision No.2018-773 DC, December 20, 2018, paragraph 9.
[598] *Update to Political content policy (June 2020)*, GOOGLE SUPPORT, (April 2020), https://support.google.com/adspolicy/answer/9832056?hl=en, (last visited Dec. 30, 2020).

put in charge of "*monitoring of the obligation for online platform operators to take the measures provided for in Article 11 of [the law]*" and to publish "*a periodic review of their application and effectiveness. To this end, it collects from these operators, …all the information necessary for the preparation of this report*. "

The CSA published its first recommendation to the platforms in May 2019.[599] It stated that the ease of access and proper visibility of the mechanism allowing users to report false information likely to disturb public order or affect the sincerity of the election, which online platform operators must put in place under the law, is ensured:

(1) by using "*a clear title to refer to the mechanism*," such as "*report content*"

(2) by placing "*the mechanism in the immediate vicinity of the content or of the account likely to be reported*"

(3) by "*providing a reporting tool that is identical across all variants of their service*"

(4) by "*promoting exchanges between the recipients of this recommendation in order to harmonise their respective reporting mechanisms*"

(5) by "*ensuring that the mechanism is as user-friendly as possible by providing a simple and logical reporting process*"

[599] *Recommendation no. 2019-03 of 15 May 2019 of the Conseil supérieur de l'audiovisuel to online platform operators in the context of the duty to cooperate to fight the dissemination of false information,* available for download in English at *Adoption de la recommandation relative à la lutte contre la manipulation de l'information : un pas de plus vers une nouvelle regulation*, CONSEIL SUPÉRIEUR DE L'AUDIOVISUEL, (May 17, 2019), https://www.csa.fr/Informer/Espace-presse/Communiques-de-presse/Adoption-de-la-recommandation-relative-a-la-lutte-contre-la-manipulation-de-l-information-un-pas-de-plus-vers-une-nouvelle-regulation.

(6) by specifying that users should be able to report content "*by following three hyperlinks at the most*" and,

(7) by "*enabling users to monitor the processing of their reports … to follow their progress and by informing them without undue delay on the actions taken as regards the content reported*."

The CSA also addressed the issue of transparency of algorithms, as users must understand how the platforms algorithms "*select… and sequenc*[*e*]… *content work."* The CSA encouraged operators to guarantee their users:

a) "traceability of their data exploited for…recommending and ranking content;

b) clear, sufficiently detailed and easily accessible information on the criteria having led to the sequencing of content offered to the user and the classification of such criteria based on their weight in the algorithm;

c) clear and detailed information on [*user's*] *ability, if any, to change settings in order to personalise how content is referenced and recommended;*

d) clear and sufficiently detailed information on the main changes made to referencing and recommendation algorithms, and their consequences;

e) an accessible communication tool allowing for real-time communication between the user and the operator and offering the user the ability to obtain personalised and detailed information on how algorithms work."

The CSA also "*encourage*[*d*]" online platform operators to set up procedures to detect "*accounts disseminating false information on a massive scale*," such as bots, and to make "*public monitoring mechanisms and statistic techniques…identify*[*ing*] and *handl*[*ing*] *such accounts*" public. The platforms were also encouraged "*to help their users to identify which sources of information are reliable and which are not*" and directed them to "*contribute towards developing users' critical thinking, particularly in younger individuals*."

c. Other Countries

i. The U.K.

The United Kingdom does not have a law regulating fake news, but this may change in the next few years.[600] The Digital, Culture, Media and Sport Committee of the U.K. Parliament explained in its *Disinformation and 'fake news' Final Report*, published in February 2019, that it had "*disregarded the term 'fake news' as it had "taken on a variety of meanings, including a description of any statement that is not liked or agreed with by the reader" and instead recommended the terms 'misinformation' and 'disinformation."*[601] The U.K. government specified that it "*defined disinformation as the deliberate creation and sharing of false and/or manipulated information that is intended to deceive and mislead audiences, either for the purposes of causing harm, or for political, personal or financial gain. 'Misinformation' refers to the inadvertent sharing of false information*."[602]

---

[600] *See Government Responses to Disinformation on Social Media Platforms: United Kingdom*, Law Library of Congress, (July 24, 2020), https://www.loc.gov/law/help/social-media-disinformation/uk.php.
[601] Digital, Culture, Media and Sport Committee of the U.K. Parliament, *Disinformation and 'fake news' Final Report*, (Feb.2019), Paragraph 11, https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf.
[602] *Disinformation and 'fake news': Interim Report: Government Response to the Committee's Fifth Report of Session 2017–19* p.2, (Oct. 23, 2018), https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1630/1630.pdf .

ii. Brazil

Brazil harbored the second highest number of COVID-19 cases, after the U.S.,[603] and many Brazilians used the Web to gather information about the new virus, but "fake news" about the virus published online hindered the fight against the disease,[604] or even endangered the health of social media users, which led to the creation by the Brazilian Minister of Health of a page debunking COVID-19-symptoms "fake news.[605] This confusion was however sown by President Bolsonaro himself, which opposed physical distancing policies[606] and even have downplayed the severity of the pandemic.[607] A lot of false information about a potentially deadly disease, with no known cure, led to a bill presented by Senator Senador Alessandro Vieira aiming at fighting "fake news," the "*Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet."*[608] The updated bill was passed by the Senate on June 30, 2020.[609] Its article 37 mandates Article 37 of the bill mandate

---

[603] *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*, JOHN HOPKINS, https://coronavirus.jhu.edu/map.html, (last visited De. 30, 2020).
[604] Pablo Uchoa, *Brazil coronavirus: 'Our biggest problem is fake news'*, BBC, https://www.bbc.com/news/world-latin-america-52739734; Bryan Harris, *Spread of fake news adds to Brazil's pandemic crisis*, FINANCIAL TIMES, (July 13, 2020), https://www.ft.com/content/ea62950e-89c0-4b8b-b458-05c90a55b81f.
[605] *Aviso! Fique atento a fraudes e informações* falsas, Ministério da Saúde, https://www.gov.br/pt-br/temas/aviso-fraudes-e-informacoes-falsas ,(last visited Dec. 30, 2020).
[606] Lorena G. Barberia, Eduardo J. Gómez, *Political and institutional perils of Brazil's COVID-19 crisis*, THE LANCET, https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)31681-0/fulltext.
[607] Katy Watson, *Coronavirus: Brazil's Bolsonaro in denial and out on a limb*, BBC News, (March 29, 2020), https://www.bbc.com/news/world-latin-america-52080830.
[608] Projeto de Lei, PL 2630/2020, https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2256735.
[609] An updated version and comment of the bill is available (in Portuguese) at https://legis.senado.leg.br/sdleg-getter/documento?dm=8127649&ts=1593563111041&disposition=inline (last visited Dec. 30, 2020).

private messaging apps and social media platforms to appoint legal representatives in Brazil having the power to remotely access users 'database.[610]

B. Denying Crimes Against Humanity

"Is it true?"

"It's history, Craig-Vyvian."

"Is history true?"

"More or less."[611]

In a Report presented to the United Nations in 2012, the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression wrote that:

> "*historical events should be open to discussion and…laws that penalize the expression of opinions about historical facts are incompatible with the obligations that the International Covenant on Civil and Political Rights imposes on States parties in relation to the respect for freedom of opinion and expression… By demanding that writers, journalists and citizens give only a version of events that is approved by the Government, States are enabled to subjugate freedom of expression to official versions of events*."[612]

---

[610] Udbhav Tiwari and Jochai Ben-Avie, *Mozilla's analysis: Brazil's fake news law harms privacy, security, and free expression*, MOZILLA, (June 29, 2020), https://blog.mozilla.org/netpolicy/2020/06/29/brazils-fake-news-law-harms-privacy-security-and-free-expression.

[611] Mark Helprin, FREDDY AND FREDERICKA, Penguin Books (2005). p. 13. The book is a fictionalized account of the misadventures of the Prince of Wales and his spouse. Some elements are similar to aspects of the real British Royal family (the Queen's husband is titled Duke, the Queen loves dogs…), while other aspects are fictionalized, such as the reliance on Merlin the Magician to provide advice in case of dire crisis.

[612] *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, A/67/357, paragraph 55, available at https://undocs.org/A/67/357.

As such, social media users, whether they are journalists or simple citizens may post messages presenting their own version of a particular event. If one of the purposes of the First Amendment is to allow people to seek truth in the marketplace of ideas, shouldn't it bar fake news, lies, and uncorroborated views of history?

John Stuart Mill wrote in *On Liberty* about the dangers how imposing one's view as infallible:

> "*… the opinion which it is attempted to suppress by authority may possibly be true. Those who desire to suppress it, of course deny its truth; but they are not infallible. They have no authority to decide the question for all mankind, and exclude every other person from the means of judging. To refuse a hearing to an opinion, because they are sure that it is false, is to assume that their certainty is the same thing as absolute certainty. All silencing of discussion is an assumption of infallibility. Its condemnation may be allowed to rest on this common argument, not the worse for being common*."[613]

However, some views must be accepted by a Democratic society. Knowing that the Shoah has occurred and being aware of its horrors is one of them, which is on the reasons the U.S. army took care to film the concentration camps it had liberated. Is believing that the Holocaust or the Armenian genocide never took place an opinion worthy of being protected?

[613] JOHN STUART MILLS, ON LIBERTY, (The Walter Scott Publishing Co., Ltd), (Projet Gutenberg eBook #34901, 2011), 31, available at https://www.gutenberg.org/files/34901/34901-h/34901-h.htm#, page 28.

a. Denying the Holocaust

The European Council's *Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems* was adopted in Strasbourg on January 28, 2003, and entered into force on January 3, 2006. Its parties both recognize:

> "*that freedom of expression constitutes one of the essential foundations of a democratic society*" and are concerned "*by the risk of misuse or abuse of … computer systems to disseminate racist and xenophobic propaganda*." Its article 6 directs each of its parties to "*adopt… legislative measures* [*which*] *may be necessary to* [*criminalize*]… *when committed intentionally and without right: distributing or otherwise making available, through a computer system to the public, material which denies, grossly minimizes, approves or justifies acts constituting genocide or crimes against humanity, as defined by international law and recognised as such by final and binding decisions of the International Military Tribunal, established by the London Agreement of 8 August 1945, or of any other international court established by relevant international instruments and whose jurisdiction is recognised by that Party*."

Therefore, it is a crime, under the Additional Protocol, to deny, not only the Holocaust, but also genocides or crime against humanity recognized as such by another court than the Nuremberg Court. Such Court could be, for instance, the International Criminal Court (ICC). Article 6 of the Rome Statute of the International Criminal Court, which created the ICC and came into force on July 1, 2002, defines genocide as " *any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical,*

racial or religious group, as such: (a) Killing members of the group; (b) Causing serious bodily or mental harm to members of the group; (c) Deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part; (d) Imposing measures intended to prevent births within the group; (e) Forcibly transferring children of the group to another group."

Article 7.1 of the Rome Statute defines crimes against humanity in a complex and broadly inclusive. They are "*any of the following acts when committed as part of a widespread or systematic attack directed against any civilian population, with knowledge of the attack:*

(a) Murder;

(b) Extermination;

(c) Enslavement;

(d) Deportation or forcible transfer of population;

(e) Imprisonment or other severe deprivation of physical liberty in violation of fundamental rules of international law;

(f) Torture;

(g) Rape, sexual slavery, enforced prostitution, forced pregnancy, enforced sterilization, or any other form of sexual violence of comparable gravity;

(h) Persecution against any identifiable group or collectivity on political, racial, national, ethnic, cultural, religious, gender as defined in paragraph 3, or other grounds that are

universally recognized as impermissible under international law, in connection with any act referred to in this paragraph or any crime within the jurisdiction of the Court;

(i) Enforced disappearance of persons;

(j) The crime of apartheid;

(k) Other inhumane acts of a similar character intentionally causing great suffering, or serious injury to body or to mental or physical health."

Denying the Holocaust is also known as "negationism." It is rooted in Anti-Semitism, as it is obvious when studying the French law making it a crime to deny the existence of the Holocaust. Anti-Semitism is on the rise in Europe,[614] and thus laws making speech denying the Holocaust illegal are enacted to protect historical truth but also to fight anti-Semitism. However, it can be argued that denying the Holocaust is not racial defamation, but an opinion crime: it is a crime to have a specific opinion, in this case, believing that the Holocaust did not take place. If negationism is an opinion, then stating that the Holocaust never took place is protected speech. If it is not protected speech, then it is not an opinion, but rather, the denial of a proven historical facts, a false statement potentially dangerous for the stability of democracy. Is negationism simply an opinion, one of the many in the "marketplace of ideas"?

Nobel prize winner and Holocaust survivor Elie Wiesel wrote that "*the Holocaust transcends history.*"[615] The ECtHR described the Holocaust, in its *Lehideux and Isorni v.*

[614] Patrick Kingsley, *Anti-Semitism Is Back, From the Left, Right and Islamist Extremes. Why?*, THE NEW YORK TIMES, (April 4, 2019), https://www.nytimes.com/2019/04/04/world/europe/antisemitism-europe-united-states.html?searchResultPosition=2.
[615] ELIE WIESEL, AGAISNT SILENCE. THE VOICE AND VISION OF ELIE WIESEL, 1995, (Holocaust Library Ed.)

France decision, as belonging to "*the category of clearly established historical facts*."[616] The General Assembly of the United Nations adopted on January 26, 2007 Resolution 61/255 on Holocaust denial condemning any denial of the Holocaust.[617] However, the Resolution was adopted by consensus, as only 103 Members States voted in favor of it out of their total number of one hundred and ninety-three. It "*condemns without any reservation any denial of the Holocaust* [*and*]… [*u*]*rges all Member States unreservedly to reject any denial of the Holocaust as a historical event, either in full or in part, or any activities to this end.* "U.N. resolutions are binding for the Member States but are not enforceable.

The European Council issued a Joint Action on July 15, 1996, which paragraph (A)(c) urged each Member State to "*take steps to see that* [*public denial of the crimes defined in Article 6 of the Charter of the International Military Tribunal*] *is punishable as a* ***criminal offense***"[618] (our emphasis). Several European countries have passed laws making negationism a crime.[619] In France, the *loi Gayssot*, of July 13, 1990 added an article 24 bis to the French Press Law and created a new offense, denying the existence of a crime against humanity, as defined by article 6 of the 1945 Statute of the Nuremberg International Military Tribunal (Nuremberg Statute) annexed in the London Accord of August 8, 1945.[620]

---

[616]Lehideux and Isorni v. France (Grand Chamber), Sept. 23, 1998, at 47.
[617] G.A. Res. 44, U.N. Doc. A/RES/61/255 (Jan. 26, 2007).
[618] Council Joint Action 96/443/JHA Concerning Action to Combat Racism and Xenophobia, Title I, § (A) (c), 1996 O.J. (L 185) 5, available at https://op.europa.eu/en/publication-detail/-/publication/4c2c7677-7bbf-4fc9-8af4-7e5e00d921b6/language-en.
[619] *See* Emanuela Fronza, *The Punishment of Negationism: The Difficult Dialogue between Law and Memory*, 30 Vt. L. Rev. 609. 616-617 (2006).The author cites, in particular, Section 130, chapter 3 of the German Penal Code (*Stratgesetzbuch*) and the Austrian Law of February 26, 1992 modifying Section 3 of the constitutional law on national-socialists (*Bundesverfassungsgesetz vom Februar 1947 über die Behandlung der Nazionlasozialisten*).
[620] Loi n°90-616 du 13 juillet 1990 Loi n° 90-615 du 13 juillet 1990 tendant à réprimer tout acte raciste, antisémite ou xénophobe [Law 90-616 of July 13, 1990 to suppress any racist, anti-Semitic or xenophobic act], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF France], July 14, 1990, p. 8333.

Therefore, the scope of the law is defined by the Nuremberg Statute and only criminalizes "*murder, extermination, enslavement, deportation, and other inhumane acts committed against civilian populations, before or during the war; or persecutions on political, racial or religious grounds in execution of or in connection with any crime within the jurisdiction of the Tribunal, whether or not in violation of the domestic law of the country where perpetrated*." The scope of the law is therefore restricted to crime committed by the European Axis countries, and could not be invoked, for example, if a person would be denying the Rwandan genocide.

The law was enacted at a time French historian Robert Faurisson was regularly featured in the press denying the existence of the Holocaust. The profanation in May 1990 of a Jewish cemetery in Carpentras,[621] in the South of France, had shocked the French opinion.[622] Protests were organized in several French cities to raise awareness on anti-Semitism, and François Mitterrand, then President of France, even joined the one organized in Paris, mingling with the crowd in a much-publicized event in the history of the French Republic.[623] The authors of this horrific act of hate were tried in 1997, following the spontaneous recognition of fact by one of the profaners in 1996, who had also designated his accomplices to the police. The *loi Gayssot* was enacted quickly after the profanation.

Defendants tried under the law regularly argued in defense that the law violated article 10 of the European Declaration of Human Rights, but the *Cour de cassation* held as

---

[621] It is the oldest Jewish cemetery in use in Europe, see *The Carpentras Affair*, THE NEW YORKER, (Oct. 30, 2000), https://www.newyorker.com/magazine/2000/11/06/the-carpentras-affair.
[622] Indeed, the desecration had been gruesome: several tombs had been opened, and the body of a recently deceased man had been dragged outside and tampered with a parasol pole.
[623] *Mitterrand Joins Marchers Protesting Anti-Semitism*, LOS ANGELES TIMES, (MAY 14, 1990 12AM), https://www.latimes.com/archives/la-xpm-1990-05-14-mn-261-story.html.

regularly that it did not.[624] The constitutionality of the law which had introduced article 24 bis to the French Press Law had not been presented to the constitutional Council before being enacted. At the time, the Council could only examine the constitutionality of a law immediately after it had been voted but prior to its promulgation, and only if sixty representatives or sixty senators presented a request to the Council for such review. Passed that window of opportunity, the constitutionality of a law, once enacted, could no longer be challenged. This changed after the Constitution was amended on July 23, 2008 by a Constitutional law which introduced a new procedure to challenge the constitutionality of laws, even after their enactment, the *question prioritaire de constitutionalité* (QPC). A party at a lawsuit can challenge the constitutionality of a law, by first asking the *Conseil d'État*, the Council of State, which is France's highest administrative court, or the *Cour de cassation*, which is France's highest civil and criminal court, to refer the case to the constitutional Council. The highest courts may decide whether such review is warranted.

The QPC became operative on March 1, 2010. Almost immediately, the constitutionality of the *loi Gayssot* was challenged by an extreme right newspaper, which

[624] Cour de cassation {Cass.] [Supreme court for judicial matters], crim., Feb. 23, 1993, no 92-83.478, Bull. crim. 86. The Court found that the *loi Gayssot* incriminated *"behavior that is detrimental to public order and the rights of individuals; [and] that, therefore, does not exceed the limits fixed by the second paragraph of Article 10 [of the European Convention on Human Rights*." *See* also Cour de cassation {Cass.] [Supreme court for judicial matters], crim., Sept. 12, 2000, No 98-88.200. In this case, author Roger Garaudy had denied the Holocaust, writing the Jewish people had merely been deported to concentration camps, not exterminated. A prolific author, he had left a rich trail of denying the Holocaust, which the Court of appeals had duly analyzed to reach the conclusion that indeed article 24bis of the French Press Law was applicable to the facts. The *Cour de cassation* approved, reasoning that "*Article 10 of the European Convention on Human Rights, which guarantees the principle of freedom of expression, includes in paragraph 2 certain limitations or sanctions, provided for by law, which constitute necessary measures in a democratic society, for the defense of order and the protection of the rights of others; that such is the object of article 24 bis of the law of July 29, 1881;… the contestation of the existence of crimes against humanity enters into the provisions of article 24 bis of the law of July 29, 1881, even if it is presented in disguised or dubious form or presented by innuendo; that it is also characterized when, under the guise of searching for a supposed historical truth, it tends to deny the crimes against humanity committed by the Nazis against the Jewish community*."

had been found guilty of denying crimes against humanity for having published an interview of Jean-Marie Le Pen, then President of the Front National, an extreme right party, where he declared that German occupation had not been "*particularly inhuman*." However, the *Cour de cassation* refused to refer the QPC to the Council, reasoning that:

> "*the question is not of a serious nature since the crime thus challenged refers to laws which have been regularly introduced into domestic law, and which define clearly and precisely the crime of denying one or several crimes against humanity as defined by Article 6 of the Charter of the International Military Tribunal annexed to the London Agreement of August 8, 1945 and which were committed either by members of a declared criminal organization pursuant to Article 9 of that statute, or by a person convicted of such crimes by a French or international court, offense of which the punishment therefore does not infringe the constitutional principles of freedom of expression and opinion*."[625]

However, in October 2015, the *Cour de cassation* asked the Constitutional Council to review the constitutionality of article 24 bis of the French Press law.[626] The Court asked the Council to answer these questions:

> "*Does article 24 bis of the law of July 29, 1881 infringe the rights and freedoms guaranteed by the French Constitution of October 4, 1958 and in particular*:

[625] Cour de cassation [Cass.] [Supreme court for judicial matters], May 7, 2010, n° 12008 (09-80.774), https://www.courdecassation.fr/jurisprudence_2/questions_prioritaires_constitutionnalite_3396/arrets_3398/12008_7_16224.html.
[626] Cour de cassation [Cass.] [Supreme court for judicial matters], crim., October 6, 2015, n° 4632, available at https://www.courdecassation.fr/jurisprudence_2/qpc_3396/4632_6_32730.html.

- the principle of equality before the law guaranteed by article 6 of the Declaration of the Rights of Man and of the Citizen of 1789 (which requires that the law be the same for all) and by article 1 of the Constitution of October 4, 1958 (which imposes equality before the law without distinction of origin, race or religion);

- the freedom of opinion guaranteed by article 10 of the Declaration of the Rights of Man and of the Citizen of 1789 (which allows its manifestation within the strict limit of disturbing public order);

- the freedom of expression guaranteed by article 11 of the Declaration of the Rights of Man and of the Citizen of 1789 (considered, unless abused, as consubstantial with democracy and the rule of law)? "

The Council found the law to be constitutional,[627] reasoning that article 24bis did not violate article 11 of the French declaration of Human Rights protecting freedom of expression because it only prohibited…

"*the negation, implicit or explicit, or the extreme exaggeration of* [*facts committed during the Second World War, qualified as crimes against humanity and sanctioned as such by a French or international jurisdiction*]… ***that the contested provisions have neither the object nor the effect of prohibiting historical debates****; that thus the infringement of the exercise of the freedom of expression which results from it is*

[627] Decision n° 2015-512 QPC , January 8 2016, https://www.conseil-constitutionnel.fr/decision/2016/2015512QPC.htm.

necessary, adapted and proportionate with the objective pursued by the legislator" (our emphasis).

In 2003, French writer Roger Garaudy was unsuccessful in his attempt to have the ECtHR declare that article 24 bis of the French Press Law violated article 10 of the ECHR. Garaudy had published in 1995 a book, *The Founding Myths of Israeli Politics,* where he denied the existence of gas chambers in Nazi concentration camps, the number of Jews killed by the Nazis, and also denied that the 'final solution' was the complete extermination of the Jews, arguing instead that it merely referred to the exile of all Jews from Europe. Garaudy self-published the same book in 1996. He was prosecuted[628] in five different cases, for five violations of the French Press Law, denying crimes against humanity, publishing racially defamatory statements, and inciting to racial or religious hatred or violence, and sentenced, for having denying crimes against humanity under article 24 bis of the French Press Law, to a suspended 6-months prison term and to a 50,000 French franc fine.[629] The civil parties were symbolically awarded one franc in damage. Garaudy appealed the five cases to the *Cour de cassation*, arguing that article 24 bis of the French Press Law did not fall within the exceptions to freedom of speech provided by Article 10 §2 of the ECHR, that his book was not racist and did not deny the existence of the Holocaust. The *Cour de cassation* confirmed the five cases in five different opinions. The Court reasoned that article 10 §2 of the ECHR indeed allows to restrict speech, if it is a necessary measure in a

[628] Garaudy tried to join the five proceedings joined but his application was rejected because even if they concerned the same author, because the proceedings were about two different editions of the same book and the five separate cases had been commenced by different civil parties, associations of former Resistance members, deportees, and human rights organizations, and by the Paris District Attorney (for denying crimes against humanity), who had each cited different passages of the book.
[629] More or less 10,000 USD at the time.

democratic society for the prevention of disorder and the protection of rights of others, and that this is also the purpose of Article 24 bis. The Court also held that article 24 bis incriminates denying the existence of crime against humanity, "*even if it is presented in a disguised or doubtful form or by insinuating such view; it is also characterized when, under the guise of looking for a supposedly historical truth, it tends to deny the crimes against humanity committed by the Nazis against the Jewish community,*" as Garaudy's book had attempted to do.[630] Garaudy then took his case to the European Court of Human Rights, which unanimously declared the application inadmissible.[631] Garaudy had claimed that article 24 bis was not one of the measures prescribed by law which are necessary in a democratic society, within the meaning of article 10(2) of the ECHR, and that article 24 bis "*created a censorship mechanism that wrongfully restricted freedom of expression*." Article 17 of the ECHR prevents however using any provisions of the Convention to "[*destroy*] … *any of the rights and freedoms set forth* [*by the Convention*]." The purpose of article 17 "*is to prevent the principles laid down by the Convention from being exploited for the purpose of engaging in any activity or performing any act aimed at the destruction of the rights and freedoms set forth in the Convention*,"[632] thus preventing people claiming extremist opinions repugnant to the principles protected by the ECHR from using the Convention to protect them from liability. In other words, article 17 prevents the Convention to be used as a sword and as a shield.

---

[630] Cour de cassation [Cass.] [supreme court for judicial matters] crim., Sept. 12, 2000, N° 98-88200, not published. Available at https://www.legifrance.gouv.fr/affichJuriJudi.do?oldAction=rechJuriJudi&idTexte=JURITEXT000007584921&fastReqId=157135238&fastPos=1.

[631] Garaudy v. France, 2003-IX 369, available at https://echr.coe.int/Documents/Reports_Recueil_2003-IX.pdf.

[632] Zdanoka v. Latvia, Application No. 58278/00) First Section (2004), at 109.

For the Court:

> "[*t*]*here can be no doubt that denying the reality of clearly established historical facts, such as the Holocaust, as the applicant does in his book, does not constitute historical research akin to a quest for the truth. The aim and the result of that approach are completely different, the real purpose being to rehabilitate the National-Socialist regime and, as a consequence, accuse the victims themselves of falsifying history. Denying crimes against humanity is therefore one of the most serious forms of racial defamation of Jews and of incitement to hatred of them. The denial or rewriting of this type of historical fact undermines the values on which the fight against racism and anti-Semitism are based and constitutes a serious threat to public order. Such acts are incompatible with democracy and human rights because they infringe the rights of others. Their proponents indisputably have designs that fall into the category of aims prohibited by Article 17 of the Convention*."

The Court further reasoned that "*the main content and general tenor*" of Garaudy's book were revisionist and thus against "*the fundamental values of the Convention, as expressed in its Preamble, namely justice and peace*." For the Court, Garaudy "*attempt*[*ed*] *to deflect Article 10 of the Convention from its real purpose by using his right to freedom of expression for ends which are contrary to the text and spirit of the Convention. Such ends, if admitted, would contribute to the destruction of the rights and freedoms guaranteed by the Convention*." Therefore, he could not rely on the protection of article 10 of free speech to deny crimes against humanity.

Social media platforms are regularly used to publish and promote anti-Semitic views, and even denying the Holocaust. For instance, in April 2016, Henry de Lesquen, a French politician, and candidate in 2017 to the French Presidential elections, published on his Twitter account in April 2016 two messages which led to his sentence of 6,000 euros fine for having contested crimes against humanity. One tweet read: "*I am amazed at the longevity of the" survivors of the Shoah "who died over 90 years old. Did they experience the horrors they told?*" and another one read: "*The busty Simone Veil*[633] *"survivor of the Shoah", is 88 years old. To my knowledge, she is fine.*" The judgment was confirmed by the Paris Court of appeals, which had found that these two tweets had called into questions the testimonies of survivors of the Shoah, and that the tweet about Mrs. Simone Veil had minimizing, by innuendo, what people deported to the camps had suffered through, the conditions of the camps, and, as such had outrageously reduced the actual number of the people deported in concentrations camps and downgraded their suffering. The tweets thus constituted the offense of contesting crime, against humanity. Lesquen argued at the Supreme Court that, as criminal laws must be strictly interpreted by the judges, article 24 bis of the French Law Press could not be applied to the facts, as article 24 bis directly refers to article 6 of the statute of the international court of Nuremberg, and that commenting on the consequences of the deportation for some individuals are outside of article 6's scope and thus outside of Article 24 bis' scope. Lesquen argued further that his tweets did not cast doubt upon these crimes. The *Cour de cassation* was not convinced by these arguments

[633] Simone Veil was a former French health minister and former President of the European Parliament. She died in June 2017 and was buried in the French Parthenon a year later, after several petitions asked the government to quickly bestow her this honor, *See* Kim Willsher, *France pays tribute to Simone Veil with hero's burial in the Panthéon*, THE GUARDIAN, (Sat 30 Jun 2018 11.33 EDT), https://www.theguardian.com/world/2018/jun/30/simone-veil-funeral-paris-pantheon.

and approved the Court of appeals for having placed their tweets into their context: Lesquen had published them after having tweeted about a particular individual, who had died at the age of 94, who had allegedly lied about having been deported. Lesquen had placed Mrs. Simone Veil's statute as Shoah survivor between quotes, thus calling into question the veracity of her experience, which had led the Court of appeals to conclude that minimizing the sufferings of the Shoah victims was indeed challenging crimes against humanity. Denying the existence of crimes against humanity, as criminalized by article 24 bis, is characterized "*even if it is presented in disguised or dubious form or even by innuendo*."[634]

A year earlier, on December 12, 2018, the Paris Court of appeals had confirmed a judgment of the lower criminal court against Lesquen for two other tweets, these about the 1942 *Vel' d'Hiv* Roundup, where the French police arrested over 13,000 Jews, including children, following the order of the German authorities, and rounded them up in Paris "Winter Velodrome" for several days, before deporting them to concentration camps.[635] Lesquen had posted two tweets in April 2017, one stating that "*To be obsessed with #veldhiv, you have to have a small bicycle in your head. Minor episode of deportation*" and another one stating that *"#veldhiv is a minor episode of deportation, which is itself a minor episode of the second world war*." The criminal court had found Lesquen guilty of denying crimes against humanity and sentenced him to 3,000 euros fine. The judgment was

[634] Cour de cassation [Cass.] [supreme court for judicial matters] crim., Nov. 13, 2019, not published on the official reporter, but available at https://www.legifrance.gouv.fr/affichJuriJudi.do?oldAction=rechJuriJudi&idTexte=JURITEXT000039418990&fastReqId=1207006220&fastPos=1.

[635] *The Vel' d'Hiv Roundup*, YAD VASHEM, https://www.yadvashem.org/holocaust/france/vel-dhiv-roundup.html.

confirmed by the Paris Court of appeals, noting that it was the Nazis who had ordered the roundup and were thus members of an organization which had been found to be a criminal one under article 9 of the statute of the international court of Nuremberg, and thus within the scope of Article 24 bis.[636] The Court of appeals noted that members of the *Schutzstaffel* (SS) had participated in meetings and interviews which Émile Hennequi, then director of the municipal police of Paris, which had taken place in Paris in the first days of July 1942, and where the scope and methods of the roundup had been decided. These SS officers had also drafted notes and reports, citing a memorandum written by Hennequin, dated July 13, 1942, instructing to arrest and regroup Jews and referring to the decision to round up French Jews taken by the occupying authorities. The Court of appeals concluded that writing that the *Vel' d'Hiv* Roundup was "*a minor episode of the deportatio*n" was an outrageous reduction of crimes against humanity.

Lesquen had used the same defense as in the case about the tweets on Shoah survivors', claiming that the tweets were outside the scope of article 24 bis of the French Press Law. The *Cour de cassation* approved however the Court of appeals to have deducted from historical facts that the round up had been decided and planned by the Nazis while the deed was carried on by the French police, with the participation of the Vichy government. The SS were members of an organization declared criminal under article 9 of the statute of the international court of Nuremberg, and thus the *Vel' d'Hiv* Roundup is withing the scope of article 24 bis. For the *Cour de cassation*, article 24 bis of the French Press Law:

[636] Cour d'appel [CA] [regional court of appeal] Paris, 7e ch. Dec. 12, 2018, Légipresse n°367, 12.

" does not require that the crimes against the humanity to have been exclusively committed either by the members of an organization declared criminal in application of article 9 of [*the statute of the international court of Nuremberg*], *either by a person convicted of such crimes by a French or international court, but it is enough that the persons thus designated have decided or organized them, regardless of whether their material execution was, partially or completely, the fact of a third party*."

This legal reasoning, which must be approved, illustrates how France is presented as one of the victorious countries of WWII, one of the four major Allied powers, with the United States, the United Kingdom and the Soviet Union, which had set up the Nuremberg International Military Tribunal, and thus acting as one of the prosecutors, yet is also a country which government, under the direction of Marshal Pétain, had actively supported the Nazi regime.[637] This inglorious past is slowly becoming widely known and accepted by the French, especially since, as noted by Professor Teachout, the public has more and more

[637] This position quickly became the official history of France. For instance, François de Menthon, France's Chief Prosecutor of the Nuremberg trials, said in France's opening statement: "*France, invaded twice in 30 years in the course of wars, both of which were launched by German imperialism, bore almost alone in May and June 1940 the weight of armaments accumulated by Nazi Germany over a period of years in a spirit of aggression…; France, which was subjected to the still more horrible grip of demoralization and return to barbarism diabolically imposed by Nazi Germany, asks you, above all in the name of the heroic martyrs of the Resistance, who are among the greatest heroes of our national legend, that justice be done. France, so often in history the spokesman and the champion of human liberty, of human values, of human progress, through my voice today also becomes the interpreter of the martyred peoples*." *See* Nuremberg Trial Proceedings Vol. 5, 36th Day, Thursday, 17 January 1946, Morning Session, THE AVALON PROJECT, YALE LAW SCHOOL, https://avalon.law.yale.edu/imt/01-17-46.asp. There have been two diverging paths taken by France since June 17, 1940, when Marshal Pétain announced that France will cease fire. On June 18, 1940, *Général* Charles de Gaulle, speaking on the BBC radio, urged all French to join him and continue to fight, the famous *Appel du 18 juin 1940*. Marshal Pétain signed the armistice with Germany and Italy on June 22, 1940, and his government, located in the town of Vichy, in the center of France, actively collaborated with the Nazi occupiers during the war. At the same time, the French *Résistance* actively fought the Nazis and de Gaulle was head of the Free French Movement, a *de facto* government. He issued on October 27, 1940, an Ordinance which created the Council of Defense of the Empire (France, while a Republic, was also at the time a vast colonial "empire"), which article 1 stated that the power would be exercised on the basis of laws predating June 23, 1940, until the restoration of freely elected representative institutions, *see* Benjamin R. Payn, *French Legislation in Exile*, 28 J. Comp. Legis. & Int'l L. 3d ser. 44, 45 (1946).

access to source of information, including movies, such as Stephen Spielberg 1993 *Schindler's List* movie or Marcel Ophüls' 1969 documentary *Le Chagrin et la Pitié* (*The Sorrow and the Pity*).[638] The Ophüls' movie, released 25 years after the Liberation of France, forced the French to confront the collaboration of their country with the occupying forces. Even though the documentary, a command from a French public television channel,[639] was produced in 1969, the network refused to broadcast it as it showed, without any doubt, that many French had collaborated during WWII, including the active participation of the French police to the *Vel' d'Hiv* Roundup.[640] The movie was finally released in French movies theaters in 1971[641] and was screened for the first time on television only in 1981. The travail of France's collective memory shows the importance of the marketplace of ideas for assessing and accepting an historical truth as such.

Article 24 of the French Press incriminates apology of war crimes and apology of crimes against humanity. The French *Cour de cassation* held in 2018 that apology of war crimes and apology of crimes against humanity, which are both criminalized by article 24 §5 of the French Press Law, are separate offenses.[642] In this case, Alain Soral, infamous in France for its racist and anti-Semitic views, had commented on Facebook about the medal given to Serge and Beate Klarsfeld, the couple who has dedicated their lives to bring Nazis war criminal to justice, in such terms:[643] "*That's what happens when you do not finish the*

---

[638] Peter R. Teachout, *Making Holocaust Denial a Crime: Reflections on European Anti-Negationist Laws from the Perspective of U.S. Constitutional Experience*, 30 Vt.L. Rev. 655, 691, note 156 (2006).
[639] There were only two television channels in France at the time, and both public channels.
[640] Stuart Jeffries, *A Nation Shamed*, THE GUARDIAN (Jan. 23, 2004 22.05 EST), https://www.theguardian.com/film/2004/jan/23/1,
[641] Jamie Russell, The Sorrow And The Pity (Le Chagrin Et La Pitié) (1971), BBC MOVIES, (May 14, 2004), http://www.bbc.co.uk/films/2004/05/14/sorrow_and_pity_1971_review.shtml.
[642] Cass crim. n° 17-82.656, May 7, 2018.
[643] Evelyne Pieiller, *The online politics of Alain Soral*, LE MONDE DIPLOMATIQUE (Nov. 2013), https://mondediplo.com/2013/11/11rightwing.

job!" He was indicted for racist insult and apology of war crimes and crimes against humanity and found guilty, because Soral had thus "*presented the genocidal enterprise of the Nazi regime in a favorable light, as a legitimate action which one must wish to be accomplished.*"[644] The *Cour de cassation* confirmed that this constitutes an apology of crimes against humanity under the French Press Law. The Court of appeals had found Soral also to be guilty of apology of war crimes. However, the *Cour de cassation* reversed, because the Court of appeals had not specified which constituent elements it had retained to find Soral guilty of apology of war crimes.

The European Court of Human Rights held in 1998 in *Lehideux and Isorni v. France* that the French law criminalizing the apology of collaboration with the enemy and war crimes was not a violation of Article 10 of the Convention. In this case, the two applicants had published in July 1984 an ad in the French newspaper *Le Monde* arguing that the French people had forgotten the good deeds of Marshal Pétain, the leader of the Vichy régime during WWII. He had been tried after the war and sentenced to death for collaboration, but his sentence had been commuted to solitary confinement for life by Charles de Gaulle. Pétain died in 1951.[645] The ECtHR found article 24 of the French Press Law to be an "*interference*" with the applicants' exercise of their right to freedom of expression, However, the interference was necessary to protect the reputation of others and to prevent disorders or crimes. The Court started its reasoning by noting that the role played by Pétain during WWII:

[644] Paris Court of Appeals, Chamber 2-7, March 2, 2017, cited by the *Cour de cassation*.
[645] *Philippe Pétain (1856 - 1951*), BBC HISTORY, http://www.bbc.co.uk/history/historic_figures/petain_philippe.shtml (last visited Dec. 30, 2020).

"*is part of an ongoing debate among historians about the events in question and their interpretation. As such, it does not belong to the category of clearly established historical facts – such as the Holocaust – whose negation or revision would be removed from the protection of Article 10 by Article 17*."

Indeed, the applicants had not denied the Holocaust and had referred in the ad to "*Nazi atrocities and persecutions*" and to "*German omnipotence and barbarism*." Their aim had instead been to argue that Pétain's policy had been "*supremely skillful*" and to support "*the so-called "double game" theory*" about Pétain's role, a theory, however, which had been recused by historians at the time of the publication of the ad. The ECtHR noted that the applicants did not act in their personal capacities, but as presidents of two legally constituted non-profit organizations seeking to promote Pétains' rehabilitation and that the word "Advertisement" appeared at the top of the newspaper page. The Court of appeals had reasoned that the applicants, by omitting essential historical facts, had made the apology of war crimes, holding that "*by putting forward an unqualified and unrestricted eulogy of the policy of collaboration* [*the applicants] were ipso facto justifying the crimes committed in furtherance of that policy*". The Court of Cassation approved this reasoning.

But the European Court of Human Rights, while acknowledging that the ad was undoubtedly "*polemical*", reiterated that "*Article 10 protects not only the substance of the ideas and information expressed but also the form in which they are conveyed."* The applicants had "*explicitly stated their disapproval of "Nazi atrocities and persecutions" and of "German omnipotence and barbarism*" and their goal had been the overturning of Pétain's conviction. The Court found no violation of Article 10. It further noted that:

"… the events referred to in the publication in issue had occurred more than forty years before. Even though remarks like those the applicants made are always likely to reopen the controversy and bring back memories of past sufferings, the lapse of time makes it inappropriate to deal with such remarks, forty years on, with the same severity as ten or twenty years previously. That forms part of the efforts that every country must make to debate its own history openly and dispassionately. The Court reiterates in that connection that, subject to paragraph 2 of Article 10, freedom of expression 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; such are the demands of that pluralism, tolerance and broadmindedness without which there is no "democratic society" "

The issue of whether it is advisable that a certain historical view should be imposed by law is still being debated. Representative Jean-Claude Gayssot, who had introduced and defended the bill which became the July 13, 1990 law adding article 24 bis to the French Press Law, had declared during the debates that "*racism is not an opinion, it is a crime*," a quote often used since when arguing that hate speech must be illegal. Denying the Holocaust is a factual transgression from facts that French Courts accept its existence as the truth in defense to a defamation suit. In one case, a French journalist had called Faurisson a "*forger of history*" and he had sued the journalist for defamation. The Paris Court of appeal admitted the exception of truth defense, which had been based in this case on testimonies

of historians, who had stated that Faurisson's intellectual approach was not the one usually followed by historians.[646]

In Germany, a Member of Parliament, and the chairperson of the National Democratic Party of Germany (NPD) in the Land Parliament of Mecklenburg-Western Pomerania, made a speech in the Land Parliament a day after January 27, 2010, Holocaust Remembrance Day, which had been marked by a ceremony at the Parliament. The politician talked about the "*so-called Holocaust*" (*sogenannte Holocaust*) and referred to the Holocaust commemoration ceremony which had been organized the day before as "*nothing more than you imposing your Auschwitz projections onto the German people in a manner that is both cunning and brutal. You are hoping… for the triumph of lies over truth*." He was sentenced for violating the memory of the dead and for defamation of Jewish people. He appealed to the ECtHR, claiming that his sentence violated article 10 of the ECHR. The Court held in October 2019, in *Pastörs v. Germany*, that the sentence, an interference to the politician freedom of speech, was prescribed by law, pursued the legitimate aim of protecting the reputation and rights of others, and was necessary in a democratic society.[647] The politician's message his message had "*show*[*ed*] *disdain towards the victims of the Holocaust and runn*[*ed*] *counter to established historical facts, alleging that the representatives of the "so-called" democratic parties were using the Holocaust to suppress and exploit Germany*."[648]

[646] Cour d'appel [CA] [regional court of appeal] Paris, 7th ch. April 12, 2018, n° 17/04449. LEGIPRESSE, 2019, 175.
[647] Pastörs v. Germany, req. n° 55225/14, (Oct. 3. 2019
[648] Paragraph 46.

For the Court:

"*the applicant intentionally stated untruths in order to defame the Jews and the persecution that they had suffered during the Second World War. Reiterating that it has always been sensitive to the historical context of the High Contracting Party concerned when reviewing whether there exists a pressing social need for interference with rights under the Convention and that, in the light of their historical role and experience, States that have experienced the Nazi horrors may be regarded as having a special moral responsibility to distance themselves from the mass atrocities perpetrated by the Nazis…, the Court therefore considers that the applicant's impugned statements* ***affected the dignity of the Jews to the point that they justified a criminal-law response****. Even though the applicant's sentence of eight months' imprisonment, suspended on probation, was not insignificant, the Court considers that the domestic authorities adduced relevant and sufficient reasons and did not overstep their margin of appreciation. The interference was therefore proportionate to the legitimate aim pursued and was thus necessary in a democratic society*"[649] (my emphasis)

The Court explained that article17 "*is only applicable on an exceptional basis and in extreme cases and should, in cases concerning Article 10 of the Convention, only be resorted to if it is immediately clear that the impugned statements sought to deflect this Article from its real purpose by employing the right to freedom of expression for ends clearly contrary to the values of the Convention*" and that statements are not protected by article 10 if they are

[649] Paragraph 48.

"*directed against the Convention's underlying values, for example by stirring up hatred or violence*" or if the statement was made "*to rely on the Convention to engage in an activity or perform acts aimed at the destruction of the rights and freedoms laid down in it*."[650] The Court thus set the tipping point of outrageous speech becoming illegal speech to the speech "*aimed at the destruction of the rights and freedoms laid down*" by the Convention. Enough is enough, no more protection, or we'll implode.

In contrast, the First Amendment invites and welcomes even statements aiming at destructing the very rights and freedoms protected by the U.S. Constitution. However, as explained by the ECTHR, the damage to the dignity of the Jews was the ultimate value to be protected, above unfettered freedom of speech. As such, the Court recognized the rights of the victims of the Holocaust and their dignity.[651] After all, the Preamble of the 1948

---

[650] Paragraph 37.

[651] The necessity of protecting the dignity of a man, or of mankind, is also the root of the right to privacy. U.S. case laws appears to be sensitive to the argument that the right of privacy may place barriers on speech. Such was the case in Catsouras v. Department of California Highway Patrol, 181 Cal. App. 4th 856 (2010), where the California Court of appeals, Fourth District, held in favor of family members of the victim of a car accident, whose images taken at the scene had been widely shared online. The Court cited its own Shulman v. Group W Productions, Inc. case: "*It is in the intrusion cases that invasion of privacy is most clearly seen as an affront to individual dignity*" (Shulman v. Group W Productions, Inc, 18 Cal.4th 469.) It was "*propriety*" and "*decency*" which had been values deemed worthy by Louis Brandeis and Samuel Warren to be protected against the press and the use of "*instantaneous photographs*" , *see* Warren & Brandeis, *The Right to Privacy* (1890) 4 Harv. L. Rev. 193, 195-196. Article 16 of the French civil Code proclaims that "*law ensures the primacy of the person, prohibits any attack on his/her **dignity** and guarantees respect for the human being from the beginning of his/her life*" (our emphasis). Article 35ter of the French Press Law, created by Ordinance in 2000, makes a crime punishable by a 15,000 Euros fine to "d*isseminat*[*e*] *by any means whatsoever and whatever the medium, … the reproduction of the circumstances of a crime or an offense, when this reproduction seriously undermines the **dignity** of a victim and is carried out without the agreement of the latter"* (our emphasis). Can the dignity of mankind as a whole be protected by privacy? The question was asked to the Paris Tribunal de Grande Instance about an advertisement of Italian brand Benetton showing closeups of human bodies marked "HIV POSITIVE." A person suffering from AIDS sued Benetton, claiming that this ad, even if the persons represented were anonymous, nevertheless constituted an invasion of the privacy of all HIV-positive persons, and therefore himself. He further argued that "*whatever the intentions of its authors, the message disseminated is not explicit and remains largely ambiguous; that it borders on provocation and constitutes, in the exercise of freedom of expression, a fault that any person who has felt the effects of it can asked for sanctions*." The court held that plaintiff had no standing and refused to recognize an invasion of privacy, reasoning that protection of private life (by article 9 of the French civil Code) "*has an individual character and relates only to the attacks suffered personally by the holder of the right concerned; that it cannot be extended to elements, certainly of a*

Universal Declaration of Human Rights starts by referring to dignity: "*[w]hereas recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world*…" which can be interpreted as meaning that humanity as a whole share a common dignity.

During the debates about the constitutionality of article 24 bis of the French Press Law,[652] some intervening parties argued that the law does not respect the principle of equality before the law, as its scope is limited to contesting the crimes against humanity committed during WWII. The government argued instead that the objective of Article 24 bis is not to protect the victims of the Holocaust but instead to protect public order and fight against racism and antisemitism, as 24 bis is placed in the French Press Law just after article 24 incriminating speech inciting violence, racial or religious discrimination, or glorifying crimes against humanity. Still, the argument can still be made that the scope of the law should be extended to denying other crimes against humanity, such as the Armenian genocide or the Rwandan genocide.

b. Denying the Armenian Genocide

The Holocaust was a genocide, a word created to describe it, depicting how the "genus" (γένος) has been killed (caedere).[653] Article 24 bis of the French Press Law only

nature to be linked to the privacy of existence, but not precisely concerning that of the person who considers himself injured, to which it is not sufficient, to benefit of the aforementioned right, to invoke a simple analogy of the situation and which must necessarily justify that it itself suffers the alleged breach." The Court, however, found that " *the right to freely express thoughts and opinions, recognized in particular by the Preamble to the Constitution, is an essential public freedom; that it must nevertheless be recognized as having a limit, when whoever uses it causes it to degenerate into abuse*." Tribunal de grande instance [TGI] Paris. Feb. 1, 1995, D. 1995. 569.

[652] The debates can be seen here (in French): https://www.conseil-constitutionnel.fr/decision/2016/2015512QPC.htm, (last visited Dec. 30, 2020).

[653] The word "genocide" has been coined from these two terms by Raphaël Lemkin, R. LEMKIN, AXIS RULE IN OCCUPIED EUROPE (1944).

incriminates denying the existence of a crime against humanity, as defined by article 6 of the statutes of Nuremberg. Therefore, even if French law does recognize other crimes against humanity than the ones perpetrated during WWII, the French Press Law cannot be used as a basis to file a complaint against an individual who would publicly make the apology of crime against humanity not perpetrated during WWII. However, there are other genocides, and article 211-1 of the French criminal code defines the crime of genocide as the fact:

> "*in execution of a common plan aimed at the partial or total destruction of a national, ethnical, racial or religious group, or determined by using any other arbitrary criterion, to incite to commit or to commit, against members of these groups, one of the following acts: Murder; Causing serious bodily or mental harm; -Inflicting conditions of living calculated to bring about the total or partial destruction of the group; -Measures intended to prevent births; Forced transfer of children.*"

This definition is directly inspired by article II of the UN Convention on the Prevention and Punishment of the Crime of Genocide.[654] The issue of the Armenia genocide is not yet settled in other countries. The European Parliament adopted on April 15, 2015 a non-legislative resolution on the centenary of the Armenian Genocide, referring to the 1915

[654] Convention on the Prevention and Punishment of the Crime of Genocide. Approved and proposed for signature and ratification or accession by General Assembly resolution 260 A (III) of 9 December 1948. Entry into force: 12 January 1951, in accordance with article XIII, available at https://www.un.org/en/genocideprevention/documents/atrocity-crimes/Doc.1_Convention%20on%20the%20Prevention%20and%20Punishment%20of%20the%20Crime%20of%20Genocide.pdf. Its article II states:
"*In the present Convention, genocide means any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group, as such: (a) Killing members of the group; (b) Causing serious bodily or mental harm to members of the group; (c) Deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part; (d) Imposing measures intended to prevent births within the group; (e) Forcibly transferring children of the group to another group.*"

massacre as a genocide through the document.[655] It calls Turkey, a candidate to the European Union since 1987, "*to come to terms with its past, to recognize the Armenian Genocide and thus to pave the way for a genuine reconciliation between the Turkish and Armenian peoples*."[656] The lower house of the German Parliament, the Bundestag, recognized in June 2016 that the 1915 massacre of Armenians by the Ottoman Empire was a genocide,[657] which led Turley to recall its Ambassador.[658] Only the Armenian genocide has been recognized by France so far, by the January 21, 2001 law on the recognition of the Armenian genocide of 1915, which has only one article: "*France publicly recognizes the Armenian genocide of 1915*."[659] A January 23, 2012 law, would have denying the existence of all genocides recognized by law a crime, but was found to be unconstitutional by the *Conseil constitutionnel* on February 28, 2012.[660] Article 1 of this law would have introduced an article 24 ter in the French Press Law , which would have made it a crime to contest or minimize "*in an outrageous manner*" ("*minimisé de façon outrancière*") the existence of one or more crimes of genocide, as defined under Article 211-1 of the criminal Code, if they are recognized as genocide under French law, irrespective of the means of expression or public communication used.[661] This new crime could have carried a sentence of one-year

---

[655] *European Parliament resolution of 15 April 2015 on the centenary of the Armenian Genocide* (2015/2590(RSP)) (April 15, 2015), http://www.europarl.europa.eu/sides/getDoc.do?type=TA&language=EN&reference=P8-TA-2015-0094.
[656] Paragraph 5 of the resolution.
[657] Alison Smale and Melissa EddyJune, *German Parliament Recognizes Armenian Genocide, Angering Turkey*, THE NEW YORK TIMES, (June 2, 2016),
[658] *Bundestag passes Armenia 'genocide' resolution unanimously, Turkey recalls ambassador*, DW, (June 2, 2016), http://www.dw.com/en/bundestag-passes-armenia-genocide-resolution-unanimously-turkey-recalls-ambassador/a-19299936.
[659] Loi no2001-70 du 29 janvier 2001 relative à la reconnaissance du génocide arménien de 1915 [Law no2001-70 of January 29, 2001 on the recognition of the Armenian genocide of 1915] JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF France], Jan.30, 2001, p.1590.
[660] Conseil constitutionnel [CC] [Constitutional Court] decision No.2012-647 DC, Feb. 28, 2012, (Fr.)
[661] "« *Art. 24 ter. - Les peines prévues à l'article 24 bis sont applicables à ceux qui ont contesté ou minimisé de façon outrancière, par un des moyens énoncés à l'article 23, l'existence d'un ou plusieurs crimes de génocide*

imprisonment and/or a fine of 45,000 euros. Sixty Representatives and sixty Senators had seized the constitutional Council to challenge the constitutionality of the law. They argued that recognizing the Armenian genocide is a violation of the principle of equality of all under the law[662] as only a few genocides and their victims would be officially recognized, not all of them. The also argued that future laws could "*determine a sort of official truth… by recognizing the existence of genocide*" and that "t*herefore, there would be battles of memory, the victors of which would be those who managed to obtain legislative recognition*, "[663] and would prevent the historians and the journalists to work without engaging their criminal responsibilities. They further argued that what constitutes an outrageous minimization of the genocide is not clear and would likely be interpreted differently by the judges.[664] The government argued in response that *"denying the existence of a genocide, which reality has been established, is therefore, necessarily, to cast reproach on a determined group whose members are, at least implicitly, designated as the authors of a collective lie*."[665] The public order can thus be endangered by a collective lie which would be denying the existence of a recognized, historical fact. The Representatives argued further in their response to the

défini à l'article 211-1 du code pénal et reconnus comme tels par la loi française », *see* http://www.senat.fr/leg/tas11-052.html.

[662] *Le principe d'égalité.*

[663] Conseil constitutionnel [CC] [Constitutional Court} decision N° 2012-647 DC, Feb. 28, 2012, Senators' arguments, available at https://www.conseil-constitutionnel.fr/les-decisions/decision-n-2012-647-dc-du-28-fevrier-2012-saisine-par-60-senateurs, (last visited Dec. 30, 2020).

[664] Conseil constitutionnel [CC] [Constitutional Court} decision N° 2012-647 DC, Feb. 28, 2012, Senators' arguments, available at https://www.conseil-constitutionnel.fr/les-decisions/decision-n-2012-647-dc-du-28-fevrier-2012-saisine-par-60-senateurs, (last visited Dec. 30, 2020).

[665] "*Nier l'existence d'un génocide, lorsque la réalité de celui-ci est établie, c'est donc, nécessairement, jeter l'opprobre sur un groupe déterminé dont les membres sont, au moins implicitement, désignés comme les auteurs d'un mensonge collectif*. », Conseil constitutionnel [CC] [Constitutional Court} decision N° 2012-647 DC, Feb. 28, 2012, Observations of the government, available at https://www.conseil-constitutionnel.fr/les-decisions/decision-n-2012-647-dc-du-28-fevrier-2012-observations-du-gouvernement., (last visited Dec. 30, 2020).

government that a law cannot recognize a historical fact.[666] We saw that the Council upheld the constitutionality of article 24 bis in 2016, noting that "*the contested provisions have neither the object nor the effect of prohibiting historical debates."* This argument is however not convincing, as the law would have made a crime to contest or minimize in an outrageous manner. Therefore, contesting or minimize in a manner which is not excessive would have remained legal. Indeed, the law did not define "outrageous manner" ("*de façon outrancière*") and, if looked from an U.S. point of view, was not precise enough and thus likely to chill speech. One can however argue that denying the Armenian genocide, an event recorded and recognized by historians, is "outrageous."

However, the *Conseil Constitutionel* found the January 23, 2012 law to be an unconstitutional limitation on the exercise of freedom of expression and communication as protected by article 11 of the 1789 Declaration of Man and the Citizen. It reasoned that restrictions on freedom of speech, as protected by article 11 "*must be necessary, appropriate and proportional having regard to the objective pursued."*[667] Article 1 of the referred law would have punished denial or minimization of the existence of genocide "*recognized as such under French law*" and thus the law would have punished denying the existence and legal classification of crimes, as defined and classified by the French legislators, which is "*an unconstitutional limitation on the exercise of freedom of expression and communication*." The legislators challenging the law had argued that, by recognizing

[666] Conseil constitutionnel [CC] [Constitutional Court} decision N° 2012-647 DC, Feb. 28, 2012, Representative's reply to the government's arguments, available at https://www.conseil-constitutionnel.fr/les-decisions/decision-n-2012-647-dc-du-28-fevrier-2012-replique-par-60-deputes, (last visited Dec. 30, 2020).
[667] Conseil constitutionnel [CC] [Constitutional Court} decision N° 2012-647 DC, Feb. 28, 2012, available in English at https://www.conseil-constitutionnel.fr/en/decision/2012/2012647DC.htm.

one genocide, the law violated the principle of equality of all in under the law. Yet, article 24 bis of the French Press Law makes it a crime to deny the Holocaust, while no similar provision has been made for a genocide which has been officially recognized as having occurred by the January 21, 2001 law. As the constitutionality of article 24 bis has been upheld in 2016, and is thus no longer challengeable in court, it is clear that the principle of equality, which the legislators had claimed wanting to protect when challenging the constitutionality of a law making it a crime to deny the Armenia Holocaust, is not respected if only one particular type of genocide denial is a crime under French law.

The European Court of Human Rights referred to this French case when it held on December 2013, in *Perinçek v. Switzerland*, that the Swiss courts had violated article 10 of the ECHR when sentencing a Turkish citizen who had publicly denied the 1915 Armenian genocide.[668] The applicant in *Perinçek* claimed that he had been wrongly convicted by Swiss courts for having stated several times publicly that the Armenian genocide was an

[668] Perinçek v. Switzerland, App. No 27510/08. The Court cited extensively a study made by the Swiss Comparative Law Institute, about how different countries addressed the issue of genocide-denial: "*[i]n France and Luxembourg, the legislation refers to denial of crimes against humanity, as defined in Article 6 of the Charter of the International Military Tribunal, annexed to the London Agreement of 8 August 1945... This limitation of the substantive scope of the offence of denial of crimes against humanity is offset in Luxembourg by the fact that there is a special provision concerning denial of crimes of genocide. Denial of such crimes is punishable by the same sentences [imprisonment from eight days to six months and/or a fine ranging from 251 to 25,000 euros] as denial of crimes against humanity but the definition of genocide used for these purposes is that of the Luxembourg Law of 8 August 1985, which is general and abstract, not being limited to acts committed during the Second World War. The limited scope of the relevant provisions in France has been criticised and it should be noted in this connection that a Bill aimed at criminalising denial of the existence of the Armenian genocide was approved at its first reading by the National Assembly on 12 October 2006. Accordingly, it appears that only Luxembourg and Spain criminalise denial of crimes of genocide in their legislation, generically and without restricting themselves to particular episodes in history. In addition, denial of crimes against humanity in general is not currently a criminal offence in any country.*" The European Court of Human Rights also noted that Spains' Constitutional Court had declared unconstitutional, on November 7, 2007, "denial" of genocide which article 607.2 of the Spanish criminal Code had made a crime. It had read: "*The pursuit of an aim of total or partial destruction of a national, ethnic, racial or religious group shall give rise to the following penalties: - a sentence of fifteen to twenty years' imprisonment for killing one of its members; a sentence of fifteen to twenty years' imprisonment for sexually assaulting one of its members or inflicting injuries as described in Article 149*."

"*international lie*." A Swiss lower court had found him guilty of racial discrimination under article 261bis § 4 of the Swiss criminal Code, which makes it a crime to "*publicly, by word, writing, image, gesture, assault or otherwise, to lower or discriminate against a person, or a group of persons, in a way that undermines human dignity, because of their race, ethnicity or religion or, for the same reason, denies, grossly minimizes or seeks to justify genocide or other crimes against humanity*."[669] The lower court reasoned that, since the Armenian genocide is considered to be a fact by the Swiss public opinion, and beyond the Swiss borders as well, deferring to legal writings and international declarations, the mobile of the applicant was racial discrimination, not contribution to a historical debate. The judgment was confirmed on appeal by the *Cour de cassation pénale du Tribunal cantonal du canton de Vaud*, noting that at the time 261bis § 4 of the criminal Code was adopted, the Armenian genocide "*a historical fact, as acknowledged s by the Swiss Parliament*" and that therefore, "*the courts did not have to use the work of historians to admit its existence*."[670] The Court of last resort, the *Tribunal Fédéral*, confirmed.

Applicant argued in front of the ECtHR that sentencing him for having denied the qualification of genocide to the 1915 massacre of Armenian population by the Turkish government was a violation of article 10 of the ECHR, and that the Swiss law incriminating denying genocide was not "*necessary in a democratic society*" under the meaning of article 10.2 of the Convention. The Strasbourg court did find a violation of article 10. While the interference prescribed by the Swiss law had the legitimate aim to "*protect […] the rights of*

[669] This article had been adopted by Switzerland when acceded in 1994 to the 1965 United Nations International Convention on the Elimination of All Forms of Racial Discrimination.
[670] Perinçek at 11.

others, namely the honor of families and relatives of the victims of the atrocities committed by the Ottoman Empire against the Armenian people from 1915 on," applicant's comments did not pose a serious risk to public order.[671] There was no "*pressing social needs*" to interfere with applicant's speech, as it was not likely to not likely to stir up hatred or violence. The Court was careful to "*further reiterate* […] *that while it is an integral part of freedom of expression to seek historical truth, it is not the Court's role to settle historical issues forming part of an ongoing debate among historians*."[672]

The case was referred to the High Chamber of the Court at the request of the Swiss government. The Court held on October 15, 2015 that Switzerland had violated article 10 of the Convention. The case is interesting for a study on online hate speech, even if the case is not about online speech, because of the role of the victims in the reasoning of the Court. The Court did not find that the interference with the applicant's right to freedom of expression was necessary to prevent disorders,[673] but could be regarded as intended to protect the rights of others.[674] The Court reasoned that the Swiss Federal Court had pointed out in its judgment that many descendants of the victims of the 1915 massacre view their community as a victim of genocide. The interference with the applicant's statements "*was intended to protect that identity, and thus the dignity of present-day Armenians*." While the intent of the applicant was not to "*cast the victims in a negative light, deprived them of their*

[671] Perinçek at 75.
[672] Perinçek at 99. The Court wrote further that it "*considers it important to make clear at the outset that it is not required to determine the actual nature of the massacres and deportations suffered by the Armenian people at the hands of the Ottoman Empire from 1915 onwards, or the appropriateness of categorising such events in legal terms as "genocide", within the meaning of Article 261 bis § 4 of the Criminal Code. It is primarily for the national authorities, notably the courts, to interpret and apply domestic law (see, among many other authorities, …. The Court's task is merely to review under Article 10 the decisions delivered by the appropriate national authorities pursuant to their power of appreciation",* Perinçek at 111.
[673] Paragraph 154.
[674] Paragraph 157.

dignity, or diminished their humanity," he nevertheless "*referred to the Armenians involved in the events as "instruments" of the "imperialist powers", and accused them of "[carrying] out massacres of the Turks and Muslims*." The Court thus agreed "*that the interference was … intended to protect the dignity of those persons and thus the dignity of their descendants*."[675] However, while "*aware of the immense importance attached by the Armenian community to the question whether the tragic events of 1915 and the following years are to be regarded as genocide*," the Court did not accept that the applicant's statements "*were so wounding to the dignity*[676] *of the Armenians who suffered and perished in these events and to the dignity and identity of their descendants as to require criminal-law measures in Switzerland.*"[677] As such, the Court assessed whether the feeling of the victims towards speech were "*so wounding to their dignity*" as to warrant an interference.

The dignity of the victim was the threshold of freedom of speech. Such standard could also be applied for victim of online hate speech. Applicant had called the 1915 massacre an international lie, "*first invented in 1915 by the imperialists of England, France and Tsarist Russia, who wanted to divide the Ottoman Empire during the First World War*." He had not however denied that these massacres took place, but considered them to be acts of war, not a genocide and had not call for hatred, violence, or intolerance towards the Armenians.[678] But the speech denied Armenians their statute as victims of a genocide. The

[675] Paragraph 156.
[676] Article 7 of the Constitution of the Swiss Confederation of 1999, on "Human dignity" provides that "*Human dignity must be respected and protected*." Article 16 of the same Constitution protects freedom of expression.
[677] Paragraph 257.
[678] Applicant had said publicly that "*the Soviet archives confirm that at the time there were occurrences of ethnic conflict, slaughter and massacres between Armenians and Muslims. But Turkey was on the side of those defending their homeland and the Armenians were on the side of the imperialist powers and their instruments …there was no genocide of the Armenians in 1915*."

Court then denied them the right to consider that this denial was an egregious attempt to their dignity.

Some of the issues hotly debated on social media are similar to the ones presented to the courts: is this event a genocide, did the Holocaust take place? We will see how Facebook recently created its Oversight Board. The first cases it accepted to review were announced in December 2020. One of the fist cases accepted, 020-003-FB-UA, had been referred to the Oversight Board by a user who had posted "*alleged historical photos showing churches in Baku, Azerbaijan, with accompanying text stating that Baku was built by Armenians and asking where the churches have gone. The user stated that Armenians are restoring mosques on their land because it is part of their history. The user said that the "m.a.z.u.k.u" are destroying churches and have no history. The user stated they are against "Azerbaijani aggression" and "vandalism". The content was removed for violating Facebook's Hate Speech policy. The user indicated in their appeal to the Oversight Board that their intention was to demonstrate the destruction of cultural and religious monument*s."[679] The photos had been posted during the Azerbaijan war with Armenia, as each country claim that the agorno-Karabakh territory is theirs. This case shows how delicate it may be for the Oversight Board to decide whether a particular speech is hateful or a mere comment on the destructive powers of wars.

Social media platforms will manage content posted by their users while keeping in mind that the equilibrium between the right of individual to speak freely, and the interest

[679] *Announcing the Oversight Board's first cases and appointment of trustees*, OVERSIGHT BOARD (Dec. 2020), https://www.oversightboard.com/news/719406882003532-announcing-the-oversight-board-s-first-cases-and-appointment-of-trustees, (last visited Dec. 30, 2020).

of the states and the platform to regulate speech must be preserved. How can it be done? We will now discuss it in the third part of our study.

3rd Part: The Police - Who Should Have the Power to Delete Speech on Social Media Sites?

Professor Jack Balkin argued in an article published in 2018 that "*Free Speech Is a Triangle*,"[680] as speech is no longer regulated by the balance of the power of the governments on one side and the individuals on the other side, as it was the case last century. Private companies, such as search engines, services providers, and social media platforms, are also now playing a role. How should the respective powers of these three groups be balanced?

The authors of a May 2019 report on regulation of social networks, which had been requested by the French Secretary of State for Digital Affairs, noted that "[*t*]*he current approach of self-regulation of social networks is interesting, as it demonstrates that platforms may be part of the solution to the problems observed. They have come up with varied and agile solutions, e.g. removal, less exposure, reminder of common rules, education and victim support. But self-regulation is still evolving, remains too reactive (after the appearance of harm)…"*[681] The authors of the report called for adoption and implemented of *ex-ante* regulations in the EU, which should respect three conditions:

---

[680] Jack M. Balkin, Free Speech Is a Triangle, 118 COLUM. L. REV. 2011 (2018).
[681] *Creating a French framework to make social media platforms more accountable: Acting in France with a European vision. Mission report "Regulation of social networks – Facebook experiment" Submitted to the French Secretary of State for Digital Affairs,* May 2019, 2,3, https://www.numerique.gouv.fr/uploads/Regulation-of-social-networks_Mission-report_ENG.pdf, (last visited Dec. 30, 2020).

"*(1) … adopt a compliance approach, according to which the regulator supervises the correct implementation of preventive or corrective measures, but does not focus on the materialisation of risks nor try to regulate the service provided, (2)… concentrate on the systemic actors capable of creating significant damages to our societies, without creating entry barriers for new European operators, (3) … stay agile to confront future challenges*."

States should thus not only have the power to ask for deletion of speech published on social media, but also prevent its posting. We will first see how governments are engaged in preventing speech being posted on social media (I), then examine how and when such speech is deleted after having been posted (II).

I. State Censorship

Censorship may occur prior to the publication or after it. If a State blocks speech on social media prior to publication, it is prior restraint. The U.S. Supreme Court, *in curiam*, stated in *New York Times v. U.S.* that "[*a*]*ny system of prior restraints of expression comes to this Court bearing a heavy presumption against its constitutional validity*."[682] There can now be "digital prior restraints"[683] If a State seek to take down speech after its publication, if may be viewed by some as censorship, while other may applaud a decision necessary to protect the public order. Often, censorship is in the eye of the beholder.

[682] New York Times Co. v. United States, 403 US 713, 714 (1971) (per curiam), citing Bantam Books, Inc. v. Sullivan, 372 U. S. 58, 70 (1963).
[683] *See* Jack M. Balkin, Free Speech Is a Triangle, 118 COLUM. L. REV. 2011 (2018), at 2017-2018: "*Imposing liability on infrastructure providers unless they surveil and block speech, or remove speech that others complain about, has many features of a prior restraint, although technically it is not identical to a classic prior restrain*t."

A. Blocking Access

The United Nations Human Rights Committee stated in 2011, in its General Comment No. 34 on Article 19 of the International Covenant on Civil and Political Rights, that:

> "[*a*]*ny restrictions on the operation of websites, blogs or any other internet-based, electronic or other such information dissemination system, including systems to support such communication, such as internet service providers or search engines, are only permissible to the extent that they are compatible with paragraph 3* [*of Article 19 of the International Covenant on Civil and Political Rights*]."

Article 19 of the Covenant states which restrictions may be imposed on the right to freedom of expression, such as respect of the rights or reputations of others, protection of national security, public order, or public health or morals. However, several states have restricted access to information published on the Web beyond the limits imposed by Article 19 of the International Covenant on Civil and Political Rights.[684]

For instance, according to a *The Citizen Labs* report published in April 2020,[685] China started to censor keywords related to the COVID-19 pandemic appearing on *YY*, a Chinese live-streaming platform, as early on December 31, 2019, that is, the day after Chinese

[684] The Electronic Frontier Foundation states that government's blocking of content is a violation of Article 19 of the Universal Declaration of Human Rights, granting the right "*to seek, receive and impart information and ideas through any media and regardless of frontiers*." See Content Blocking, THE ELECTRONIC FRONTIER FOUNDATION, https://www.eff.org/issues/content-blocking, (last visited Dec. 30, 2020).
[685] Lotus Ruan, Jeffrey Knockel, and Masashi Crete-Nishihata, *Censored Contagion How Information on the Coronavirus is Managed on Chinese Social Media* (March 3, 2020), https://citizenlab.ca/2020/03/censored-contagion-how-information-on-the-coronavirus-is-managed-on-chinese-social-media/.

doctors started to warn the public[686] about the virus. *WeChat*, a Chinese messaging and social media platform also censored content related to the virus. Chinese censors appear to be more tolerant if the social media pots direct their ire at private parties, not at the government. Such was the case in 2015 when the accidental death of a 31-year-old mother on a faulty mall escalator was relayed by many Chinese on the micro-blogging platform *Weibo*.[687]

The U.S. government launched a national security review in November 2019 about the acquisition, for one billion dollars, by the Beijing ByteDance Technology Company, who owns Tik Tok, of the U.S. social media app Musical.ly. Several U.S. lawmakers had expressed their concerns about TikTok allegedly censoring politically sensitive content.[688] On March 12, 2020, Senator Josh Hawley [R-MO] introduced the "No TikTok on Government Devices Act" which would have prohibited federal employees and members of Congress to download or use TikTok *"on any device issued by the United States or a government corporation*."[689] On August 6, 2020, President Trump signed two Executive Orders:[690]

---

[686]As noted by the report, the late Dr. Li Wenliang was among these doctors. He died in February 2020 of the virus. The Chinese police forced Dr. Li to sign a statement that his warning about the virus was "illegal behavior, *" see* Chris Buckley and Steven Lee Myers, *As New Coronavirus Spread, China's Old Habits Delayed Fight*, THE NEW YORK TIMES, (Feb. 1, 2020, Updated Feb. 7, 2020), https://www.nytimes.com/2020/02/01/world/asia/china-coronavirus.html.

[687] Javier C. Hernández, *Escalator Death in China Sets Off Furor Online*, THE NEW YORK TIMES, (July 28, 2015), https://www.nytimes.com/2015/07/29/world/asia/escalator-death-in-china-sets-off-furor-online.html. The article quoted Zhan Jiang, professor of journalism at Beijing Foreign Studies University: "*The media frenzy and the demands of the relatives are pointed at the company rather than the government, so the authorities are more open*."

[688] Greg Roumeliotis, Yingzhi Yang, Echo Wang, Alexandra Alper, Exclusive: U.S. opens national security investigation into TikTok – sources, REUTERS, (Nov. 1, 2019 11:21 AM), https://www.reuters.com/article/us-tiktok-cfius-exclusive/exclusive-u-s-opens-national-security-investigation-into-tiktok-sources-idUSKBN1XB4IL.

[689] S.B 3455, 116th Congress (2019-2020). The bill carved out an exception for *"any investigation, cybersecurity research activity, enforcement action, disciplinary action, or intelligence activity*."

[690] Exec. Order No. 13942, 85 FR 48637 (2020), available at https://www.federalregister.gov/documents/2020/08/11/2020-17699/addressing-the-threat-posed-by-

EO 13942 expressed concerns that "*TikTok… reportedly censors content that the Chinese Communist Party deems politically sensitive, such as content concerning protests in Hong Kong and China's treatment of Uyghurs and other Muslim minorities. This mobile application may also be used for disinformation campaigns that benefit the Chinese Communist Party, such as when TikTok videos spread debunked conspiracy theories about the origins of the 2019 Novel Coronavirus*" and ordered the prohibition of the app. 45 days after the EO.

EO 13942 stated that "[*I*]*ike TikTok, WeChat automatically captures vast swaths of information from its users. This data collection threatens to allow the Chinese Communist Party access to Americans' personal and proprietary information*" and ordered the prohibition of the app. 45 days after the EO.[691] The U.S. Commerce Department issued a statement from U.S. Department of Commerce Secretary Wilbur Ross on September 18, 2020, about the prohibitions on buying and downloading the *WeChat* and *TikTok* apps "*to safeguard the national security of the United States.*"[692] The move was presented as necessary to protect the privacy of American citizens, whose personal data was maliciously collected by China. It became prohibited, as of September 20, 2020, "*to distribute or*

---

tiktok-and-taking-additional-steps-to-address-the-national-emergency, and Exec. Order No. 13943,. 85 FR 48641, (2020), available at https://www.federalregister.gov/documents/2020/08/11/2020-17700/addressing-the-threat-posed-by-wechat-and-taking-additional-steps-to-address-the-national-emergency.

[691] The President had invoked his authority under the International Economic Powers Act (IEEPA), which gives the President the power "*to deal* [*during peace time*] *with any unusual and extraordinary*" foreign "threat" to national security, if the President declares such emergency. The President invoked his authority under the IEEPA to issue E.O. No. 13873 on May 2019, allowing him then to identity Tik Tok as posing a risk to national security. *See* Executive Order No. 13873, Securing the Information and Communication Technology and Supply Chains, 84 Fed. Reg. 22689 (May 15, 2019).

[692] Wilbur Ross, *Commerce Department Prohibits WeChat and TikTok Transactions to Protect the National Security of the United States*, U.S. COMMERCE DEPARTMENT, (Sep. 18, 2020), https://www.commerce.gov/news/press-releases/2020/09/commerce-department-prohibits-wechat-and-tiktok-transactions-protect.

maintain the WeChat or TikTok mobile applications, constituent code, or application updates through an online mobile application store in the U.S." and using the *WeChat* app to transfer funds or processing payments within the U.S. was also prohibited. Tik Tok and ByteDance filed a lawsuit against Donald Trump, in his official capacity, on September 18, 2020, claiming, inter alia, violation of the First Amendment and of the International Economic Powers Act. U.S. District Judge Carl Nichols, from the District Court for the District of Columbia, granted plaintiffs 'motion for preliminary injunction on December 7, 2020.[693] Judge Nichols did not address the First Amendment issue, but an Amicus Curiae brief filed by NetChoice argued that the EO violated the First Amendment rights of TikTok users in the United States and of U.S. companies and developers.[694] The brief cited an article relating that "*TikTok has become a prominent venue for ideological formation, political activism*…" giving as example the use of the social media platform to organize a mass false-registration drive to skew figures of registered attendants of a June 2020 Trump rally in Tulsa, Oklahoma.[695]

On May 5, 2008, the Ankara Criminal Court of First Instance blocked access to YouTube after ten pages of the video-sharing website had allegedly insulted the memory of Atatürk, the founder of modern Turkey, thus infringing Law no. 5816 of 25 July 1951 prohibiting insults to the memory of Atatürk, which is illegal to do so online under of Law no. 5651 of 4 May 2007 on regulating Internet publications and combating Internet offences. Article 8 gives of the May 4, 2007 law gives the Turkish government the power to

[693] Order, TikTok Inc. v. Donald J. Trump, 1: 20-cv- 02658 (D.D. C. Dec. 7, 2020).
[694] Brief for NetChoice as Amici Curiae Supporting Plaintiffs, TikTok Inc. v. Donald J. Trump, 1: 20-cv- 02658 (D.D. C. Sept. 26, 2020).
[695] John Herrman, *TikTok Is Shaping Politics. But How?,* N.Y.TIMES(June, 28, 2020), https://www.nytimes.com/2020/06/28/style/tiktok-teen-politics-gen-z.html.

block Internet access to some publications. Such access blocking (*erişimin engellenmesi*) can be done "*where there are sufficient grounds to suspect that, by their content, they constitute offenses*" such as incitement to suicide, sexual abuse of minors, facilitating the use of narcotics, providing a product dangerous for health, obscenity, prostitution, gambling and insulting the memory of Atatürk. Article 8(2) of the law specifies that it is a judge who blocks access during the investigation stage and a court of law if the case being prosecuted. A copy of the decision of the judge or the court is provided to the head of Turkey's Information and Communications Technologies Authority, so that it may execute the decision However, if the ISP or the content provider is not located in Turkey, then the President of Turkey's Information and Communications Technologies Authority makes a statutory decision and inform the ISP of it, asking it to block the access to the content. Following the blocking of YouTube's access, three Turkish law professors filed suit, claiming that blocking access to YouTube infringed their own freedom to receive and impart information and ideas, as protected by the ECHR. The Ankara criminal court dismissed their case, arguing they lacked standing. They lost on appeals and then lodged their case with the ECtHR, which held on December 1, 2015, that Turkey had violated article 10 of the European Convention on Human Rights.[696] Turley blocked access to the *WikiLeaks* website in July 2016 after the site published some 300,000 emails sent by the political party of Turkey's President Recep Tayyip Erdoğan, the Justice and Development

[696] Cengiz and others v. Turkey, Application nos. 48226/10 and 14027/11 (Dec. 1, 2015).

Party (AKP).[697] In 2015, Facebook blocked access in Turkey to pages that a Turkish court had found to be insulting to the Prophet Mohammad.[698]

In November 2019, the Iran government enforced an Internet blackout apparently to stifle protests over anti-government protests about increases in oil prices.[699] In Belarus, the legitimacy of the election was contested by Belarusians after Alexander Lukashenko was declared the winner of the Presidential election of August 2020, with 80 per cent of the vote, and citizens organized protests using the Web. The elected President then cut down Internet access around the country to stifle these organizing efforts and chill speech.[700] The protesters then started using the *Telegram* messaging app, which was still accessible.[701] A social media technology thus allowed protests against a contested election to move ahead. A few days later, the President-elect shut down the Internet again to stop the viral spread of a video showing him being booed by factory workers, shouting "Go Away!"[702]

---

[697] Kareem Shaheen and agencies, *Turkey blocks access to WikiLeaks after Erdoğan party emails go online*, THE GUARDIAN (July 20, 2016 5: 38 EDT), https://www.theguardian.com/world/2016/jul/20/turkey-blocks-access-to-wikileaks-after-erdogan-party-emails-go-online.

[698] Sebnem Arsu and Mark Scottjan*, Facebook Is Said to Block Pages Critical of Muhammad to Avoid Shutdown in Turkey,* THE NEW YORK TIMES (Jan. 26, 2015), http://www.nytimes.com/2015/01/27/world/europe/facebook-said-to-block-pages-on-muhammad-to-avoid-ban-in-turkey.html. The article states that Facebook removed the pages at the demand of a court, not at the direct demand of the Information and Communications Technologies Authority.

[699] Chris Baraniuk, *Iran's internet blackout reaches four-day mark*, BBC, (Nov. 13, 2019), https://www.bbc.com/news/technology-50490898; Farnaz Fassihi, Iran Blocks Nearly All Internet Access, THE NEW YORK TIMES, (Nov. 17, 2019), https://www.nytimes.com/2019/11/17/world/middleeast/iran-protest-rouhani.html.

[700] Yan Auseyushkin and Andrew Roth, *Will knocking Belarus offline save president from protests?* THE GUARDIAN, (Aug. 11, 2020), https://www.theguardian.com/world/2020/aug/11/belarus-president-cuts-off-internet-amid-widespread-protests.

[701] Max Seddon, *Protesters find way round Belarus's internet blackout*, FINANCIAL TIMES, (Aug, 13, 2020), https://www.ft.com/content/3466da92-946e-4d29-81b3-e96ba599a63e.

[702] David Gilbert, *Belarus Shut Down the Internet to Stop a Wild Viral Video of Its President*, VICE, (August 17, 2020, 7:59am),

In Europe, the Declaration on freedom of communication on the Internet adopted on May 28, 2003 by the Committee of Ministers of the Council of Europe[703] states in its Principle 3 that "[*p*]*ublic authorities should not, through general blocking or filtering measures, deny access by the public to information and other communication on the Internet, regardless of frontiers*."[704] The right of the states to block access to content illegal under their laws is not without limits. The ECtHR held in June 2020, in *Kharitonov v. Russia*,[705] that Russia had violated the ECHR by blocking applicant's website about cannabis. This case was one of two other cases where the ECtHR addressed the issue of Russia blocking access to websites[706] The Court:

> "*reiterate*[*d*] *that owing to its accessibility and capacity to store and communicate vast amounts of information, the Internet has now become one of the principal means by which individuals exercise their right to freedom of expression and information. The Internet provides essential tools for participation in activities and discussions concerning political issues and issues of general interest, it enhances the public's access to news and facilitates the dissemination of information in general. Article 10 of the Convention guarantees "everyone" the freedom to receive and impart information and ideas. It applies not only to the content of information but also to the means of its*

[703] Declaration on freedom of communication on the Internet, (Adopted by the Committee of Ministers on 28 May 2003 at the 840th meeting of the Ministers' Deputies).
[704] Principle 3 went on by stating that "[*t*]*his does not prevent the installation of filters for the protection of minors, in particular in places accessible to them, such as school or libraries*."
[705] Vladimir Kaharitonov v. Russia, Application n° 10795/14 (June 23, 2020). In this case, the IP address of Applicant's website had been blocked pursuant to a decision of the Federal Drug Control Service , which intended, however, to block access to another website, featuring cannabis-themed folk stories .
[706] Flavus and Others v. Russia, Application n° 12468/15 (June 23, 2020); Engels v. Russia, Application n° 61919/16 (June 23, 2020).

dissemination, for any restriction imposed on the latter necessarily interferes with that freedom."[707]

The Court also "*reiterat*[*ed*] *that the wholesale blocking of access to an entire website is an extreme measure which has been compared to banning a newspaper or television station, the Court considers that a legal provision giving an executive agency so broad a discretion carries a risk of content being blocked arbitrarily and excessively*."[708]

In this case, applicant's website had been blocked, albeit by mistake, because its topic was cannabis, a substance now legally sold in some U.S. states. However, possession of such substance is still a criminal offense in many countries, France included. What constitutes a crime is more fluid than what would expect.

B. When Social Media Posts Lead to Imprisonment

Publishing pictures of a couple kissing on social media is a mundane activity in the U.S. and the European Union but may lead to arrest and even prison terms in other countries. Article 638 of the Iranian criminal Code makes it a crime punishable by ten days to two months imprisonment or a fine of fifty thousand to five hundred Rials[709] for a woman to appear in a public place without a hijab.[710] Article 639 of the same Code makes facilitation or encouragement of immorality or prostitution a crime punishable by one to ten years imprisonment. These articles were the legal basis for the arrest in Tehran, in May

[707] Kharitonov v. Russia, § 33.
[708] Kharitonov v. Russia, § 38.
[709] A little less than two US dollars to a little less than two US cents in current exchange rate.
[710] *See Islamic Penal Code of the Islamic Republic of Iran – Book Five*, IRAN HUMAN RIGHTS DOCUMENTATION CENTER,(July 15, 2013), https://iranhrdc.org/islamic-penal-code-of-the-islamic-republic-of-iran-book-five/#17

2020, of Iranian "parkour"[711] amateur Alireza Japalaghi. His crime was to have posted on his Instagram account photographs of him kissing on the roofs of skyscrapers a young woman dressed in a bathing suit and not wearing a veil.[712] Three Iranian women who had published dance videos on Instagram were arrested in October 2019 for having published obscene material.[713] What is viewed as art in the Occident is obscene in other countries… In Egypt, a prominent belly dancer, Sama el-Masry, was sentenced in June 2020 to three years in prison and was fined 300,000 Egyptian pounds[714] for having incited debauchery and immorality when posting videos on *TikTok* deemed to be too suggestive.[715] The 2018 Egyptian Anti-Cybercrime law makes it a crime to post content online that "*violates the family principles and values upheld by Egyptian society*," which is punished by a minimum of six-months' imprisonment and/or a fine of 50,000 to 100,000 Egyptian pounds.[716]

When French philosopher Blaise Pascal stated, "*Truth this side of the Pyrenees, error this side*," he alluded to the geographical barrier of the Pyrenees mountains, establishing a physical barrier between two ways of looking at an issue. But social media platforms allow users to share content all around the world. Twitter explains that it offers its services *"in*

---

[711] The Merriam-Webster dictionary defines "parkour" as *"the sport of traversing environmental obstacles by running, climbing, or leaping rapidly and efficiently*", see *Parkour*, MERRIAM WEBSTER, https://www.merriam-webster.com/dictionary/parkour (last visited Dec. 30, 2020).
[712] *En Iran, une star de parkour arrêtée pour des photos de baiser trop suggestives*, FRANCE 24 LES OBSERVATEURS, (MAY 19, 2020), https://observers.france24.com/fr/20200519-iran-parkour-photo-baiser-islamique-alireza-japalaghy.
[713] *Iran : trois femmes arrêtées et emprisonnées pour avoir dansé sur Instagram*, FRANCE 24 LES OBSERVATEURS, (Nov. 11, 2019), https://observers.france24.com/fr/20191101-iran-danseuses-femmes-arretees-emprisonnees-instagram.
[714] Approximately 18, 600 U.S. dollars.
[715] *Egyptian belly-dancer given three-year jail term for 'inciting debauchery'*, THE GUARDIAN, (Jun. 7, 2020 17.01 EDT), https://www.theguardian.com/world/2020/jun/27/egyptian-belly-dancer-given-three-year-jail-term-for-inciting-debauchery.
[716] *Egypt: President Ratifies Anti-Cybercrime Law*, LIBRARY OF CONGRESS LEGAL MONITOR, (Oct. 5, 2018), https://www.loc.gov/law/foreign-news/article/egypt-president-ratifies-anti-cybercrime-law.

order to give everyone the power to create and share ideas and information instantly, without barriers."[717] Do social media platforms allow for the unfettered sharing of ideas, without borders?

II. The Platforms

The European Commission noted in 2016 that "[*o*]*nline platforms have dramatically changed the digital economy over the last two decades and bring many benefits in today's digital society*."[718]

The March 3, 2017 *Joint declaration on freedom of expression and "fake news", disinformation and propaganda*[719] warned that if intermediaries restrict third party speech, by deleting it or moderating it "*beyond legal requirements,*" they should then "*adopt clear, pre-determined policies governing those actions…based on objectively justifiable criteria rather than ideological or political goals and should, where possible, be adopted after consultation with their users*." Do social media platforms heed to this advice?

A. How Laws Regulate Social Media Platforms

Article 15 of Thailand's Computer Crime Act provides that internet service providers are liable for their users' activities.[720] This is not a view shared by the U.S. and the E.U. who consider instead that they are not publishers. Should social media platforms

---

[717] *Twitter, our services, and corporate affiliates*, TWITTER, https://help.twitter.com/en/rules-and-policies/twitter-services-and-corporate-affiliates, (last visited Dec. 30, 2020).
[718] *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Region, Online Platforms and the Digital Single Market Opportunities and Challenges for Europe*, COM(2016) 288 final, THE EUROPEAN COMMISSION, (May 25, 2016), p.2, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016DC0288&from=EN.
[719] *Joint declaration on freedom of expression and "fake news", disinformation and propaganda,* ORGANIZATION FOR SECURITY AND CO-OPEARATION IN EUROPE (March 3, 2017), https://www.osce.org/fom/302796.
[720] https://www.article19.org/data/files/medialibrary/1739/11-03-14-UPR-thailand.pdf

be viewed as publishers? The answer to this question is not without interest for democracy, as many social media users are getting their news from them, or, rather, through them. A Pew Research Center survey conducted in 2016 found that 62% of U.S. adults get news on social media, and 18% even do it often, while only 49% of U.S. adults had reported in 2012 getting news on social media.[721] The survey also found that 66% of Facebook users got news from the world's biggest social networking site, while only 59% of Twitter users did so.

Social media facilitates the spread of news, any kind of news: fresh or old, local or of worldwide interest, real or fake. Broadcasting news using new technologies is not a recent invention: it has been argued that the word "phony" stems from "telephone," an instrument which was considered, when first introduced, to spread untruth.[722] Former President Barak Obama said in a November 2020 interview, about social media platforms:

> "*The degree to which these companies are insisting that they are more like a phone company than they are like The Atlantic, I do not think is tenable. They are making editorial choices, whether they've buried them in algorithms or not. The First Amendment doesn't require private companies to provide a platform for any view that is out there*."[723]

[721] Jeffrey Gottfried and Elisa Shearer, *News Use Across Social Media Platforms 2016*, PEW RESEARCH CENTER, (May 26, 2016), http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/.
[722] This theory is mentioned by Lawrence Lessing, CODE AND OTHER LAWS OF CYBERSPACE, Basic Books (1999), p. 171.
[723] Jeffrey Goldberg, *Why Obama Fears for Our Democracy*, THE ATLANTIC (Nov. 17, 2020), https://www.theatlantic.com/ideas/archive/2020/11/why-obama-fears-for-our-democracy/617087.

The law in this field is likely to evolve in the next years, in the U.S. and in the E. U., and the scope and legality of the platforms' power to take down speech is likely to dramatically change.[724]

a. Overview of US Law

When the Second Circuit Court of appeals affirmed in July 2019 the Southern District of New York's judgment which had found that the "interactive space of a tweet" published by Donald Trump from his @realDonaldTrump Twitter account was a public forum, the court carefully noted that it had not "*consider*[*ed*] *or decide*[*d*] *whether private social media companies are bound by the First Amendment when policing their platforms.*[725]

Section 230 of the Communications Decency Act of 1996 (CDA)was passed[726] as part as Title V of the Telecommunications Act of 1996,[727] following the *Stratton Oakmont, Inc., v. Prodigy Services Co.* case[728] where the New York Supreme Court[729] held that the operator of a computer bulletin board, on which a third party had posted defamatory allegations, was a publisher, as it played an active role in monitoring its bulletin boards. The CDA originally contained provisions aiming at protecting minors from "*indecent*" and "*patently offensive*"

---

[724] For an explanation on how the platforms control speech see Tim Wu, *Will Artificial Intelligence Eat the Law: The Rise of Hybrid Social-Ordering Systems*, 119 COLUM. L. REV. 2001, 2013-2018 (2019).
[725] Knight First Amendment Institute Columbia v. Trump, 928 F. 3d 226, 229 (2019). We will discuss the case further on.
[726] See H.R. Doc. No. 104-458 (1996), available at https://www.congress.gov/104/crpt/hrpt458/CRPT-104hrpt458.pdf: "[*one*] *of the specific purposes of* [*section 230*] *is to overrule Stratton-Oakmont v. Prodigy and any other similar decisions which have treated such providers and users as publishers or speakers of content that is not their own because they have restricted access to objectionable material*."
[727] Telecommunications Act of 1996, Pub. L. 104-104, 110 Stat. 56.
[728] Stratton Oakmont, Inc. v. Prodigy Services Co., (N.Y. Sup. Ct., 1995 WL 323710). The Court was aware of the possibility that the CDA may become law, noting that "*… the issues addressed herein may ultimately be preempted by federal law if the Communications Decency Act of 1995, several versions of which are pending in Congress, is enacted*.")
[729] The New York Supreme Court is a court of first jurisdiction.

speech online by providing that knowingly making, creating, soliciting or initiating the transmission "*in interstate or foreign communications*, *by means of telecommunications device*… *any comment, request, suggestion, proposal, image, or other communication which is obscene or indecent, knowing that the recipient of the communication is under 18 years of age, regardless of whether the maker of such communication placed the call or initiated the communication*" was punishable by a fine and two years in jail. The law also provided that using "*an interactive computer service*" in interstate or foreign communications to send to a "*specific*" minor or to minors "*any comment, request, suggestion, proposal, image, or other communication that, in context, depicts or describes, in terms patently offensive as measured by contemporary community standards, sexual or excretory activities or organs, regardless of whether the users of such service placed the call or initiated the communication*" was a crime.

The Supreme Court, however, found these provisions to violate the First Amendment as overbroad:[730]

> "*We are persuaded that the CDA lacks the precision that the First Amendment requires when a statute regulates the content of speech. In order to deny minors access to potentially harmful speech, the CDA effectively suppresses a large amount of speech that adults have a constitutional right to receive and to address to one another. That burden on adult speech is unacceptable if less restrictive alternatives would be at least*

[730] Reno v. American Civil Liberties Union, 521 U.S. 844 (1997). Writing for the Court, Justice Stevens noted that "*each of the two parts of the CDA uses a different linguistic form*," as one uses the word "*indecent*," while the other mentions speech that "*in context, depicts or describes, in terms patently offensive as measured by contemporary community standards, sexual or excretory activities or organs*." The Court noted that these terms are not defined and that "*this difference in language will provoke uncertainty among speakers about how the two standards relate to each other and just what they mean*", adding that "[*t*]*his uncertainty undermines the likelihood that the CDA has been carefully tailored to the congressional goal of protecting minors from potentially harmful materials*". Reno, at 871.

as effective in achieving the legitimate purpose that the statute was enacted to serve."[731]

The aim of the CDA was to protect freedom of speech, as explained in the statutory findings of the law: "[*t*]*he Internet and other interactive computer services offer a forum for a true diversity of political discourse, unique opportunities for cultural development, and myriad avenues for intellectual activity*."[732]

As explained by the Fourth Circuit Court of Appeals in 1997, shortly after the enactment of Section 230:

> "*Congress made a policy choice ... not to deter harmful online speech through the separate route of imposing tort liability on companies that serve as intermediaries for other parties' potentially injurious messages. Congress' purpose in providing the § 230 immunity was thus evident. Interactive computer services have millions of users. The amount of information communicated via interactive computer services is therefore staggering. The specter of tort liability in an area of such prolific speech would have an obvious chilling effect. It would be impossible for service providers to screen each of their millions of postings for possible problems. Faced with potential liability for each message republished by their services, interactive computer service providers might choose to severely restrict the number and type of messages posted. Congress*

[731] Reno, at 874.
[732] 47 U.S.C. § 230 (a)(3)

considered the weight of the speech interests implicated and chose to immunize service providers to avoid any such restrictive effect."[733]

Under 230(c)(1) of the CDA,[734] "[*n*]*o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider*." Section 230 of CDA defines *"interactive computer service*" as an information service, system, or access software provider providing or enabling computer access by multiple users to a computer server.[735] As such, social media platforms are included in the definition and are protected by the CDA safe harbor, thus immune from liability because they only act as intermediaries of third-party content and cannot as such be held liable for content posted through their services.[736] They are intermediaries, not publishers. The Section 230 of the CDA thus provides social media platforms broad immunity from liability for content created by users. The Ninth Circuit explained in 2003, in *Carafano v. Metrosplash. com*, that "*an "interactive computer service" qualifies for immunity so long as it does not also function as an "information content provider" for the*

[733] Zeran v. America Online, Inc., 129 F. 3d 327, 328-29 (4th Circ.1997)
[734] 47 U.S.C. § 230(c)(1).
[735] 47 U.S.C. § 230(f)(2): "*The term "interactive computer service" means 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*."
[736] See for instance Sikhs for Justice" SFJ", Inc. v. Facebook, Inc., 144 F. Supp. 3d 1088, 1093 (N.D. California 2015): "[*t*]*he Court…agrees that Defendant "provides or enables computer access by multiple users to a computer service" as required by § 230*." In Fraley v. Facebook, Inc., 830 F. Supp. 2d 785, 801 (N.D. California 2011), the Northern district Court of California had found that Facebook met the definition of an interactive computer service under the CDA, but that also that " *in the context of Plaintiffs' claims, it also [met] the statutory definition of an information content provider… 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.*" The Court cited Fair Housing Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1162 (9th Cir.2008) (en banc): "*A website operator can be both a service provider and a content provider.... [A]s to content that it creates itself, or is `responsible, in whole or in part' for creating or developing, the website is also a content provider*."

portion of the statement or publication at issue."[737] Therefore, if a social media platform creates or develop content, it is an information content provider as to that content and is not immune from liability.[738] Section 230 of the CDA defined "*information content provider*" 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*."[739]

What does "development of information" mean? The Ninth Circuit cautioned in 2003, *in Fair Housing Council of San Fernando Valley v. Roommates. com,* that it should not be broadly interpreted:

> "*…it's true that the broadest sense of the term "develop" could include the functions of an ordinary search engine — indeed, just about any function performed by a website. But to read the term so broadly would defeat the purposes of section 230 by swallowing up every bit of the immunity that the section otherwise provides*."[740]

---

[737] Carafano v. Metrosplash. com. Inc., 339 F. 3d 1119, 1123 (9th Circ. 2003).

[738] See Jones v. Dirty World Entertainment Recordings LLC, 755 F. 3d 398, 408, (6th Circ. 2014). In that case, Plaintiffs had sued a website and its owner for defamation over publications presenting her as a promiscuous person of ill virtue The District Court had agreed with the plaintiff, holding that the website and its owner were not immune under the CDA because they were information content providers with respect to the information underlying plaintiff defamation claims because they developed that information, Jones v. Dirty World Entertainment Recordings, LLC, 965 F. Supp. 2d 818, 823 (ED Kentucky 2013). The Court proposed to measure if a defendant indeed developed this information as such: "*a website owner who intentionally encourages illegal or actionable third-party postings to which he adds his own comments ratifying or adopting the posts becomes a `creator' or `developer' of that content and is not entitled to immunity*", Id. at 821.

[739] 47 U.S.C. § 230(f)(3).

[740] Fair Housing Council of San Fernando Valley v. Roommates. com, LLC, 521 F. 3d 1157, 1167 (9th Circ. 2003). In this case, Roomates.com was a website allowing users to find…roommates. Its subscribers could create a profile, and had to disclose their sex, sexual orientation, and whether children were living with them. They could also provide "Additional Comments" which could only be seen by paid subscribers, not by the persons using the site for free. The Fair Housing Councils of the San Fernando Valley and San Diego sued the website, claiming that asking these questions violated the federal Fair Housing Act. The District Court found the site to be immune under Section 230 of the CDA and dismissed the federal claims. However, the Ninth Circuit found that the website was not immune, as, "[*b*]*y requiring subscribers to provide the information as a condition of accessing its service, and by providing a limited set of pre-populated answers, Roommate becomes*

However, the 9th Circuit "*interpret*[*ed*] *the term "development" as referring not merely to augmenting the content generally, but to materially contributing to its alleged unlawfulness. In other words, a website helps to develop unlawful content, and thus falls within the exception to section 230, if it contributes materially to the alleged illegality of the conduct*."[741]

As noted by the Second Circuit in 2019, "[*i*]n *light of Congress's objectives, the Circuits are in general agreement that the text of Section 230(c)(1) should be construed broadly in favor of immunity*."[742] Section 230(c)(1) "*immunizes decisions to delete user*

much more than a passive transmitter of information provided by others; it becomes the developer, at least in part, of that information. And section 230 provides immunity only if the interactive computer service does not "creat[e] or develop[]" the information "in whole or in part", Fair Housing, at 1166.

[741] Fair Housing, at 1167-68.

[742] Force v. Facebook, Inc., 934 F. 3d 53, 64 (2nd Circuit 2019).

profiles."[743] It also immunized platforms against defamation claims,[744] federal anti-terrorism claims,[745] or housing discrimination claims.[746]

---

[743] Riggs v. MySpace, Inc., 444 Fed. App'x 986, 987 (9th Cir. 2011). *See* also Federal Agency of News LLC. V. Facebook , Inc (N.D. California 2019), about a Russian-language Facebook account and page shut down because it violated Facebook's Terms of Service, one of the 270 or so Russian language accounts and pages Facebook shut down by on April 3, 2018 because they were allegedly controlled by the Russia-based Internet Research Agency (IRA). Facebook's move was part of its effort to protect the integrity of the elections, as explained by Marc Zuckerberg in an April 3, 2018 blog post: FACEBOOK (April 3, 2018), https://www.facebook.com/zuck/posts/10104771321644971.

[744] *See* for example, Brikman v. Twitter, Inc., 2020 WL 5594637 (E.D.N.Y. Sept. 17, 2020), available at https://www.govinfo.gov/content/pkg/USCOURTS-nyed-1_19-cv-05143/pdf/USCOURTS-nyed-1_19-cv-05143-0.pdf. Plaintiffs, the Rabbi of Kneses Israel of Seagate, a Brooklyn synagogue, and other members of the synagogue, had sued Twitter *pro se*, claiming that, by failing to take down content allegedly defamatory, posted by an unknown party using the @KnesesG Twitter account, Twitter had "*knowingly and with malice… allowed and helped non-defendant owners of Twitter handle @KnesesG, to abuse, harras[sic], bully, intimidate, [and] defame" plaintiffs*." The Eastern District of New York Court found that Plaintiffs' claims were foreclosed by the CDA. Twitter an interactive computer service, Plaintiffs' claims were based on information provided by another information content provider, the unknown party who had created the @KnesesG Twitter account, and Plaintiffs' claims would treat Twitter as the publisher or speaker of the information provided by another information content provider.

[745] Force v. Facebook, Inc., 934 F. 3d 53 (2nd Circuit 2019). Plaintiffs alleged that Facebook had unlawfully provided Hamas with a communications platform enabling terrorist attacks, posting messages advocating kidnapping Israeli soldiers, or encouraging car-ramming attacks at light rail stations. Facebook had raised Section 230(c)(1) immunity, as an affirmative defense and the Southern District of New York granted Facebook's motion to dismiss. The Second Circuit Court of appeals affirmed. Plaintiff had claimed that Facebook was an "information content provider," withing the meaning of Section 230, as its algorithms had developed Hamas's content "*by directing such content to users who are most interested in Hamas and its terrorist activities, without those users necessarily seeking that content*"(Force, at 68). The Second Circuit uses the " material contribution test" to find out whether a defendant has developed content: the defendant must have directly and "materially" contributed to what made the content itself "unlawful, " *see* LeadClick, 838 F.3d at 174, quoting Roommates.Com, 521 F.3d at 1168. The Ninth Circuit uses for that purpose the "material contribution" test, which "*draw[s] the line at the `crucial distinction between, on the one hand, taking actions... to... display... actionable content and, on the other hand, responsibility for what makes the displayed content [itself] illegal or actionable*.'" Kimzey v. Yelp! Inc., 836 F.3d 1263, 1269 n.4 (9th Cir. 2016), (quoting Jones, 755 F.3d at 413-14). Facebook does not edit or suggest edits for content published by its user, including Hamas, which is a "*practice… consistent with Facebook's Terms of Service, which emphasize that a Facebook user "own[s] all of the content and information [the user] post[s] on Facebook, and [the user] can control how it is shared through [the user's] privacy and application settings*." The platform does note acquire information from its users, beyond their names, telephone number and email addresses and thus is not a developer under Section 230, but a "*neutral intermediary*, " LeadClick, 838 F.3d at 174. Facebook's algorithms are "*content neutral*" as they "*take the information provided by Facebook users and "match" it to other users—again, materially unaltered— based on objective factors applicable to any content, whether it concerns soccer, Picasso, or plumbers*, "Force at 70.

[746] Chicago Lawyers' Comm. Civ. Rights v. Craigslist, 461 F. Supp. 2d 681 (N.D. Illinois 2006). Plaintiff alleged that Craigslist published, in violation of the Fair Housing Act, housing ads indicating "*a preference, limitation, or discrimination, or an intention to make a preference, limitation, or discrimination, on the basis of race, color, national origin, sex, religion and familial status*, " providing examples such as "*African Americans and Arabians tend to clash with me so that won't work out*." The Court noted that "*Congress made a policy choice… not to deter harmful online speech through the separate route of imposing tort liability on companies that serve as intermediaries for other parties' potentially injurious messages.*" Chicago Lawyers' Comm. Civ. Rights, at 689.

The CDA is a shield, but also provides a sword to providers and users of interactive computer services, as its Section 230(c)(2), the so-called "Good Samaritan" provision, makes them immune from civil responsibility if for "*any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected*." As such, platforms may engage in voluntary self-regulation by taking down material, even protected speech. Professor Danielle Keats Citron noted in her book on hate crimes in cyberspace that some sites "*make a mockery*" of Section 230's safe harbor provision as they take down speech only if paid to do so, by a fee or even monthly dues,[747] and suggested that "*Congress should amend the CDA to exclude these bad actors from its protection.*"[748] However, the major social platforms do not ask to be financially compensated to take down speech, but do so pursuant to their own private guidelines, as we will see further on. The scope of Section 230 is broad, and not precisely defined, as it contains a "catchall" provision allowing providers to take down speech which is "*otherwise objectionable*." The Ninth Circuit Court of appeal recognized in 2019 in *Enigma Software Group USA, LLC v. Malwarebytes*, that this provision "*establishes a subjective standard whereby internet users and software providers decide what online*

[747] *See* for example People v. Bollaert, 248 Cal. App. 4th 699 (4th Appellate Dist., Calif. 1st Div. 2016), a case about the conviction of a man who operated both the <UGotPosted.com> website, where users could post private, intimate photographs of others along with that person's name, location and social media profile links, and the <ChangeMyReputation.com> which could be used by victims to have the information removed, for a fee. As explained by the court, at 716, "*the evidence showed that when many of the victims e-mailed the webmaster contact at UGotPosted.com and pleaded to have their personal information removed, they received no response. For at least one victim, Bollaert used a false name when he acted as the UGotPosted.com contact person. But victims who contacted and paid ChangeMyReputation.com were successful in their efforts. One victim, Brian, communicated with the e-mail associated with ChangeMyReputation.com to remove his photos and asked whether he could submit payment after his photographs were removed; he received a response telling him, "[W]e can't remove it until you pay*."

[748] DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE (Harvard University Press), 25, (2014),

material is objectionable,"[749] adding that "[*t*]*he history of § 230(c)(2) shows that access to pornography was Congress's motivating concern, but the language used in § 230 included much more, covering any online material considered to be "excessively violent, harassing, or otherwise objectionable*."[750]

Section 230 has its detractors. One of the arguments against Section 230 is that it provides immunity to platforms, and that this "*isolation from liability via Section 230 will increase the prevalence of low-value speech, as well as speech that causes dignitary harm*."[751] The Allow States and Victims to Fight Online Sex Trafficking Act and Stop Enabling Sex Traffickers Act[752] (FOSTA-SESTA Act), signed into law in 2018, narrowed the scope of Section 230 by adding a paragraph (5) to it specifying that Section 230 has:

> "*no effect on sex trafficking law*," as "[*n*]*othing in this section (other than subsection (c)(2)(A)) shall be construed to impair or limit—(A) any claim in a civil action brought*

[749] Enigma Software Group USA, LLC v. Malwarebytes, 946 F. 3d 1040, 1044 (9th Cir. 2009), citing Zango Inc. v. Kaspersky Lab, Inc., 568 F.3d 1169, 1173 (9th Cir. 2009). In a concurring opinion in Zango, Judge Fisher, warned that "[*t*]*he risk inheres in the disjunctive language of the statute — which permits blocking of "material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected" — and the unbounded catchall phrase, "otherwise objectionable*, " and called for Congress to clarify the statute, Zango, at 1178.
[750] Enigma Software Group at 1047.
[751] THE OFFENSIVE INTERNET, 156 (Brian Leiter, CLEANING CYBER-CESSPOOLS, GOOGLE AND FREE SPEECH), Saul Levmore and Martha C. Nussbaum, eds., 2010). Professor Leiter defines "*dignitary harm*" as "*harms to individuals that are real enough to those affected and recognized by ordinary standards of decency, though not generally actionable*." He contrasts these harms, which are not actionable, with tortious harms, such as defamation, which are actionable. In contrast with U.S. law, French law provides a right of action for "dignitary harm", such as having been insulted. The author writes about the now infamous pre-law chat room Auto-Admit, which administrators wrote, in response to one of Professor Eugene Volokh's post on his website about Auto Admit , that they were "*very strong believers in the freedom of expression and the marketplace of ideas. This is why we allow off-topic discussion and almost never censor content, no matter how abhorrent it may be*." Professor Leiter relates further that, after the *Washington Post* published an article about women's harassment on Auto Admit, naming the administrators, the law firm which had extended an offer for permanent employment to one of them rescinded the offer, writing that the firm "*expect[ed] any lawyer associated with [it] , when presented with the kind of language exhibited on the message board, to reject it and disavow it*."
[752] Allow States and Victims to Fight Online Sex Trafficking Act, Public Law No: 115-164; 47 U.S.C. 230 (e)(5).

under section 1595 of title 18, if the conduct underlying the claim constitutes a violation of section 1591 of that title; (B) any charge in a criminal prosecution brought under State law if the conduct underlying the charge would constitute a violation of section 1591 of title 18; or (C) any charge in a criminal prosecution brought under State law if the conduct underlying the charge would constitute a violation of section 2421A of title 18, and promotion or facilitation of prostitution is illegal in the jurisdiction where the defendant's promotion or facilitation of prostitution was targeted."

As such, a provider of an interactive computer service can no longer claim Section 230 civil immunity if sued by a victim of sexual exploitation in federal court[753] and in state court as well. The amendments are retroactive.

Indeed, Section 230(e)(3) on state laws provides that Section 230 cannot "*be construed to prevent any State from enforcing any State law that is consistent with [the] section. No cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with [Section 230]."* State laws aiming at fighting sexual exploitation had been found to be preempted by Section 230 For instance, a Washington State bill, SB 6251, proposed to make a felony to knowingly publish, disseminate, or display or to *"directly or indirectly"* cause content to be published, disseminated or displayed, if it contained *a "depiction of a minor*" and any *"explicit or implicit offer"* of sex for *"something of value."* The bill was to go into effect on June 7, 2012, but Backstage.com (Backstage) filed a

[753] This right is granted by 18. U.S.C. 1595, which provides victims of trafficking the right to bring a private civil action for restitution against "*whoever knowingly benefits, financially or by receiving anything of value from participation in a venture which that person knew or should have known has engaged in an act in violation of* [*Chapter 77 on peonage, slavery, and trafficking in persons*]."

lawsuit on June 4, 2012, asking the Court to enjoin enforcement of SB 6251, claiming, inter alia, that it violated Section 230 of the CDA. Backstage is a website[754] allowing users to post classified ads, in several categories, among them "Adult Entertainment," which includes an "Escorts" subcategory. Users of the site only pay to post "Adult Entertainment" ads, which can be searched by geographic area. The United States District Court for the Western District of Washington found[755] that SB 6251 was "*likely inconsistent*" with Section 230 and therefore expressly preempted by the federal law for two reasons. First, under Section 230, an online service provider cannot be treated as the publisher or the speaker of an information provided by another information content provider. However, Backstage is an online service provider, within the meaning of the law, and SB 6251 would have treated the site as the publisher or speaker of information created by third parties, by imposing liability on Backpage for information created by third parties, if knowing that it was publishing, disseminating, displaying, or causing to be published, disseminated, or displayed such information. The Court also found the Washington bill to be inconsistent with Section 230 as it criminalized the "*knowing*" publication, dissemination, or display of specified content, and thus "*created an incentive for online service providers **not** to monitor the content that passes through its channels. This was precisely the situation that the CDA was enacted to remedy*" (emphasis in the original text). Furthermore, SB 6251 "*likely*

[754] According to a 2017 staff report of the U.S. Senate Permanent Subcommittee on Investigations, Committee on Homeland Security and Governmental Affairs, "*Backpage is involved in 73% of all child trafficking reports that the National Center for Missing and Exploited Children (NCMEC) receives from the general public (excluding reports by Backpage itself*)" The Report quotes the National Association of Attorneys General describing Backpage as a "*hub*" of "*human trafficking, especially the trafficking of minors*" (Letter from the National Association of Attorneys General to Samuel Fifer, Esq., Counsel for Backpage.com LLC (Aug. 31, 2011), see United States Senate Permanent Subcommittee on Investigations Committee on Homeland Security and Governmental Affairs, *Backstage.com's Knowing Facilitation of Online Sex Trafficking* (2017), at 4.
[755] Backpage. Com, LLC v. McKenna, 881 F. Supp. 2d 1262 (W.D. Washington 2012).

conflict[*ed*] *with the CDA because "the challenged state law stands as an obstacle to the accomplishment and execution of the full purposes and objectives of Congress,*"[756] citing *Arizona v. U.S.*[757]

Before the enactment of the FOSTA-SESTA Act, a Texas attorney, Annie McAdams[758] represented anonymous Plaintiffs in several Texas cases against Facebook, claiming that Facebook had facilitated sex crimes, and that it was a tortious conduct, thus trying to avoid the issue of preemption by Section 230, and her strategy appears to have been successful, at least so far. In a case filed in Texas,[759] Facebook was sued for negligence, gross negligence, and statutory damages. The Plaintiff claimed that Facebook "*facilitates and/or was used by predators to find, groom, target, recruit and kidnap children into sex trade*" and that "*Facebook profits from the collection of data and the use of the data to target and promote interactions between Facebook users*," including interactions between minors and sex predators. Facebook moved to dismiss under Texas Rules of Civil Procedure, claiming it was not liable to Plaintiff under Section 230 of the CDA. On May 23, 2019, the Court rejected the claim, albeit laconically, simply writing that Plaintiff's action could not be

[756] Backstage v. McKenna, at 1273.
[757] Arizona v. US, 132 S. Ct. 2492, 2501 (2012).
[758] Ms. McAdams participated to the DoJ February 19, 2020, roundtable *Section 230 — Nurturing Innovation or Fostering Unaccountability?* In her written submission, she argued that "[*t*]*he overly expansive judicial interpretation of Section 230, which began almost immediately after its enactment, has provided internet-based companies nearly absolute immunity from tort (and criminal) liability for injuries they inflict upon their users*." *See* https://www.justice.gov/file/1286206/download, (last visited Dec. 30, 2020).
[759] *See* for instance Jane Doe v. Facebook, (District Court of Harris County, Texas), N°2018-69816. A copy of the 2019 judgment and of the decision of the Court of Appeals are available at Mike Masnick, *Texas Appeals Court Brushes Off Section 230 In Allowing Lawsuit Over Sex Trafficking Against Facebook To Continue*, TECHDIRT, (May 4th 2020 12:08pm), https://www.techdirt.com/articles/20200428/22031844399/texas-appeals-court-brushes-off-section-230-allowing-lawsuit-over-sex-trafficking-against-facebook-to-continue.shtml . For more background on these cases, see Jack Nicas, *Sex Trafficking via Facebook Sets Off a Lawyer's Novel Crusade*, THE NEW YORK TIMES, (Dec. 3, 2019). https://www.nytimes.com/2019/12/03/technology/facebook-lawsuit-section-230.html .

barred under Section 230. A Texas Court of Appeals upheld the decision as succinctly in May 2020, which does not allow to comment on the ways the court reached its conclusion and does not provide a solid ground for further similar litigation.

The FOSTA-SESTA Act was enacted following the dismissal of yet another case against Backstage.[760] The site had been sued, inter alia,[761] for sex trafficking by three women who had been advertised as escorts on Backstage. They claimed they had been victims of sex trafficking when minors,[762] through advertisements posted on Backpage, directly posted by sex traffickers or by the victims forced to do so by the traffickers. They further claimed that Backstage was facilitating the efforts of sex traffickers to advertise their victims on the website to gain further advertising profit. Plaintiffs also argued that the site's rules and processes about the content of advertisements were designed to encourage sex trafficking, such as not requiring verification of phone numbers or email addresses, allowing to post phone numbers in alternative formats, and giving users the option to hide their e-mail addresses in postings, as Backpage provides message forwarding and auto-replies services on behalf of the user. The District Court dismissed the complaint, albeit reluctantly,[763] because the site is immune under Section 230 of the CDA. The First Circuit

[760] Doe v. Backpage.com, LLC, 817 F.3d 12 (1st Cir. 2016).
[761] Plaintiffs also sued for unfair and deceptive business practices, unauthorized use of likeliness and copyright infringement.
[762] Sex trafficking of minors is a major societal concern: according to a 2011 report by the Bureau of Justice Statistics on the characteristics of suspected human trafficking incidents, almost 95 percent of sex trafficked victims are female, and 54 percent were 17 years old or younger, see U*.S. Department of Justice, Office of Juvenile Justice & Delinquency Prevention, Literature Review: Commercial Sexual Exploitation of Children/Sex Trafficking*, OFFICE OF JUVENILE JUSTICE AND DELINQUENCY PREVENTION, at 3 (2014) (citing Bureau of Justice Statistics data) http://www.ojjdp.gov/mpg/litreviews/CSECSexTrafficking.pdf .
[763] The Court wrote: "*Putting aside the moral judgment that one might pass on Backpage's business practices, this court has no choice but to adhere to the law that Congress has seen fit to enact*." Doe ex rel. Roe v. Backpage. com, LLC, 104 F. Supp. 3d 149", 165 (Massachusetts 2015).

Court of Appeals affirmed.[764] Appellants claimed they did not argue that Backpage was the publisher or speaker of the content of the ads at stake, but that 18 U.S.C.1595 provides victims the right to bring a civil suit against a perpetrator or against "*whoever knowingly benefits, financially or by receiving anything of value from participation in a venture which that person knew or should have known has engaged in an act*" of sex trafficking, and that the court could find the site liable without treating it as the publisher or speaker of any of the content of the sex trafficking ads. The First Circuit was not convinced by these arguments, as the website's practices are "*traditional publisher functions under any coherent definition of the term*" and that "the "*publisher or speaker" language of section 230(c)(1) extends to the formulation of precisely the sort of website policies and practices that the appellants assail.*"[765] For the First Circuit, 18 U.S.C. 1585 cannot be invoked to make Backpage a publisher with respect to third-party content, because, while a website "*might display a degree of involvement sufficient to render its operator both a publisher and a participant in a sex trafficking venture*,"[766] giving as example an instance where the website operator helped to procure the sexual trafficking victims, it was not the case here. The First Circuit held that claiming "*that a website facilitates illegal conduct through its posting rules necessarily treat the website as a publisher or speaker of content provided by third parties and, thus, are precluded by section 230(c)(1).*"[767]

Some advocated that the passage of the FOSTA-SESTA Act was not enough to fight child trafficking online, and that the CDA must be further modified to this effect to prevent

[764] Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12 (1st Cir. 2016).
[765] Doe No. 1 v. Backpage.com , at 20.
[766] Doe No. 1 v. Backpage.com , at 21.
[767] Doe No. 1 v. Backpage.com , at 22.

the immunity provided to platforms to be used from shielding them from responsibility.[768] The "Eliminating Abusive and Rampant Neglect of Interactive Technologies Act of 2020" or "EARN IT Act of 2020" (EARN ACT),[769] is a bipartisan bill introduced in the Senate on March 5, 2020 by U.S. Senators Lindsey Graham (R-South Carolina), Richard Blumenthal [D-CT], Josh Hawley [R-MO] and Dianne Feinstein [D-CA].[770] The House Version of the bill was introduced on September 30, 2020 by Representatives Sylvia Garcia [D-TX] and Ann Wagner [R-MO].[771] The EARN ACT aims at establishing a nineteen-member "National Commission on Online Child Sexual Exploitation Prevention " which purpose would be

> "*to develop recommended best practices that providers of interactive computer services may choose to implement to prevent, reduce, and respond to the online sexual exploitation of children, including the enticement, grooming, sex trafficking, and sexual abuse of children and the proliferation of online child sexual abuse material*."

The best practices would aim, inter alia, at "*preventing, identifying, disrupting, and reporting online child sexual exploitation*." As such, providers of interactive computer services would have to filter and take down child sexual abuse material (CSAM). The

---

[768] This is the opinion of the National Association of Attorneys General (NAAG) , which wrote on May 23, 2019, in a letter to several Congressional leaders, that "[*c*]*urrent precedent interpreting the CDA, … continues to preclude states and territories from enforcing their criminal laws against companies that, while not actually performing these unlawful activities, provide platforms that make these activities possible*.

[769] S.3398 - EARN IT Act of 2020, 116th Congress (2019-2020). The NAAG argued that it was necessary to amend § 230(e)(1) of the CDA, as such (added language in bold): "*Nothing in this section shall be construed to impair the enforcement of section 223 or 231 of this title, chapter 71 (relating to obscenity) or 110 (relating to sexual exploitation of children) of title 18, or any other Federal,* ***State, or Territorial*** *criminal statute*."

[770] *Graham, Blumenthal, Hawley, Feinstein Introduce EARN IT Act to Encourage Tech Industry to Take Online Child Sexual Exploitation Seriously*, U.S. SENATE, COMMITTEE ON THE JUDICIARY, https://www.judiciary.senate.gov/press/rep/releases/graham-blumenthal-hawley-feinstein-introduce-earn-it-act-to-encourage-tech-industry-to-take-online-child-sexual-exploitation-seriously, (last visited Dec. 30, 2020).

[771] H.R.8454 To establish a National Commission on Online Child Sexual Exploitation Prevention, and for other purposes, 116th Congress (2019-2020).

Commission would develop these best practices and submit them to the Attorney General, the head of the Commission.[772] Both Senator Graham and Senator Blumenthal stated that the tech companies must "earn" their immunity (hence the title of the bill….)[773]

In its current version, which differs significantly from its original version[774], it would also add the following to Section 230 (e):

"(6) NO EFFECT ON CHILD SEXUAL EXPLOITATION LAW.—Nothing in this section (other than subsection (c)(2)(A)) shall be construed to impair or limit—

---

[772] As noted by the American Civil Rights Union (ACLU), in a letter to the Senators, the personality of whomever is the Attorney General would likely influence the Commission: "*This means the best practices will overwhelmingly reflect the preferences of whoever is AG at the time the best practices are submitted and, because the practices will need to be updated every five years, they could change based upon the preferences of the individual in the AG's office*." See *ACLU Opposition to S. 3398, the EARN It Act*, March 9, 2020, (last visited?), https://www.aclu.org/letter/aclu-opposition-s-3398-earn-it-act. , President Trump's AG at the time was William Barr, who does not hide that his Christian faith influences him. For instance, he stated in its Remarks at Notre Dame Law School in October 2019 that religion "*gives us the right rules to live by. The Founding generation were Christians*." See *Attorney General William P. Barr Delivers Remarks to the Law School and the de Nicola Center for Ethics and Culture at the University of Notre Dame*, (Oct. 11, 2019), https://www.justice.gov/opa/speech/attorney-general-william-p-barr-delivers-remarks-law-school-and-de-nicola-center-ethics (last visited Dec. 30, 2020). However, Senator Blumenthal assured, in his opening statement in a March 11, 2020 hearing of the Senate Committee on the Judiciary about the EARN IT bill, that the bill "*was not about… Attorney General William Barr*." *The EARN IT Act: Holding the Tech Industry Accountable in the Fight Against Online Child Sexual Exploitation* (March 11, 2020), https://www.judiciary.senate.gov/meetings/the-earn-it-act-holding-the-tech-industry-accountable-in-the-fight-against-online-child-sexual-exploitation (@35.09). Senator Blumenthal explained that AG Barr would get one vote only on the Commission, and that the vote of the Commission would have to be at a 14 out of 19 majority. Senator Blumenthal explained further the power of the AG would be "*purely negative,*" and that the AG would not have the power to support best practices.

[773] Senator Blumenthal said in his March 11, 2020 hearing's opening statement in the hearing of the Senate Committee on the Judiciary about the EARN IT bill that "*immunity is not a right, it should be earned or deserved and certainly should not be continuing if these companies fail to observe and follow the basic moral obligations that they have under the law, but it is a matter of simple morality.*" (@36.56.). During the hearing, Senator Hawley called the immunity "*a gift from* [*the*] *government.*" Senator Hawley took the view that the immunity had been originally provided to tech companies by the CDA so that they would monitor "*certain contents*" in exchange for their immunity, but "*the content they were originally supposed to monitor is gone*", following Reno v. American Civil Liberties Union, 521 U.S. 844 (1997), and that they have now "*basically free immunity for all of these years and they've used it to their own purposes*." He further stated that "*the days of this Congress giving out free stuff to big tech, without asking or expecting anything in return are coming to an end… and rightly so because the parents of this country are demanding it*" (@1:59.00).

[774] Senator Lindsey Graham (R-SC), then Chair of the Senate Judiciary Committee, who had introduced the bill, had introduced a manager's amendment to the bill which was unanimously approved in July.

"(A) any claim in a civil action brought against a provider of an interactive computer service under section 2255[775] *of title 18, United States Code, if the conduct underlying the claim constitutes a violation of section 2252*[776] *or section 2252A of that title;*

"(B) any charge in a criminal prosecution brought against a provider of an interactive computer service under State law regarding the advertisement, promotion, presentation, distribution, or solicitation of child sexual abuse material, as defined in section 2256(8)[777] *of title 18, United States Code; or*

"(C) any claim in a civil action brought against a provider of an interactive computer service under State law regarding the advertisement, promotion, presentation, distribution, or solicitation of child sexual abuse material, as defined in section 2256(8) of title 18, United States Code.

"(*7) CYBERSECURITY PROTECTIONS DO NOT GIVE RISE TO LIABILITY.—Notwithstanding paragraph (6), a provider of an interactive computer service shall not be deemed to be in violation of section 2252 or 2252A of title 18, United States Code, for the purposes of subparagraph (A) of such paragraph (6), and shall not otherwise be subject to any charge in a criminal prosecution under State law under subparagraph (B) of such*

[775] 18.U.S.C. 2255 protects the rights of a minor victim, inter alia, of a violation of 18 U.S.C. 2252, and having suffered persona injury, to sue in a federal court of law and recover damages.
[776] 18 U.S.C. 2252 makes a crime, inter alia, to produce, receive or distribute in interstate commerce visual depictions of minors engaged in sexually explicit conduct.
[777] Defining "child pornography" as "*any visual depiction, including any photograph, film, video, picture, or computer or computer-generated image or picture, whether made or produced by electronic, mechanical, or other means, of sexually explicit conduct, where—(A) the production of such visual depiction involves the use of a minor engaging in sexually explicit conduct; (B) such visual depiction is a digital image, computer image, or computer-generated image that is, or is indistinguishable from, that of a minor engaging in sexually explicit conduct; or (C) such visual depiction has been created, adapted, or modified to appear that an identifiable minor is engaging in sexually explicit conduct.*" The bill aims also at replacing the term "child pornography" through the U.S. Code by the term "child sexual abuse material" (CSAM).

paragraph (6), or any claim in a civil action under State law under subparagraph (C) of such paragraph (6), because the provider—

"(A) utilizes full end-to-end encrypted messaging services, device encryption, or other encryption services;

"(B) does not possess the information necessary to decrypt a communication; or

"(C) fails to take an action that would otherwise undermine the ability of the provider to offer full end-to-end encrypted messaging services, device encryption, or other encryption services."

As such, providers of interactive computer services would have to follow these best practices to retain their Section 230 immunity.

Yet another bill, the Platform Accountability and Consumer Transparency (PACT) Act,[778] introduced by Senators Brian Schatz [D-HI] and John Thune [R-SD], proposes to amend Section 230 of the CDA. The bill would introduce a mechanism system similar to the DMCA, designed to report online copyright infringement, to report illegal content or activity. The bill follows the same pattern to moderate content than followed by the German NetzGz and the French Avia, the obligation to report illegal content in the brief 24 hours delay. Providers of interactive computer services would not be protected by Section 230(c)(1) if the provider:

[778] S.4066, 116th Congress (2019-2020).

"*has knowledge of the illegal content or illegal activity*"[779]occurring of its services and "*does not remove the illegal content or stop the illegal activity within 24 hours of acquiring that knowledge, subject to reasonable exceptions based on concerns about the legitimacy of the notice*."

Providers of interactive computer services would be deemed to:

"*have knowledge" of illegal content "only if…receiv*[*ing*] *a notification*" that such content is illegal, which must include "[*i*]*dentification of the illegal content or illegal activity, and information reasonably sufficient to permit the provider to locate the content or each account involved" and "to permit the provider to contact the complaining party*."

The person complaining about a particular content would have to provide a statement, under penalty and perjury, that "*the content in the notification is accurate… and …the content or activity described in the notification has been determined by a Federal or State court to be illega*l." The Bill provides an exemption for small business providers, defined as businesses receiving than 1,000,000 monthly active users or monthly visitors and having accrued revenue of less than $25,000,000. They would only have to act "*with respect to illegal content or illegal activity…within a reasonable period of time based on the size and capacity of the provider*," nor within the 24-hour frame. Interactive computer service "*used by another interactive computer service for the management, control, or operation of that other interactive computer service, including for services such as web*

[779] The bill defines 'illegal activity' as "*activity conducted by an information content provider that has been determined by a Federal or State court to violate Federal criminal or civil law*" and defines 'illegal content' as information "*provided by an information content provider that has been determined by a Federal or State court to violate Federal criminal or civil law or State defamation law*."

hosting, domain registration, content delivery networks, caching, back-end data storage, and cloud management" would be exempt.

Senator Josh Hawley [R-MO] introduced in June 2019 yet another bill, S.1914 , the Ending Support for Internet Censorship Act, aiming at amending the CDA " *to encourage providers of interactive computer services to provide content moderation that is politically neutral.*"[780] Social media companies would have no longer be immune from liability for having removed content, unless they had obtained an "*immunity certification*" from the Federal Trade Commission that it does not moderate information provided by other information content providers in a manner that is biased against a political party, political candidate, or political viewpoint. To obtain this certification, providers would have had to prove to the FTC "*by clear and convincing evidence that* [*it*] [*did*] *not (and, during the 2-year period preceding the date on which the provider submits the application for certification, did not) moderate information provided by other information content providers in a politically biased manner*." The bill was fortunately not passed, as application would have likely triggered a multitude of First Amendment lawsuits, even though S.1914 provided a "business necessity" exception, stating that moderation practices which would be politically biased under the Act would have been so if "*necessary for business*" or if the speech had not been protected by the First Amendment and "*there is no available alternative that has a less disproportionate effect, and the provider does not act with the intent to discriminate based on political affiliation, political party, or political viewpoint*."

[780] S. 1914, 116th Cong. (2019-2020).

This law would have given the FTC unchecked power to decide what speech is politically biased, whatever this means, even as the mission of the FTC is not to censor speech.

A month after the introduction of the Ending Support for Internet Censorship Act, Representative Paul Gosar [R-AZ-4] introduced H.R.4027, the Stop the Censorship Act, which proposed to amend section 230 of the CDA "*to stop censorship, and for other purposes*."[781] It would have changed the title of Section 230 from "*Protection for private blocking and screening of offensive material*" to "*Protection for private blocking ~~and screening~~ of* ***unlawful or objectionable material***" (my emphasis). The bill would also have limited the safe harbor provisions of Section (c)(2) to the removal of "*unlawful material*," striking "*material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected'* from the law. It also would have extended the safe harbor provisions of the CDA to "*any action taken to provide users with the option to restrict access to any other material, whether or not such material is constitutionally protected*." The bill did not define what is "*unlawful or objectionable material.*" Who would have been in charge of making this decision? Would only the laws of the U.S. be taken into account, or would social media platforms be able to respond to demands from other countries, such as France, asking, for instance, to take down content denying the Holocaust? Also, as noted by Professor Eric Goldman, if platforms are only authorized to remove "unlawful material,"

[781] H.R. 4027, 116th Cong. (2019-2020).

there is a great risk for the web to be overwhelmed by trolls and spammers,[782] to the detriment of the "*marketplace of ideas*."

On May 28, 2020, President Trump signed an *Executive Order on Preventing Online Censorship*,[783] (EO) which aims at preventing "*a limited number of online platforms*[784] *to hand pick the speech that Americans may access and convey on the internet*." The EO refers to the platforms as "*the 21st century equivalent of the public square."* The EO was issued two days after Twitter added, for the first time, a civic integrity notice under two tweets posted by President Trump on May 26, 2020, alleging mail-in ballot fraud in California.[785] Twitter added a link under both tweets which read: *"! Get the facts about mail-in ballots*" and directed to a dedicated page titled "*Trump makes unsubstantiated claim that mail-in ballots will lead to voter fraud*" and listing tweets presenting counter views. As such, it can be argued that Twitter's reaction to the President's tweet was not censorship, but active participation in the marketplace of ideas. The Twitter Safety team account explained that the labels had been added "*as part of* [*Twitter's*] *efforts to enforce [its] civic integrity policy.*

---

[782] *Comments on Rep. Gosar's "Stop the Censorship Act," Another "Conservative" Attack on Section 230*, TECHNOLOGY AND MARKETING LAW BLOG, (Aug. 15, 2019), https://blog.ericgoldman.org/archives/2019/08/comments-on-rep-gosars-stop-the-censorship-act-another-conservative-attack-on-section-230.htm.

[783] Exec. Order No. 13925: Preventing Online Censorship, 85 Fed. Reg. 34,079 (June 2, 2020)(E.O. 13925), available at https://www.whitehouse.gov/presidential-actions/executive-order-preventing-online-censorship.

[784] Hint: the Executive Order refers to Twitter, Instagram, YouTube and Facebook…

[785] One tweet read: "*There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forged & even illegally printed out & fraudulently signed. The Governor of California is sending Ballots to millions of people, anyone....*", the other "*.... living in the state, no matter who they are or how they got there, will get one. That will be followed up with professionals telling all of these people, many of whom have never even thought of voting before, how, and for whom, to vote. This will be a Rigged Election. No way!*", see @realDonaldTrump, Twitter (May 26, 2020, 8:17 AM), https://twitter.com/realDonaldTrump/status/1265255835124539392 and @realDonaldTrump, Twitter (May 26, 2020, 8:17 AM), https://twitter.com/realDonaldTrump/status/1265255845358645254.

We believe those Tweets could confuse voters about what they need to do to receive a ballot and participate in the election process."[786]

We saw that Section 230 (c)(2)(A) allows social media platforms to take down speech by excluding them from civil liability for "*any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected*." The EO argues, however, that:

> "[*i*]t *is the policy of the United States to ensure that, to the maximum extent permissible under the law, this provision is not distorted to provide liability protection for online platforms that — far from acting in "good faith" to remove objectionable content — instead engage in deceptive or pretextual actions (often contrary to their stated terms of service) to stifle viewpoints with which they disagree.*

The EO seemingly argued that social media platforms should no longer be protected by the CDA safe harbor provisions, as they are now "*content creators*":

> "*In a country that has long cherished the freedom of expression, we cannot allow a limited number of online platforms to hand pick the speech that Americans may access and convey on the internet. This practice is fundamentally un-American and anti-democratic. When large, powerful social media companies censor opinions with which*

[786] @TwitterSafety, (May 27, 2020, 10:54 PM), https://twitter.com/TwitterSafety/status/1265838824451694597.

they disagree, they exercise a dangerous power. They cease functioning as passive bulletin boards, and ought to be viewed and treated as content creators."

The EO further argued that the platforms are engaging in "*deceptive or pretextual actions stifling free and open debate by censoring certain viewpoints*." This view is commonly expressed by conservatives, believing that the media favors the publication of liberal views, from the mundane, such as choice of footwear,[787] to general bias attributed to big tech companies,[788] and this view is shared by many in the U.S., most of them conservatives.[789] The White House even launched on May 15,2019 its Tech Bias Reporting Tool, asking users to report instances of suspensions of their social media accounts banned, or fraudulently reported by social media platforms "*for unclear 'violations' of user policies*."[790] Numerous U.S. conservatives believe that platforms are exercising a bias against conservative views.[791] For instance, when U.S. Representative Devin Nunes filed in 2019 a defamation

---

[787] Tristan Justice, *Media: Timberland Boots Only Look Fabulous On Democrats,* THE FEDERALIST, (Sept. 17, 2020), https://thefederalist.com/2020/09/17/media-timberland-boots-only-look-fabulous-on-democrats.
[788] Chris Mills Rodrigo, *Jordan confronts tech CEOs over claims of anti-conservative bias*, THE HILL, (July, 29 2020 02:36 PM EDT), https://thehill.com/homenews/house/509619-jordan-confronts-tech-ceos-with-claims-of-anti-conservative-bias.
[789] Emily A. Vogels, Andrew Perrin and Monica Anderson, *Most Americans Think Social Media Sites Censor Political Viewpoints*, (Aug. 19, 2020), https://www.pewresearch.org/internet/2020/08/19/most-americans-think-social-media-sites-censor-political-viewpoints. According to a June 2020 Pew Research Center survey 69% of Republicans and Republican leaners believe major technology companies generally support liberal over conservative views, while only 25% of Democrats and Democratic leaners believed so.
[790] Emily Birnbaum, *White House launches tool for reporting social media 'bias'*, THE HILL, (May 15, 2019, 5:45 PM EDT), https://thehill.com/policy/technology/443934-white-house-launches-tool-for-reporting-social-media-bias. The platform is no longer active, and the landing page reads: "*This typeform isn't accepting new responses. SOCIAL MEDIA PLATFORMS should advance FREEDOM OF SPEECH. Yet too many Americans have seen their accounts suspended, banned, or fraudulently reported for unclear "violations" of user policies. On May 15, President Trump asked Americans to share their stories of suspected political bias. The White House received thousands of responses—thank you for lending your voice!*" https://whitehouse.typeform.com/to/Jti9QH (last visited Dec. 30, 2020).
[791] *See* for instance the testimony of Representative Jim Jordan (Ohio), Ranking Member of the House Judiciary Committee, which said during a hearing on competition in digital marketplaces: "*I'll just cut to the chase, Big Tech is out to get conservatives… That's not a hunch, that's not a suspicion, that's a fact,* " quoted in Chris Mills Rodrigo - *Jordan confronts tech CEOs over claims of anti-conservative bias*, THE HILL, (July 29, 2020, 02:36 PM EDT), https://thehill.com/homenews/house/509619-jordan-confronts-tech-ceos-with-claims-of-

suit against Twitter and the alleged authors of two parody accounts mocking him,[792] he described Twitter as an "*information content provider*" who "*creates and develops content… by explicit censorship of viewpoints with which it disagrees… , by shadow banning conservatives, such as Plaintiff,… by completely ignoring lawful complaints about offensive content and by allowing that content to be accessible to the public*." The complaint was dismissed in June 2020 and Representative Nunes then started promoting the Parler platform, writing, for instance "*Parler will set you free!"*[793] Parler is a social media platform which invite people to join to "[*s]peak freely and express yourself openly, without fear of being "deplatformed" for your views. Engage with real people, not bots*."[794] Anybody can join the platform if providing a phone number,[795] but the platform is only available to registered users. It has been reported that it allows hate speech to foster, as such speech is not removed by the platform.[796] However, some users have nevertheless been banned by this platform.[797] According to the May 28, 2020 EO, "[*w*]*hen an interactive computer service provider removes or restricts access to content and its actions do not meet the criteria of subparagraph (c)(2)(A), it is engaged in editorial conduct. It is the policy of the United States*

anti-conservative-bias. The entire hearings are available at https://judiciary.house.gov/newsroom/watch-live.htm. Rep. Jordan also argued that "*no one is safe from the cancel culture*"(at 11:30 of the video).
[792] Devin Nunes v. Twitter, Inc., Elizabeth L. "Liz" Mair, Mair Strategies, LLC et al, No CL 19-1715-00.
[793] Devin Nunes (@DevinNunes), Twitter, (June 23, 2020. 10:54PM), https://twitter.com/DevinNunes/status/1275623353101856768.
[794] PARLER, https://parler.com (last visited Dec. 30, 2020).
[795] Anybody, including the person, or persons, behind the @DevinCow parody account, which tweeted at Devin Nunes (@DevinNunes), the day after the defamation suit was dismissed: "I'm at Parler! Come say hi!" with a screen capture of a Parler post of @Therealdevincow showing a cartoon cow holding a sign which read "Devin is a loser", @DevinCow, Twitter, (June 25, 2020, 12:22 AM), https://twitter.com/DevinCow/status/1276007860439138304.
[796] *See* Andrew Blake, *Parler CEO says up-and-coming social media service will not ban users for hate speech*, THE WASHINGTON POST, (Aug. 5, 2020), https://www.washingtontimes.com/news/2020/aug/5/parler-ceo-says-up-and-coming-social-media-service.
[797] Poppy Noor, *New rightwing free speech site Parler gets in a tangle over … free speech*, THE GUARDIAN, (JULY 1, 2020, 11.40 EDT), https://www.theguardian.com/technology/2020/jul/01/parler-conservative-twitter-new-free-speech-social-network

that such a provider should properly lose the limited liability shield of subparagraph (c)(2)(A) and be exposed to liability like any traditional editor and publisher that is not an online provider."

The EO directed the National Telecommunications and Information Administration (NTIA) to file a petition for rulemaking with the Federal Communications Commission (FCC) seeking clarification about Section 230 of the CDA. FCC Commissioner Brendan Carr approved the move, arguing that Congress had passed Section 230 "*to empower parents to protect their children from material on Internet sites like the then-popular Prodigy messaging board. And it acted to protect the 'good faith' steps taken by those computer services providers*" but that now "*Internet giants and social media companies … benefit from those Section 230 protections when other speakers do not.*"[798] However, FCC Commissioner Mike O'Rielly, interviewed a few days after the publication of the EO, expressed doubts about the power of the FCC to regulate Section 230 of the Communications Decency Act:

> "*I have deep reservations they provided any intentional authority for this matter, but I want to listen to people…I do not believe it is the right of the agency to read into the statute authority that is not there.*"[799]

---

[798] Carr Welcomes Executive Order On Online Censorship, FEDERAL COMMUNICATIONS COMMISSION, (May 28, 2020), https://www.fcc.gov/document/carr-welcomes-executive-order-online-censorship.

[799] Margaret Harding McGill, *FCC Republican voices doubts about Trump's executive order*, AXIOM, (June 12, 2020), https://www.axios.com/fcc-republican-doubts-trumps-executive-order-11335664-f932-43ed-8d2e-bb803c112246.html. Commissioner O'Rielly spoke about the First Amendment on July 29, 2020, two days after the NTIA had filed its Petition for Rulemaking with the FCC, as he had been invited to speak at The Media Institute's Luncheon Series: " *To be clear, the following critique is not in any way directed toward President Trump or those in the White House, who are fully within their rights to call for the review of any federal statute's application, the result of which would be subject to applicable statutory and constitutional guardrails. Rather, I am very troubled by certain opportunists elsewhere who claim to be the First Amendment's biggest heroes but only come to its defense when convenient and constantly shift its meaning to fit their current political objectives… The First Amendment protects us from limits on speech imposed by the government—not private actors—and we should all reject demands, in the name of the First Amendment, for private actors to curate or*

Pursuant to the EO, the NTIA (NTIA) filed a Petition for Rulemaking with the FCC on July 27, 2020, requesting that the Commission institute a rulemaking to interpret Section 230 of the CDA.[800] It argued that artificial intelligence and "*automated methods of textual analysis*" which were not available at the time the CDA was enacted, can now be used to flag harmful content automatically, without having to manually review each post.[801] The NTIA argued that "[*t*]*he FCC should use its authorities to clarify ambiguities in section 230 so as to make its interpretation appropriate to the current internet marketplace and provide clearer guidance to courts, platforms, and users."* It argued further that "[*n*]*ew regulations guiding the interpretation of section 230 are necessary to facilitate the provisions' interpretation in a way that best captures one of the nation's most important Constitutional freedoms."*[802] The NTIA appeared to agree with President Trump's belief that the social media platforms are biased in favor of liberal views[803] as it took the view that that the FCC has the authority to

publish speech in a certain way." The full text is available at *Remarks of FCC Commissioner Michael O'Rielly Before The Media Institute's Luncheon Series, July 29, 2020*, https://docs.fcc.gov/public/attachments/DOC-365814A1.pdf. It has been suggested that the statement led to the White House withdrawal O'Rielly's nomination for a second term, even though his nomination had been approved by the Senate Commerce Committee on July 22, see *The Making – and Unmaking – of an FCC Commissioner*, THE NATIONAL LAW REVIEW, (Aug. 10, 2020), https://www.natlawreview.com/article/making-and-unmaking-fcc-commissioner.

[800] National Telecommunications and Information Administration, Petition for Rulemaking, Docket No. RM-11862 (Filed July 27, 2020) (Petition), available at https://www.ntia.gov/files/ntia/publications/ntia_petition_for_rulemaking_7.27.20.pdf.

[801] NTIA Petition, p.4-5.

[802] NTIA Petition, p.6.

[803] *"Unfortunately, large online platforms appear to engage in selective censorship that is harming our national discourse*." NTIA Petition, p.7. The NTIA expressed its regret that *"few academic empirical studies exist of the phenomenon of social media bias*." NTIA Petition, p.8. A majority of Americans believe that social media sites are biased and censor political viewpoints, according to a Pew Research Center survey conducted in June 2020: some three-quarters of U.S. adults say it is very (37%) or somewhat (36%) likely that social media sites intentionally censor political viewpoints that they find objectionable, while only 25% of them it not to be likely. *See* Emily A. Vogels, Andrew Perrin and Monica Anderson, *Most Americans Think Social Media Sites Censor Political Viewpoints,* PEW RESEARCH CENTER, (Aug. 19, 2020), https://www.pewresearch.org/internet/2020/08/19/most-americans-think-social-media-sites-censor-political-viewpoints.

issue regulations interpreting section 230 and to show "*how regulations are necessary to resolve the statute's ambiguities that the E.O. identified*" and asked the agency to:

- "*clarify the relationship between 230(c)(1) and (c)(2);*
- *explain the meaning of "good faith" and "otherwise objectionable" in section 230(c)(2);*
- *specify how the limitation on the meaning of "interactive computer service" found in section 230(f)(2) should be read into section 230(c)(1); and,*
- *explicate the meaning of "treated as a speaker or publisher" in section 230(c)(1).*"[804]

For the NTIA:

"*Congress did not intend a vehicle to absolve internet and social media platforms[805]…from all liability for their editorial decisions*"[806] and contented that "[*i*]*n public comments,*[807] *Representative Cox explained that the section 230 would reverse Stratton Oakmont and advance the regulatory goal of allowing families greater power to control online content. The final statute reflected his stated policy: "to encourage the development of technologies which maximize user control over what information is received by individuals, families, and schools who use the Internet and other interactive computer services."*

[804] NTIA Petition, p.15.
[805] Let's note that there were no social media platforms in 1995…
[806] NTIA Petition, p.21.
[807] Citing 141 Cong. Rec. H8469-70 (daily ed. Aug. 4, 1995) (statement of Rep. Cox).

However, Representative Cox had said, during the debate cited by the NTIA, that Section 230:

> *"will protect computer Good Samaritans, online service providers, anyone who provides a front end to the Internet, let us say, who takes steps to screen indecency and offensive material for their customers. It will protect them from taking on liability such as occurred in the Prodigy case in New York that they should not face for helping us and for helping us solve this problem*."[808] Therefore, it is disingenuous to state that Section 230 was not intended as "*a vehicle to absolve internet and social media platforms… from all liability for their editorial decisions*", as stated by the NTIA.

i. Clarifying the Relationship Between 230(c)(1) and (c)(2)

The NTIA asked the FCC to "*make clear that section 230(c)(1) applies to liability directly stemming from the information provided by third-party users*," claiming that "Section 230(c)(1) does not immunize a platform's own speech, its own editorial decisions or comments, or its decisions to restrict access to content or its bar user from a platform." It asked the FCC to add "*Subpart E. Interpreting Subsection 230(c)(1) and Its Interaction With Subsection 230(c)(2) to 47 CFR Chapter I,*" redacted as such:

"(a) 47 U.S.C. 230(c)(1) applies to an interactive computer service for claims arising from failure to remove information provided by another information content provider. Section 230(c)(1) has no application to any interactive computer service's decision, agreement, or action to restrict access to or availability of material provided by another

[808] 141 Cong. Rec. H8469-70 (daily ed. Aug. 4, 1995) (statement of Rep. Cox).

information content provider or to bar any information content provider from using an interactive computer service. Any applicable immunity for matters described in the immediately preceding sentence shall be provided solely by 47 U.S.C. § 230(c)(2).

(b) An interactive computer service is not a publisher or speaker of information provided by another information content provider solely on account of actions voluntarily taken in good faith to restrict access to or availability of specific material in accordance with subsection (c)(2)(A) or consistent with its terms of service or use."

As such, Section 230(c)(1) would only protect a social media platforms and other interactive computer service providers from being considered a publisher or speaker if they remove "*obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable speech*."

ii. The Meaning of Section 230(c)(2)

We saw earlier that the meaning of what is *"otherwise objectionable"* is a crucial element of the law, a rare example on the vagueness of a term protecting the expanse of the First Amendment, and the way the notion is interpreted may or may not ensure that Section 230 may still be used by platforms to regulate users' speech. The NTIA argued that "[*i*]*f* "*otherwise objectionable" means any material that any platform "considers" objectionable, then section 230(b)(2) offers* ***de facto*** *immunity to all decisions to censor content*" (emphasis in the text). This demand is made to prevent social media platforms to exercise an allegedly liberal views-biased control over speech, and the NTIA argued that

"*section 230(c)(2) only applies to obscene, violent, or other disturbing matters*."[809] The NTIA asked the FCC to add this definition of "other objectionable" to Section 230, a meaning "*any material that is similar in type to obscene, lewd, lascivious, filthy, excessively violent, or harassing materials*."

The NTIA also argued that the phrase "good faith" in section 230(c) is "*ambiguous*," citing the often-quoted dissent of Judge Fisher in *Zango, Inc. v. Kaspersky Lab, Inc*., who warned that "[*t*]*he risks inheres in the disjunctive language of the statute — which permits blocking of "material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected,*"[810] a view recognized as valid in 2019 by the Ninth Circuit, in *Enigma*, holding that "i*nterpreting* [*Section 230*] *to give providers unbridled discretion to block online content would, as Judge Fisher warned, [would] enable and potentially motivate internet-service providers to act for their own, and not the public, benefit*."[811]

---

[809] Citing Darnaa, LLC v. Google, Inc., 2016 WL 6540452 at *8 (N.D. Cal. 2016) ("*The context of § 230(c)(2) appears to limit the term to that which the provider or user considers sexually offensive, violent, or harassing in content.*"), and Song fi Inc. v. Google, Inc., 108 F. Supp. 3d 876,883 (N.D. California 2015), which noted, when interpreting Section 230(c)(2) that "*when a statute provides a list of examples followed by a catchall term (or "residual clause") like "otherwise objectionable," the preceding list provides a clue as to what the drafters intended the catchall provision to mean*." YouTube had argued unsuccessfully in this case that an artificially inflated view count associated with a video uploaded on YouTube was "*objectionable content*." The Court argued further that *"both the context in which "otherwise objectionable" appears in the Communications Decency Act and the history and purpose of the Act support this reading. Section 230 is captioned "Protection for `Good Samaritan' blocking and screening of offensive material," yet another indication that Congress was focused on potentially offensive materials, not simply any materials undesirable to a content provider or user*. Song fi, at 883.

[810] Zango, Inc. v. Kaspersky Lab, Inc., 568 F. 3d 1169, 1178 (9th Circuit 2009) (Fisher, J., concurring): "…under the generous coverage of § 230(c)(2)(B)'s immunity language, a blocking software provider might *abuse that immunity to block content for anticompetitive purposes or merely at its malicious whim, under the cover of considering such material "otherwise objectionable*."

[811] Enigma Software Group USA, LLC v. Malwarebytes, 946 F. 3d 1040, 1051 (9th Circ. 2019).

The NTIA proposed an extensive definition of "good faith" when asking the FCC to modify section 230. It would be defined as such:

"(e) A platform restricts access to or availability of specific material (including, without limitation, its scope or reach) by itself, any agent, or any unrelated party in "good faith" under 47 U.S.C. § (c)(2)(A) if it:

i. *restricts access to or availability of material or bars or refuses service to any person consistent with publicly available terms of service or use that state plainly and with particularity the criteria the interactive computer service employs in its content-moderation practices, including by any partially or fully automated processes, and that are in effect on the date such content is first posted;*
ii. *has an objectively reasonable belief that the material falls within one of the listed categories set forth in 47 U.S.C. § 230(c)(2)(A);*
iii. *does not restrict access to or availability of material on deceptive or pretextual grounds, and does not apply its terms of service or use to restrict access to or availability of material that is similarly situated to material that the interactive computer service intentionally declines to restrict; and*
iv. *supplies the interactive computer service of the material with timely notice describing with particularity the interactive computer service's reasonable factual basis for the restriction of access and a meaningful opportunity to respond, unless the interactive computer service has an*

objectively reasonable belief that the content is related to criminal activity or such notice would risk imminent physical harm to others."

If adopted, this definition would significantly limit the power of the platform to take down material and such power would likely be strictly limited within the confines of obscene or violent speech. Its interpretation would also certainly create a wealth of interpretation, possible a Circuit split necessitating a Supreme Court holding.

iii. What About "Information Content Providers"?

"Information content providers," that, is, the persons or entities responsible, *"in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service*"[812] are not immune under the CDA. The Ninth Circuit, in *Fair Housing Council of San Fernando Valley v. Roommates.com*, interpreted '*development'*, withing the meaning of Section 230, as "*materially contributing*" to the material at stake's "*alleged unlawfulness*", explaining that "*a website helps to develop unlawful content, and thus falls within the exception to section 230, if it contributes materially to the alleged illegality of the conduct*."[813] While the NTIA cited the case, it found that the definition offered by the Ninth Circuit "*has failed to provide clear guidance, with courts struggling to define "material contribution*,"[814] citing a case where the operator of a "revenge porn" site and of a "reputation" site[815] had been found guilty of extortion unlawful

[812] 47 U.S.C. § 230(f)(3).
[813] Fair Housing Council of San Fernando Valley, 521 F.3d at 1166.
[814] NTIA Petition, p.40.
[815] The victim of the "porn revenge" site could not get their personal information taken down unless making such demand via the reputation site and paying a fee.

use of personal identifying information under California law.[816] He appealed, claiming that he was immunized from liability as an "interactive computer service" or "access software provider" within the meaning Section 230, and the jury had been instructed at his request that "[*a*]*n interactive computer service or access software provider can become an `information content provider' to the extent a website is designed to require users to post illegal or actionable content as a condition of use.*" The California Court of Appeals did not find him entitled to benefit from Section 230 immunity, as users of the <UGotPosted.com> "revenge porn" website had:

> "*to answer a series of questions with the damaging content in order to create an account and post photographs. That content — full names, locations, and Facebook links, as well as the nude photographs themselves — exposed the victims' personal identifying information and violated their privacy rights" and that the website was "designed to solicit… content that was unlawful, demonstrating that Bollaert's actions* ***were not neutral****, but rather materially contributed to the illegality of the content and the privacy invasions suffered by the victims. In that way, he developed in part the content, taking him outside the scope of CDA immunity*"[817] (my emphasis).

In this case, the court seemed to imply that neutrality is necessary to be granted immunity. In our current discussion, this is not enough as President Trump's EO started the debate claiming social media platforms are biased. The NTIA asked the FCC to add to Section 230 a definition of what is" *creation or development of information" as*

[816] People v. Bollaert, 248 Cal. App. 4th 699, 717 (California Court of Appeal, 4th Appellate Dist., 1st Div. 2016).
[817] Bollaert, at 721,

"substantively contributing to, modifying, altering, presenting or prioritizing with a reasonably discernible viewpoint, commenting upon, or editorializing about content provided by another information content provider."[818]

On August 3, 2020, FCC Chairman Ajit Pai announced that the agency would seek public comments about the Petition.[819] On September 17, 2020, the NTIA filed reply comments on its petition addressed to the FCC.[820] It acknowledged that "*many comments*" had claimed that the FCC did not have the authority "*to prescribe implementing regulations under section 230*, *"* but argued that it does have this authority, quoting Justice Scalia's in *City of Arlington v. FCC*[821] who wrote that Section 201(b) of the Communications Act of 1934 empowers the FCC to "*prescribe such rules and regulations as may be necessary in the public interest to carry out* [its] *provisions*,",] adding "[*o*]*f course, that rulemaking authority extends to the subsequently added portions of the Act*." As the Telecommunications Act of 1996 was incorporated in the Communications Act of 1934, and incorporated Section 230 into the Communications Act of 1934, and as the FCC's section 201(b) rulemaking "*extends to the subsequently added portions of the Act*" as held by the Supreme Court, the NTIA concluded that the FCC had the authority to issue regulations implementing Section 230.[822]

---

[818] NTIA Petition, p.42.
[819] *Chairman Pai on Seeking Public Comment on NTIA's Sec. 230 Petition*, FEDERAL COMMUNICATIONS COMMISSION, (Aug. 3, 2020), https://www.fcc.gov/document/chairman-pai-seeking-public-comment-ntias-sec-230-petition.
[820] National Telecommunications and Information Administration, Reply Comments, Docket No. RM- 11862 (Filed Sept. 17, 2020) (Reply Comments), available at https://www.ntia.gov/files/ntia/publications/ntia_reply_comments_in_rm_no._11862.pdf.
[821] City of Arlington, Tex. v. FCC, 133 S. Ct. 1863, 1866 (2013).
[822] Non-profit organization Public Knowledge had argued in its filed comments about the Petition that Congress has not delegated an explicit nor an implicit authority over Section 230 to the FCC. The comments are available at
Shiva Stella, *Public Knowledge Urges FCC to Reject Unlawful Trump Administration Request to Rewrite Section 230*, PUBLIC KNOWLEDGE (Sept. 2, 2020), https://www.publicknowledge.org/press-release/public-knowledge-urges-fcc-to-reject-unlawful-trump-administration-request-to-rewrite-section-230.

On September 23, 023, the U.S. Department of Justice (DoJ) issued recommendations to Congress to amend Section 230,[823] which including a copy of Section 230 law with the DoJ's proposed changes in redline. It proposed to add to take out of the scope of Section 230 (c)(1) providers or users of an interactive computer service who have agreed, decided or act to *"restrict access to or availability of material provided by another information content provider*"[824] They could only claim immunity as provider or user of an interactive computer service under Section 230 (c)(2). A provider or user of an interactive computer service would not be considered publisher or speaker "*for all other information on its service provided by another information content provider solely on account of actions voluntarily taken in good faith to restrict access to or availability of specific material that the provider or user has an objective reasonable belief violates its terms of service or use*."[825] Therefore, a social media platform taking down speech, or restricting access to it, would have to do so (1) in good faith, (2) if it has an objective and reasonable belief that (3) the speech violates its terms of service. Social media platforms have comprehensive terms of services or use, which are their private laws. It seems thus that taking down speech or restricting access to speech pursuant to this private law would be done because of an objective and reasonable belief that these rules have been violated, carried out in good faith. Some speech blatantly violates terms of use, such as speech advocating terrorism[826]

---

[823] *Department of Justice's Review of Section 230 of the Communications Decency Act of 1996*, DEPARTMENT OF JUSTICE (Sept. 23, 2020), https://www.justice.gov/ag/department-justice-s-review-section-230-communications-decency-act-1996?utm_medium=email&utm_source=govdelivery.
[824] Proposed Section 230 (c)(1)(B).
[825] Proposed Section 230 (c)(1)(C).
[826] *See* Anton Battesti, *Lutte contre les contenus haineux : l'exemple de Facebook*, LÉGIPRESSE HORS SÉRIE No. 63 (2020-1), 33. Mr. Battesti is head of the public affairs for Facebook France and explained that 99% of terrorism-related speech were proactively taken down, without having to rely on users reporting it. He also noted that, while "*there is no need to assess the context*" of pedo-pornographic speech, it is more delicate to assess whether a speech signaled as being hate speech is indeed hateful, in bad taste, or "*a bad joke*." He

or pedo-pornography, and the platforms have developed technologies allowing for their automatic detection.

The DOJ proposed a new standard for the "good Samaritan" immunity standard. Section 230 (c)(2) (A) protects now providers or users of an interactive computer service from civil liability if taking speech down voluntarily and in good faith, if considering it to be "*obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable*." The DoJ proposed to strike down "*if considering it to be*" and replace the terms by "*has an objectively reasonable belief is.*" The standard would thus move from a subjective one, a "consideration," which is defined as a "*continuous and careful thought*"[827] to an "*objectively reasonable belief.*"

The "objective reasonableness" standard is a multi-factor analysis allowing courts to determine whether excessive force has been applied by law enforcements when making an arrest, investigatory stop, or any other seizures within the meaning of the Fourth Amendment, which protects the people "*against unreasonable searches and seizures*." It is therefore a standard used to protect against governmental conduct, not to be protected against private parties. The Supreme Court held in 1989, in *Graham v. Connor*,[828] that "*all claims that law enforcement officers have used excessive force — deadly or not — in the course of an arrest, investigatory stop, or other "seizure" of a free citizen should be analyzed*

---

explained that 60% to 70 % of hate speech has been detected proactively, that is, human beings made the decision to take them down or to leave them published. He further explained how machines can moderate speech, by using machine learning and having a machine assign a score to a speech, allowing it to decide that a particular speech is likely to be hateful, as it resembles highly speech formerly published and deemed hateful.

[827] *Consideration*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/consideration, (last visited Dec. 30, 2020).

[828] Graham v. Connor, 490 U.S. 386, 395 (1989).

under the Fourth Amendment and its "reasonableness" standard." In this case, the District Court had applied as "objective reasonableness" standard test : (1) *the need for the application of force; (2) the relationship between that need and the amount of force that was used; (3) the extent of the injury inflicted; and (4)* "[*w*]*hether the force was applied in a good faith effort to maintain and restore discipline or maliciously and sadistically for the very purpose of causing harm*."[829] Therefore, only actions deemed malicious and sadistic, performed "*for the very purpose of causing harm*, were deemed not objectively reasonable when applying force. This provided law enforcement the power to search and seize at leisure, if their actions were not sadistic and malicious… The Fourth Circuit Court of Appeals had affirmed, even endorsing the test "*as generally applicable to all claims of "constitutionally excessive force" brought against governmental officials*."[830] The Supreme Court reversed, holding that "the "*malicious and sadistic" factor puts in issue the subjective motivations of the individual officers, which our prior cases make clear has no bearing on whether a particular seizure is "unreasonable" under the Fourth Amendment*."

The Supreme Court had explained in *Terry v. Ohio* that test "*for determining reasonableness…* [*is*]*balancing the need to search [or seize] against the invasion which the search* [*or seizure*] *entails*."[831] Does the DOJ suggest that the "objective reasonableness" test of Section 230 must be analyzed under the First Amendment? If so, the decision to take down speech would be reasonable only if the "invasion" that the take down entails is balanced. As such, the right of the speaker would have to be taken in consideration. It may

[829] Graham at 390.
[830] Graham at 391.
[831] Terry v. Ohio, 392 US 1, 21 (1968), citing Camara v. Municipal Court, 387 U. S. 523, 534-535, 536-537 (1967).

very well have been the goal of the DOJ, as its aim when suggesting changes to Section 230 was likely to protect President Trump against further reining of his speech on social media.

The scope of what is considered offensive speech under Section (c)(2)(A) is also modified by the DOJ redline suggestions. While users and providers of an interactive computer service would keep the right to block and screen "*obscene, lewd, lascivious, filthy, excessively violent, harassing*" speech, they would also be provided immunity if blocking and screening speech "*promoting terrorism or violent extremism*" or "*promoting self-harm*." However, the catch-all provision providing them good Samaritan protection for blocking or screening "*otherwise objectionable*" speech, is taken down, and replaced by a protection for blocking and screening "*unlawful*" speech. The standard would thus no longer be subjective and expansive ("*objective*") but objective and restricted by the scope of laws ("*unlawful*.")

The DOJ proposes further to "carve out" the good Samaritan protection of Section 230(c)(1) if the provider of an interactive computer service[832] is criminally prosecuted under federal or state law, or if an state or federal civil action has been brought against him, her, or it,[833] "*if, at the time of the facts giving rise to the prosecution or action, the service provider acted purposefully with the conscious object to promote, solicit, or facilitate material or activity by another information content provider that the service provider knew*

[832] Not the user of an interactive computer service.
[833] It is not surprising that federal civil actions are mentioned, as Attorney General William Barr said in his opening remarks at the DOJ workshop on Section 230 on February 19, 2020, that "*civil tort law can act as an important complement to our law enforcement efforts. Federal criminal prosecution is a powerful, but necessarily limited tool that addresses only the most serious conduct. The threat of civil liability, however, can create industry-wide pressure and incentives to promote safer environments.*" See *Attorney General William P. Barr Delivers Opening Remarks at the DOJ Workshop on Section 230: Nurturing Innovation or Fostering Unaccountability*, DEPARTMENT OF JUSTICE, (Feb. 19, 2020), https://www.justice.gov/opa/speech/attorney-general-william-p-barr-delivers-opening-remarks-doj-workshop-section-230 (last visited Dec. 30, 2020).

or had reason to believe would violate Federal criminal law, if knowingly disseminated or engaged in."[834] As such, the promotion of material posted by a third party, if reasonably believed to violated Federal law, would make an interactive computer service a "bad Samaritan."

President Trump threatened in December 2020 to veto the annual defense policy bill if Congress did not agree to repel the CDA,[835] thus showing his personal animosity towards a law viewed as allowing platforms to check his speech. While Donald Trump was not reelected in 2019, it does not necessarily mean that the CDA is safe, as then candidate Joe Biden said in an interview that he was in favor of repelling Section 230.[836]

b. Overview of European Law

Members States of the European Union cannot impose general obligations on intermediaries, which include hosting service providers.[837] The European Court of Justice

---

[834] This would be the news Section 230 (1) (D), "Exclusion from "Good Samaritan" Immunity. "Bad Samaritan Carve Out." The DoJ appears to take the position that Section 230 has a sinister side, because it can be used to shield criminals, among them the vilest, from liability. See *Deputy Attorney General Jeffrey A. Rosen Speaks at the Free State Foundation's 12th Annual Telecom Policy Conference*, DEPARTMENT OF JUSTICE, (March 10, 2020), https://www.justice.gov/opa/speech/deputy-attorney-general-jeffrey-rosen-speaks-free-state-foundations-12th-annual-telecom, (last visited Dec. 30, 2020), stating that ("…*now, a quarter century after its enactment, there also is recognition that Section 230 immunity has not always been a force for good, particularly in light of some of the extraordinarily broad interpretation given to it by some courts. For example, platforms have been used to connect predators with vulnerable children, to facilitate terrorist activity, and as a tool for extreme online harassment*.

[835] Melissa Quinn, *Trump threatens to veto defense bill unless social media shield is repealed*, CBS NEWS, (updated on Dec. 2, 2020 10:02 AM), https://www.cbsnews.com/news/trump-threatens-veto-defense-bill-section-230-repeal.

[836] Editorial Board of The New York Times, *Joe Biden*, THE NEW YORK TIMES, (Jan. 17, 2020), https://www.nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html?smid=nytcore-ios-share : "*It should be revoked because it is not merely an internet company. It is propagating falsehoods they know to be false, and we should be setting standards not unlike the Europeans are doing relative to privacy. … There is no editorial impact at all on Facebook. None. None whatsoever. It's irresponsible. It's totally irresponsible*."

[837] *See* Kletia Noti, *Injunctions and Article 15(I) of the E-Commerce Directive: The Pending Glawischnig-Piesczek v. Facebook Ireland Limited Preliminary Ruling*, TTLF Newsletter on Transatlantic Antitrust and IPR Developments, Stanford-Vienna Transatlantic Technology Law Forum, (Nov.21, 2018),

held in its *Sabam* case[838] that an online social networking platform which stores on its servers information provided by the users of the platform, relating to their profile, is a hosting service provider within the meaning of Article 14 of Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services (the e-Commerce Directive.)[839] Recital 18 of the e-Commerce Directive specifies that information society services "*span a wide range of economic activities which take place on-line*" and can consist, for example, of selling goods on-line, delivering goods or providing services, offering on-line information or commercial communications, even for free, or providing data searching tools. As such, social media platforms are information services providers within the meaning of the e-Commerce Directive. However, the European Parliament noted in a recent report that:

> "*it remains unclear to what extent the new type of online services, such as social media companies that have appeared since the adoption of the E-commerce Directive, fall within the definition of 'information society services' providers that can benefit from the liability exemption*."[840]

Section 4 of the e-Commerce Directive is comprised of articles 12 to 15 and deals with liability of intermediary service providers and states when they can, or cannot, be held

https://ttlfnews.wordpress.com/2018/11/21/injunctions-and-article-15i-of-the-e-commerce-directive-the-pending-glawischnig-piesczek-v-facebook-ireland-limited-preliminary-ruling.

[838] C-360/10, EU:C:2012:85, paragraph 27.

[839] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, 2000 O.J. (L 178), 1,16.

[840] *Reform of the EU liability regime for online intermediaries - Background on the forthcoming digital services act*, EUROPEAN PARLIAMENT, (May 2020), p.2, https://www.europarl.europa.eu/RegData/etudes/IDAN/2020/649404/EPRS_IDA(2020)649404_EN.pdf.

liable under applicable national law.[841] Recital 42 of the Directive explains that the information society service providers are exempted from liability only if their activities are "*of a mere technical, automatic and passive nature*," meaning that the information society service provider does not have knowledge of nor control over the information transmitted or stored.

Article 12 of the Directive grants a "mere conduit" exemption to information society service providers providing services of "*transmission in a communication network of information provided by a recipient of the service, or* [*providing]… access to a communication network, Member States shall ensure that the service provider is not liable for the information transmitted, on condition that the provider*:

(a) does not initiate the transmission;

(b) does not select the receiver of the transmission; and

(c) does not select or modify the information contained in the transmission."

Recital 43 of the e-Commerce Directive specifies that "[*a*] *service provider can benefit from the exemptions for "mere conduit" …when he is in no way involved with the information transmitted; this requires among other things that he does not modify the information that he transmits; this requirement does not cover manipulations of a technical nature which take*

[841] The European Court of Justice specified in Google France and Google, that "*Section 4 of* [*the e-Commerce Directive*], *comprising Articles 12 to 15 and entitled 'Liability of intermediary service providers', seeks to restrict the situations in which intermediary service providers may be held liable pursuant to the applicable national law. It is therefore in the context of that national law that the conditions under which such liability arises must be sought, it being understood, however, that, by virtue of Section 4 of that directive, certain situations cannot give rise to liability on the part of intermediary service providers*." Joined Cases C-236/08 to C-238/08, Google France and Google, paragraph 117.

place in the course of the transmission as they do not alter the integrity of the information contained in the transmission." A Member State's court or administrative authority court or administrative authority may however require the service provider "*to terminate or prevent an infringement.*"[842]

However, under article 14 of the e-Commerce Directive, an hosting provider, that is, a provider of an "*information society service…consist*[*ing*] *of the storage of information provided by a recipient of the service,*" is not liable "*for the information stored at the request of a recipient of the service*" if it "*does not have actual knowledge of illegal activity or information and … upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information.*" As such, an immunity is provided to an information service provider if (1) it has not knowledge of the illegal activity or speech and (2) it "*expeditiously*" removes such content or disable access to it as soon as it is aware of it. The European Court of Justice (ECJ) stated that an internet service provider is within the scope of article 14 of the e-Commerce Directive only if its conduct is limited to that of an intermediary provider within the meaning of Section 4 of the e-Commerce Directive.[843] However, it is not the case if the service provider does not "*confin*[*e*] *itself to providing that service neutrally by a merely technical and automatic processing of the data provided by its customers*, [*but*] *plays an active role of such a kind as to give it knowledge of, or control over, those data*."[844] Courts must thus examine whether the role played by a service provider is

---

[842] Article 12(3) of the e-Commerce Directive.
[843] Joined Cases C-236/08 to C-238/08, Google France and Google, paragraph 112. Section 4 of the e-Commerce Directive comprises of articles 12 to 15 and is entitled 'Liability of intermediary service providers.'
[844]Case C-324/09, L'Oréal v. eBay, paragraph 113, citing Joined Cases C-236/08 to C-238/08 Google France and Google, paragraphs 114 and 120.

merely technical, automatic and passive, thus a pledge of neutrality. Article 14, however, applies to the operator of an online marketplace having an active role providing him knowledge[845] or control of the data stored. Data can be, of course, speech. As the European Court of Justice specified that the operator of an online marketplace plays an active role if it provides assistance, which includes "*optimizing the presentation of the offers for sale… or promoting them*,"[846] one can wonder if this applies only to operators of online marketplaces, such as eBay, or if a social media platform may lose its immunity under Article 14 of the e-Commerce Directive if, more than merely hosting data, including speech created by its users, it also optimizes its presentation and promotes it.[847]

Article 15 of the e-Commerce Directive states that Member States cannot impose on providers a general obligation to monitor information transmitted or stored when providing the services covered by Articles 12, 13 and 14 of the Directive. Article 15 §1 of the directive provides that information society service providers do not have a general obligation to monitor the information they transmit or store, nor do they have a general obligation *"actively to seek facts or circumstances indicating illegal activity*." Article 15 §2, however, gives Member States the right to enact legislation obliging them to "*promptly… inform the competent public authorities of alleged illegal activities undertaken or information provided by recipients of their service*" or to oblige them "*to communicate to the competent*

[845] Case C-324/09, L'Oréal v. eBay, paragraph 119 specified that a provider cannot be exempt from liability if it had been 'aware of facts and circumstances from which the illegal activity or information is apparent' or , if it has "*obtained such knowledge or awareness, it had acted expeditiously to remove, or disable access to, the information*."
[846] Case C-324/09, L'Oréal v. eBay, paragraph 123.
[847] For instance, article L. 111-7-I-1° of the French consumers Code defines an online platform operator as *"any natural or legal person offering, in a professional capacity, against payment or free of charge an online communication service to the public based on…* [*t*]*he classification or referencing, by means of computer algorithms, of content, goods or services offered or put online by third parties*."

authorities, at their request, information enabling the identification of recipients of their service with whom they have storage agreements." Also, recital 47 of the Directive states that the prohibition of general monitoring laid out by article 15 does not concern monitoring in a specific case: while generally monitoring cannot be imposed, specific monitoring, within the confine of a particular issue, such as defamation or infringement, can be required from providers. In the *L'Oréal and Others* case, the European Court of Justice held that the operator of an online marketplace can be ordered by a court[848] to make measures aiming at ending an infringement and at preventing "*further infringements of that kind*." It thus appears than the operator cannot engage in a 'fishing expedition', whose scope would be too broad. It also appears that the measures must be confined to a particular type of infringement of the same nature, perpetrated by the same person.[849]

The European Commission explained in its Recommendation on measures to effectively tackle illegal content online,[850] that, following the *L'Oréal v. eBay* case, a mechanism to submit notices must be sufficiently precise "*to trigger actual knowledge or awareness for the purposes of Article 14of the e-Commerce Directive*" and stated further that these mechanisms:

[848] The European Court of Justice specified that the court's injunction 'must be effective, proportionate, dissuasive and must not create barriers to legitimate trade." Case C-324/09, L'Oréal v. eBay, paragraph 144.
[849] *See* Opinion of Advocate General Szpunar, June 4, 2019, Case C-18/18, Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 68: "*…it is apparent from the judgment in L'Oréal and Others that an information society service provider may be ordered to take measures that help to prevent* ***new infringements*** *the* ***same kind*** *of the same rights* "(emphasis in the original). He added however, at paragraph 72, that, in his opinion, "*a host provider may be ordered to identify information equivalent to that characterized as illegal and originating from the same user*."
[850] Commission Recommendation of 1.3.2018 on measures to effectively tackle illegal content online, EUROPEAN COMMISSION (March 1, 2018), C(2018) 1177 final., Point 6.

"*should allow for and encourage the submission of notices which are sufficiently precise and adequately substantiated to enable the hosting provider concerned to take an informed and diligent decision in respect of the content to which the notice relates, in particular whether or not that content is to be considered illegal content and is be removed or access thereto is be disabled. Those mechanisms should be such as to facilitate the provision of notices that contain an explanation of the reasons why the notice provider considers that content to be illegal content and a clear indication of the location of that content*."

The e-Commerce Directive has been implemented by the Member States, for instance, in France, by the LCEN which Article 6 provides that hosting providers and ISPs, which include social media platforms, must fight online apology for crimes against humanity, incitement to racist hatred and child pornography, by providing an "*easily accessible and visible*" mechanism allowing any person to bring this type of content to their attention (article 6-7 of the LCEN). They also must "*promptly*" inform competent public authorities of any of such illicit content which are reported to them and which is carried out by their users. They also make public the resources they devote to fight these illegal activities Failing to do so is punishable by one-year imprisonment and a 75,000 Euros fine (Article 6-VI-1 of the LCEN). Under article 6.7 of the LCEN, the Internet Service Providers (ISPs) "*are not subject to a general obligation to monitor the information which they transmit or store, nor do they have a general obligation to seek facts or circumstances indicating illegal activity*." However, under the same article, ISPs have nevertheless the obligation to put in place "*an easily accessible and visible system so that anyone can alert them about data related to the apology of crimes against humanity, incitement to racial hate*

and child pornography. They also have the obligation "to promptly inform the competent public authorities of all [such] illegal *activities*."

Social media companies are considered by French law to be online platforms. Article L.111-7- I of the French Consumer Code defines an online platform operator as:

> "*any natural or legal person offering, in a professional capacity, paid or unpaid, an online communication service to the public based on:*
>
> *1 ° The classification or referencing, by means of computer algorithms, of content, goods or services offered or put online by third parties;*
>
> *2 ° Or the bringing together of several parties with a view to the sale of a good, the supply of a service or the exchange or sharing of content, a good or a service*."

We saw earlier during our discussion on fake news that article L.111-7- II of the French Consumer Code requires online platform operators "*to provide consumers with fair, clear and transparent information on:*

1 ° The general conditions of use of the intermediation service that it offers and the terms of referencing, classification and de-referencing of the content, goods or services to which this service provides access;

2 ° The existence of a contractual relationship, of a capital link or of remuneration for its benefit, since they influence the classification or referencing of the content, goods or services offered or posted online;

3 ° The quality of the advertiser and the rights and obligations of the parties in civil and fiscal matter, when consumers are put in contact with professionals or non-professionals."

Whether intermediaries have or not a duty to monitor content is of interest for the European Court of Human Rights. The facts which led to the Delfi AS v. Estonia ECtHR case are as follow. Delfi was the largest news portal on the Internet in Estonia. It did not edit nor moderate comments, and some 10,000 comments were posted every day by readers, mostly under a pseudonym. Some of these comments were defamatory, some were insulting. Readers had the opportunity to mark a comment as insulting, and Delfi would then take it down, following a notice and take down system voluntarily put in place by the news portal. Messages containing obscene words were automatically deleted.

In January 2006, Delfi published an article about the alleged destruction of ice roads by the Saaremaa Shipping Company, which provides a public ferry transport service between the Estonian mainland and some islands. Ice roads are public roads, open between the mainland and some islands in the winter, when the water is frozen, and can be used for free, unlike the commercial ferry. Many comments published by readers under the article threatened and insulted L., the company's sole majority shareholder. Delfi took down the comments at the demand of L's attorney, but L. nevertheless filed a suit against Delfi. The claim was first dismissed, as the court found that Delfi's liability was excluded by the e-Commerce Directive, and that Delfi was not the publisher of the comments. The Court of appeals reversed and remanded, and the case was eventually heard by the Estonian Supreme Court, which held in June 2009 that Delfi could not be protected by the e-

Commerce Directive's "mere conduit" exemption and that Delfi was liable under the Estonian Obligations Act, because it should have prevented the publication of clearly unlawful comments.

The Estonian Supreme Court reasoned that:

> "[*t*]*he objective of* [*Delfi*] *is not merely the provision of an intermediary service*. [*Delfi*] *has integrated the comments section into its news portal, inviting visitors to the website to complement the news with their own judgments… and opinions …. In the comments section*, [*Delfi*] *actively calls for comments on the news items appearing on the portal. The number of visits to* [*Delfi*]*'s portal depends on the number of comments; the revenue earned from advertisements published on the portal, in turn, depends on the* [*number of visits*]. *Thus, [Delfi] has an economic interest in the posting of comments*."[851]

Delfi then changed its content moderation practices. Readers who had posted offensive comments were asked to read and accept Delfi's rules[852] on posting comments (the Rules) before being authorized to post new comments, and Delfi set up a team of moderators to moderate its site and to monitor user's compliance with the Rules. Delfi then

---

[851] Delfi SA v. Estonia, Grand Chamber, § 13.

[852] They stated, inter alia, that "The Delfi message board is a technical medium allowing users to publish comments. Delfi does not edit *comments. An author of a comment is liable for his/her comment. It is worth noting that there have been cases in the Estonian courts where authors have been punished for the contents of a comment ... Delfi prohibits comments the content of which does not comply with good practice.These are comments that:*
- contain threats;
- contain insults;
- incite hostility and violence;
- incite illegal activities ...
- contain obscene expressions and vulgarities ...
Delfi has the right to remove such comments and restrict their authors' access to the writing of comments ..."

applied the case to the ECtHR, claiming that Article 10 of the ECHR had been violated and that an intermediary was not considered a publisher of content. Delfi also argued that the interference with its freedom of speech was not necessary in a democratic society, and that it could only either hire "*an army of highly trained moderators to patrol (in real time) each message board (for each news article) to screen any message that could be labelled defamatory (or that could infringe intellectual property rights, inter alia*" who would remove every day sensitive comments and moderate discussions so that they are "*limited to the least controversial issues,*" or decide to shut down the forum of comments entirely. Either way, readers would no longer be provided the opportunity "*to comment freely on daily news and assume responsibility independently for their own comments*."[853] As such, according to Delfi, the decision of the Estonian Supreme Court had a chilling effect on speech.[854] The news portal also argued that the comments were directed at the decision of the Saaremaa company about the ice road, not at the article itself, which had addressed an important issue. It had taken sufficient measures to prevent or remove defamatory comments, even doing so the day it had been notified of them. For Delfi, it is the authors of the comments who should bear responsibility for them, not the news portal. The Estonian Government argued that the interference with Delfi's freedom of expression had been prescribed by law, and the interference had for legitimate aim to protect the reputation of others, and that the interference is necessary in a democratic society. The First Section of the ECtHR held on October 10, 2013, in Delfi AS v. Estonia,[855] there had been no violation of article 10 of the

[853] Delfi SA v. Estonia, Grand Chamber, § 72.
[854] Delfi SA v. Estonia, Grand Chamber, § 73.
[855] Delfi AS v. Estonia, Application no. 64569/09), (Oct. 10, 2013).

ECHR. The case was then referred to the Grand Chamber of the ECtHR, which held on June 16, 2015 that article 10 of the ECHR had not been violated.

The Grand Chamber first noted "*that user-generated expressive activity on the Internet provides an unprecedented platform for the exercise of freedom of expression*."[856] It also noted that the Estonian Supreme Court had rightfully recognized that posting comments on an online portal is a journalistic activity, but that one can reasonably not require the portal operator to edit comments before publishing them, as a printed media editor would do. Instead, online, it is the author of the comments which is the initiator of the comment, while it is the editor in the printed media environment. The ECtHR also noted that the Estonian Supreme Court had been right to assess that the comments at stake in the case had been defamatory, threatening, and generally illegal. For the Court, the case was not a case about:

> "*other fora on the Internet where third-party comments can be disseminated, for example an Internet discussion forum or a bulletin board where users can freely set out their ideas on any topic without the discussion being channeled by any input from the forum's manager; or a social media platform where the platform provider does not offer any content and where the content provider may be a private person running the website or blog as a hobby*."[857]

It thus appears that the *Delfi* cannot be invoked against a social media platform, unless it "*offer any content*."

[856] Delfi SA v. Estonia, Grand Chamber, § 110.
[857] Delfi SA v. Estonia, Grand Chamber, § 116.

The Grand Chamber further noted that it is not disputed that the comments posted by readers in reaction to the news article published in the comments section on the applicant company's Internet news portal were of a clearly unlawful nature. Indeed, the applicant company removed the comments once it had been notified by the injured party. The Court also found that most of these comments were hate speech or incitements to violence, which are not protected by Article 10 of the Convention. Therefore, the right to free speech of the authors of these comments had not been violated. The issue presented to the Court was thus merely whether holding Delfi liable for comments posted by third parties had been a violation of Article 10.

> The First Section of the Court had identified several relevant issues in the case[858] and examined them when assessing whether the interference with freedom of expression could be justified:
>
> (1) the context of the comments;
>
> (2) the measures applied by Delfi to prevent or remove defamatory comments;
>
> (3) the liability of the actual authors of the comments as an alternative to the applicant company's liability, and:
>
> (4) the consequences of the domestic proceedings for Delfi.

Regarding the context of the comments, while the article about the ferry company was balanced and did not contain illegal speech, it was published on Delfi's platform which had been designed to attract many comments under each article. The comments section

[858] Delfi SA v. Estonia, Grand Chamber, § 142.

was integrated to the news portal and "[*t*]*he number of visits to the* [*Delfi*]*'s portal depended on the number of comments; the revenue earned from advertisements published on the portal, in turn, depended on the number of visits*."[859] Delfi's Rules prohibited posting threatening, insulting or obscene language, and had the sole power to modify or remove comments once posted. As such, the Grand Chamber agreed with the Chamber's finding that Delfi had to be "*considered to have exercised a substantial degree of control over the comments published on its portal*,"[860] and that it had been sufficiently established that Delfi's *"involvement in making public the comments on its news articles on the Delfi news portal went beyond that of a passive, purely technical service provider*."[861] For the Grand Chamber, the comments posted on Delfi's site had been made in reaction to an article published on the site, which is a "*professionally managed news portal run on a commercial basis*." However, the measures taken by Delfi to remove "*without delay after publication comments amounting to hate speech and speech inciting violence and to ensure a realistic prospect of the authors of such comments being held liable*" had not been sufficient and thus the local court's imposition of liability on Delfi had been based on relevant and sufficient grounds and was not a disproportionate restriction on the applicant company's right to freedom of expression.[862]

The issue of content moderation was addressed by the European Court of Justice when asked by the *Oberster Gerichtshof*, the Austrian Supreme Court, in *Eva Glawischnig-Piesczek v. Facebook Ireland Limited,* to review which obligations may be imposed on a host

[859] Delfi SA v. Estonia, Grand Chamber, § 144.
[860] Delfi SA v. Estonia, Grand Chamber, § 145.
[861] Delfi SA v. Estonia, Grand Chamber, § 146.
[862] Delfi SA v. Estonia, Grand Chamber, § 162. The right to one's honor is protected by Article 17 of the Estonian Constitution, under which "*No one's honor or good name shall be defamed*." Article 19(2) of the Constitution states that "*Everyone shall honor and consider the rights and freedoms of others and shall observe the law in exercising his or her rights and freedoms and fulfilling his or her duties*."

provider without a general obligation as provided by article 15 of the e-Commerce Directive, which has been transposed by Austria in its *E-Commerce Gesetz*. The referring court also asked the European Court of Justice to rule on whether a host provider can be ordered by a court to remove content for internet users in the Member States where the court issuing the order is located, but also worldwide.

In this case, Plaintiff was a member and federal spokesperson of the Austrian Green Party. In April 2016, a Facebook user publicly shared an article about the party which had been published online, thus creating a thumbnail of the article on the platform, which featured the photograph of Plaintiff. The Facebook user had added insulting and defamatory comments in connection with the article, calling Plaintiff, among other insults, a member of a 'fascist party.' Plaintiff contacted Facebook Ireland in July 2016, asking it to remove the comments. As it did not comply with her request, she filed a suit in Austria, asking the *Handelsgericht*, the Commercial Court of Vienna, Austria, to issue an injunction ordering Facebook to remove the comments and her photograph from the platform. The Court issued such order in December 2016 and Facebook complied, then appealed the decision and requested to limit the interlocutory order to Austria. This request was, however, not granted by the *Oberlandesgericht Wien*, the Hight Regional Court in Vienna, which held instead that Facebook Ireland was only obliged to cease dissemination of allegations brought to its knowledge by the applicant, main parties, or otherwise. Both parties then appealed to the *Oberster Gerichtshof*, which stayed the proceeding and referred its questions to the European Court of Justice.

In his opinion, Advocate General Szpunar reviewed how article 14 and article 15 of the e-Commerce Directive two articles have been interpreted over the years.[863] He noted that, would a Member State impose to a service provider, which merely stores information, a general obligation to monitor such stored information, the provider would lose its status of intermediary service provider and the immunity granted to these providers by the e-Commerce Directive, which would *"undermine the practical effect*"[864] of article 14 of the Directive. Interestingly for our topic of study, Advocate General Szpunar wrote that "*the identification of information equivalent to that characterized as illegal originating from other users would require the monitoring of all information disseminated via a social network platform*,"[865] which would not be neutral, and the service provider would even "*exercis*[*e*] *a form of censorship…*[*by*] *becom*[*ing*] *an active contributor to that platform.*" Advocate General Szpunar proposed to interpret article 15 of the Directive as not preventing a social network platform from being enjoined by a court "*to seek and identify, among all the information disseminated by users of that platform, the information that was characterized as illegal*" by the court issuing the injunction.[866] Advocate General Szpunar noted that there is not harmonized EU defamation law, nor is there harmonized EU conflict-of-law rules in this field, and that, therefore, a Member State cannot require an electronic platform to delete defamatory speech worldwide, as it would lead to information being illegal in one

[863] Opinion of Advocate General Szpunar, June 4, 2019, Case C-18/18, Eva Glawischnig-Piesczek v. Facebook Ireland Limited, http://curia.europa.eu/juris/document/document.jsf;jsessionid=33E522865BBB5473F2B3D87EBC84F0B2?text=&docid=214686&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=2458372.
[864] Opinion of Advocate General Szpunar, Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 38.
[865] Opinion of Advocate General Szpunar, Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 73.
[866] Opinion of Advocate General Szpunar, Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 75.

Member States to possibly be illegal in other States,[867] noting further that protection of personal data is outside the scope of the E-Commerce Directive.[868]

The European Court of Justice ruled on October 3, 2019, that a host provider can be enjoined by a court of a Member States to remove or block illegal information stored if monitoring this information, worldwide and "*within the framework of the relevant international law*."[869] The Court took care to state preliminary that "*directives covering the supply of information society services must insure* that *this activity may be engaged in freely in the light of* [*article 10 of the European Convention on Human Rights*]."[870] For the Court, article 15 of the E-Commerce Directive, prohibiting Member States from imposing a general obligation to monitor to host providers, must not be interpreted as prohibiting obliging them to monitor "*in a specific case*,"[871] such as cases where "*a particular piece of information stored by the host provider concerned at the request of a certain user of its social network*"[872] which has been found illegal by a court of law after analysis.[873] The Court pointed out that, because of the way social media networks operate, "*there is a genuine risk that information which was held to be illegal is subsequently reproduced and shared by another user of that network*."[874] Indeed, once a content has been flagged as illegal, it may

[867] Opinion of Advocate General Szpunar, Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 80.
[868] Opinion of Advocate General Szpunar, Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 90.
[869] Case C-18/18, Eva Glawischnig-Piesczek v. Facebook Ireland Limited ()CT. 3, 2019), European Court of Justice (Third Chamber), available at http://curia.europa.eu/juris/document/document.jsf?text=&docid=218621&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=9989845.
[870] Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 9.
[871] Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 34.
[872] This is a rather convoluted way to describe user-generated content stored on a platform.
[873] Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 35.
[874] Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 36.

be taken down or blocked "*expediously*," but this swift action does not prevent the content to be reposted on the same platform, or even on other platforms, forcing host providers to engage in a wild goose chase. This was recently the case when the image of the body of a French teacher, who had been decapitated in October 2020 in front of the school where he taught, had been published on Twitter by the killer, taken down almost immediately by the platform, yet remained online on other websites several weeks after the murder.[875] To prevent such dissemination, a court may legitimately require host providers to block access to content "*previously declared to be illegal,*"[876] and this encompasses content which is not exactly the same that the content found to be illegal, but is essentially similar to it, so that the injunction of a court cannot be circumvented by claiming that the content is not the same than the illegal content. The Court explained that an information is similar to a previously deemed illegal information if the host provider is not obliged "*to carry out an independent assessment of that content*."[877] How could it be assessed that a particular information is equivalent to an illegal information? The case was about defamation, and the Court specified that an equivalent information:

> "*contains specific elements which are properly identified in the injunction, such as the name of the person concerned by the infringement determined previously, the circumstances in which that infringement was determined and equivalent content to that which was declared to be illegal. Differences in the wording of that equivalent content, compared with the content which was declared to be illegal, must not, in any*

[875] *Conflans : une enquête vise un site néonazi, après la diffusion de la photo du professeur décapité*, LE PARISIEN, (Oct. 19 2020 23h39), https://www.leparisien.fr/faits-divers/conflans-une-enquete-vise-un-site-neonazi-apres-la-diffusion-de-la-photo-du-professeur-decapite-19-10-2020-8404028.php.
[876] Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 37.
[877] Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 53.

event, be such as to require the host provider concerned to carry out an independent assessment of that content."[878]

However, it can be argued that this guidance may very well be followed for defamation cases but does not provide guidelines precise enough in other cases of illegal speech, such as incitement to violence or even hate speech, and the e-Commerce Directive "*does not comprehensively regulate the permissible scope of injunctions*."[879]

The European Union Law on online intermediaries' liability is set to change soon as the European Commission published on December 15, 2020 its Digital Service Act package (DSA) proposal,[880] which would amend the e-Commerce Directive, and which aims, inter alia, at addressing the growing influence on online platforms in our lives.[881] It "*defines clear responsibilities and accountability for providers of intermediary services, and in particular online platforms, such as social media and marketplaces*."[882] One of the issues addressed by the DSA will be the dissemination of illegal content online, including hate speech, incitement to terrorism, and child sexual abuse material. An internal note of officials in the Commission's Directorate General for Communications Networks, Content

[878] Eva Glawischnig-Piesczek v. Facebook Ireland Limited, paragraph 45.
[879] Anja Hoffmann& Alessandro Gasparotti, *Liability for illegal content online Weaknesses of the EU legal framework and possible plans of the EU Commission to address them in a "Digital Services Act", CENTRUM FÜR EUROPÄISCHE POLITIK,* (March 2020), p. 10, https://www.cep.eu/fileadmin/user_upload/cep.eu/Studien/cepStudie_Haftung_fuer_illegale_Online-Inhalte/cepStudy_Liability_for_illegal_content_online.pdf.
[880] European Commission, *Proposal for a Regulation of the European Parliament and of the Council on a single market or Digital Services (Digital Services Act) and amending Directive 2000/31/EC*, COM(2020) 82 5final (Dec. 15, 2020), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en
[881] *How do online platforms shape our lives and businesses?*, THE EUROPEAN COMMISSION, (Sept. 18, 2019), https://ec.europa.eu/digital-single-market/en/news/how-do-online-platforms-shape-our-lives-and-businesses-brochure. The brochure, which provides figures on uses of online platforms in the European Union, reveals that only 30% of European Union citizens never use social media platforms.
[882] Digital Service Act, p. 2.

and Technology had been leaked online in July 2019.[883] It shows that the DSA meant to address intermediary liability, and:

> "*could mean that the notions of mere conduit, caching and hosting service could be expanded to include explicitly some other services*," and that "*the concept of active/passive hosts would be replaced by more appropriate concepts reflecting the technical reality of today's services, building rather on notions such as editorial functions, actual knowledge and the degree of control*."[884]

The leaked document had also revealed that general monitoring would remain prohibited, but that the DSA would address the issue of algorithms for automated filtering technologies, how they should be used in a transparent manner and how automated content moderation systems should be made accountable. The leaked document stated also that uniform rules for the removal of illegal content, including hate speech, "*would be made binding across the EU, building on the Recommendation on illegal content and relevant case-law, and include a robust set of fundamental rights safeguards*," and were described as "*notice-and action rules*" which would be tailored to the type of services to which they apply, such as a social network.

Before publishing its DSA proposal, the EU Commission initiated a public consultation, open from June 2, 2020, to September 8, 2020, aiming at helping it *"analyzing and collecting evidence for scoping the specific issues that may require an EU-level*

---

[883] The leaked document is available at https://www.cep.eu/fileadmin/user_upload/cep.eu/Studien/cepStudie_Haftung_fuer_illegale_Online-Inhalte/cepStudy_Liability_for_illegal_content_online.pdf./

[884] Internal note of officials in the Commission's Directorate General for Communications Networks, Content and Technology, p. 5.

intervention."[885] The public comments were published online.[886] In its comments, Facebook Ireland identified four areas: harmful content, election integrity, privacy and data portability, which it was "*very keen to continue to explore both in the Digital Services Act (DSA) process and beyond*."[887] Facebook Ireland commented on the apparent wish of the Commission to have "*large online platforms acting as gatekeepers*, " noting that it was "*striking… that the Commission seems to be proposing that a company is considered a 'gatekeeper'* (*and therefore within the scope of the framework*) *on the basis of criteria one could summarize as relating to the 'size' of the company and its potential multimarket activity*."[888] It proposed that, if the Commission does not opt for a case-by-case approach to assess which companies are within the scope of the framework, but instead provides a list of such companies, the list should be reviewed periodically, as "*the digital sector is dynamic and fast-changing* [*and*] [*t*]*hat needs to be accounted for in the framework*."[889] In contrast, the non-profit organization *Article 19* wrote in its comments that the small number of large online platforms identified by the European Commission as gatekeepers are not only "*economic gatekeepers, but also as 'fundamental rights' gatekeepers*," adding that "[*t*]*hrough their business models, their terms of services and community guidelines, these platforms set standards in the market with regards to, among others, consumers' rights to privacy, data*

---

[885] *The Digital Services Act package*, THE EUROPEAN COMMISSION, https://ec.europa.eu/digital-single-market/en/digital-services-act-package (last visited Dec. 30, 2020).
[886] *Feedback received on: Digital Services Act package – ex ante regulatory instrument of very large online platforms acting as gatekeepers* , https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12418-Digital-Services-Act-package-ex-ante-regulatory-instrument-of-very-large-online-platforms-acting-as-gatekeepers/feedback?p_id=7937460.
[887] *Facebook observations to the Inception Impact Assessment on the DSA Ex-Ante Instrument of very large Online Platforms acting as Gatekeepers* (June 30, 2020), available at https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12418-Digital-Services-Act-package-ex-ante-regulatory-instrument-of-very-large-online-platforms-acting-as-gatekeepers/F535672 (last visited Dec. 30, 2020).
[888] Facebook observations p.7.
[889] Facebook observations p. 10.

protection and freedom of expression."[890] They have the power to do it because the barriers to market entry are so high that new players may only enter the market with great difficulties and also "*because consumers do not have viable alternatives to switch to*." Would data portability become a right for EU consumers, thus allowing them to leave a social media platform with their data and upload it to a new platform, new players may be willing to challenge the strong hold on social media market held by Facebook, Instagram and Twitter.[891]

B. The Private Law of the Platforms

Franck LaRue, then U.N. Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, wrote in 2011 that "*censorship measures should never be delegated to a private entity, and… no one should be held liable for content on the Internet of which they are not the author. Indeed, no State should use or force intermediaries to undertake censorship on its behalf*."[892]

---

[890] Article 19, *Joint statement in response to the inception impact assessments on a new competition tool and an ex ante regulatory instrument of very large online platforms acting as gatekeepers*, available at https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12418-Digital-Services-Act-package-ex-ante-regulatory-instrument-of-very-large-online-platforms-acting-as-gatekeepers/F535671 (last visited Dec. 30, 2020).

[891] We saw that Representative Nunes started promoting the *Parler* platform, as "Parler will set you free", after having been criticized and mocked by an anonymous Twitter user. The *Parler* platform does not seem, however, to have been much successful at carving a viable niche for its product, a social media for conservative-leaning users. However, the *Parler* app was downloaded 500,000 downloads per week at one point during the summer of 2020, after President Trump publicly toyed with the idea to use it as his social platform of choice, and its user base was almost 2 million, see Abram Brown, *The App That The Proud Boys Used To Celebrate Donald Trump's Debate Performance*, FORBES, (Sep. 30, 2020,12:43am EDT), https://www.forbes.com/sites/abrambrown/2020/09/30/the-app-that-the-proud-boys-used-to-celebrate-donald-trumps-debate-performance/#63eece5f64fe.

[892] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 16 May 2011, A/HRC/17/27, para. 43, available at https://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf. The Special Rapporteur gave as example the Korea Communications Standards Commission of the Republic of Korea, which he described as "a *quasi-State and quasi-private entity tasked to regulate online content*."

Is moderating speech censorship? This practice is necessary for social media platforms not to become unmanageable cesspools of hateful and illegal speech. A June 1, 2011 Joint declaration on freedom of expression and the Internet Self-regulation stated that "[*s*]*elf-regulation can be an effective tool in redressing harmful speech, and should be promoted*.[893] The European Commission noted in October 2015 its conclusions paper to its first colloquium on hate speech that "[*e*]*nforcement and implementation of EU legislation obliging Member States to penalise hate speech inciting racist hatred or violence, including on ethnic, national or religious grounds, needs to be rigorously monitored."*[894] This colloquium led to a code of conduct on countering illegal hate speech online developed in 2016 by the European Commission in cooperation with Twitter, Facebook, Microsoft and YouTube.[895] Instagram, Google+, Dailymotion, Snap and Jeuxvideo.com joined the Code of Conduct scheme in 2018 and 2019.[896] The code of conduct cited the European Union Council Framework Decision 2008/913/JHA of November 28, 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law. By signing the code of conduct, the companies agreed to remove illegal hate speech in less than twenty-four hours

[893]Joint Declaration by The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organisation for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organisation of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information, article 1(e), available at https://www.osce.org/fom/78309.

[894] European Commission, *Joining forces against anti-Semitic and anti-Muslim hatred in the EU: outcomes of the first Annual Colloquium on Fundamental Rights*, EUROPEAN COMMISSION, (Oct.9, 2015), p.5. http://ec.europa.eu/justice/events/colloquium-fundamental-rights-2015/index_en.htm.

[895] *European Commission and IT Companies announce Code of Conduct on illegal online hate speech*, (May 31, 2016), http://europa.eu/rapid/press-release_IP-16-1937_en.htm (last visited Dec. 30, 2020). TikTok joined the European Union's Code of Conduct on Countering Illegal Hate Speech in September 2020. See Natasha Lomas, *TikTok joins Europe's code on tackling hate speech*, TECHCRUNCH, (Sept. 8, 2020 6:23 am EDT), https://techcrunch.com/2020/09/08/tiktok-joins-europes-code-on-tackling-hate-speech.

[896] *Information note - Progress on combating hate speech online through the EU Code of conduct 2016-2019*, 12522/19, COUNCIL OF THE EUROPEAN UNION, (Sept. 27, 2019), p.2, https://ec.europa.eu/info/sites/info/files/aid_development_cooperation_fundamental_rights/assessment_of_the_code_of_conduct_on_hate_speech_on_line_-_state_of_play__0.pdf.

and to remove or disable access to it. The Code of Conduct is, however, a mere list of public commitments, a "*soft law*"[897] which is not binding. For instance, Twitter does not take down every single message which can be interpreted as hate speech, frustrating some users, who once noted, for instance, that the platform was keener at removing content infringing the copyright of the Olympic Committee than rape threats.[898] However, it was reported in 2018, in an EU Commission Staff Report, that it had been at the origin of an increasing percentage of removals of content notified under it: 28% as of December 2016, 59% as of June 2017, and 70 % as of January 2018.[899]

Platforms have their own private rules, which may or may not be similar to the laws a country or to a group's morale or religion. Blocking, censoring, controlling speech on a particular platform can be made at the sole initiative of the provider, or may be triggered by laws. When UK Culture Secretary Jeremy Wright presented in April 2019 the government *Online Harms White Paper*,[900] he wrote: "*We propose a duty of care for those online companies which allow users to share or discover user-generated content, or that allow users to interact with each other online*."[901] But laws and private rules do not always intersect perfectly, far from it. For instance, Facebook's policy on nudity led to some highly

---

[897] Anja Hoffmann& Alessandro Gasparotti, Liability for illegal content online Weaknesses of the EU legal framework and possible plans of the EU Commission to address them in a "Digital Services Act", CENTRUM FÜR EUROPÄISCHE POLITIK, (March2020), p. 22.
[898] Rosalyn Warren, *Why Isn't Twitter Taking Down Harassment As Fast As It Takes Down Olympics Content?*, BUZZFEED (Aug. 17, 2016, 1:43 pm), https://www.buzzfeed.com/rossalynwarren/i-cant-believe-they-deleted-that-perfect-santana-tweet?utm_term=.myXZZpq0a#.sl4yynDjK.
[899] Commission Staff, Regulation Impact Assessment, 135 (Sept. 12, 2018), https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-swd-408_en.pdf.
[900] *Open consultation, Online Harms White Paper*, GOV.UK, https://www.gov.uk/government/consultations/online-harms-white-paper.
[901] *Jeremy Wright speaking at the launch of the Online Harms White Paper*, GOV.UK, (April 8, 2019), https://www.gov.uk/government/speeches/jeremy-wright-speaking-at-the-launch-of-the-online-harms-white-paper.

publicized cases where accounts have been suspended for having posted artistic nudes. Several of these cases stemmed from France, a country known for its open-mindedness, if not its outright enthusiasm, over nude pictures.

We will now discuss how platforms are regulating, or not regulating, three categories of speech, (a) nudity, (b) fake information and (c) "hate speech".

a. Nudity and the Platforms

In 2013, Facebook suspended for twenty-four hours the Facebook account of the *Musée du Jeu de Paume* in Paris, because it had published on its Facebook page a photography of a female nude by Laure Albin Guillot, an artist who was exhibited at the time by the *Jeu de Paume* museum.[902] The museum decided after that to stop publishing pictures of nude art on its Facebook page to avoid its account to be terminated. These episodes led French legal editorialist Félix Rome to write that there is now being "*quietly pus up on the Web, an uncultivated private justice which allows champions of sexual morality, which render it without any control or recourse and without, of course, ensuring that the rights of the defense are being respected far and near, to close the accounts of their users arbitrarily and discretionally*."[903] Félix Rome specifically referred to the suspension of two Facebook accounts which had published arts works featuring female nudity. *L'origine du Monde* (The Origin of the World) is a painting by Gustave Courbet, the title of which refers to almost everybody's port of entry into this world, the vagina. The painting shows the lower half of the abdomen of a woman, spreading her legs widely apart to revel the *corpus*

[902] Amandine Schmitt, *Facebook : le musée du Jeu de Paume ne publiera plus de nus*, L'OBS, (March 6, 2013, 18 :46), http://tempsreel.nouvelobs.com/les-internets/20130305.OBS0862/facebook-le-musee-du-jeu-de-paume-ne-publiera-plus-de-nus.html.
[903]Félix Rome, *Réseau Moral*, D. 2013, 633.

delicti. If it can be argued that, while *L'origine du Monde* may be "obscene," its title reflects on the dual role a woman may play, sexual object or mother, one leading sometimes to the other, and is thus more philosophical and thought inducing than obscene. A French Facebook user, which account had been taken down by Facebook after he had published on his wall a reproduction of the Gustave Courbet painting, unsuccessfully asked Facebook to reinstate his account. He then filed a suit in France to have it reinstated. Facebook moved to have the suit dismissed for lack of personal jurisdiction. The French judge in charge of reviewing the validity of the cause of action pre-trial[904] found that the exclusive jurisdiction clause was void and held that the French courts had jurisdiction over the case. Facebook appealed, but the Paris Court of appeals held in favor of the consumer on February 12, 2016,[905] noting that Facebook and the user were linked by a standard agreement (*contrat d'adhésion*), with Facebook being the professional party of the contract and the user the consumer party of the contract, who had merely adhered to the contract without being able to negotiate any of its terms.[906] Under French law, such consumers are granted special protection.[907] Facebook had argued that the contract was not a consumer's contract, as users do not have to pay a fee to use the service. But this argument did not convince the court, which reasoned that, while the service was indeed free of charge, Facebook was able to garner vast amount of revenue from the sale of advertising.

---

[904] *Le juge de la mise en état.*
[905] Cour d'appel [CA] [regional court of appeal] Paris, Pôle 2, 2eme ch., Feb. 12, 2016.
[906] Article 1110 of the French civil Code defines such contract as a contract "*which terms and conditions, exempt from the negotiations, are determined in advance by one of the parties*."
[907] *See* article L. 132-1 of the French consumers 'Code (*Code de la consommation*), which defines unfair clauses of such contracts as clauses having "*for purpose or effect to create, at the prejudice of the non-professional party or the consumer, a significant imbalance between the rights and obligations of the parties to the contract*." Such clauses are deemed invalid under article L. 132-1 ("*réputées non écrites*.")

Facebook's ban of female nudity used to encompass pictures of breastfeeding[908] and even the breasts of the 30,000 years old paleolithic statue of a naked woman.[909] Facebook even took down in 2016 a post by Norwegian author Tom Egeland which featured the iconic 1972 Nick Ut "*The Terror of War*" photograph of a frightened Vietnamese naked girl, running away from a Napalm attack, crying. Mr. Egeland's Facebook account was even temporarily suspended.[910] The *Aftenposten*, Norway's largest newspaper, published an article commenting on the suspension, used the iconic image to illustrate it and even featured the photograph on its front page It then posted the article on its own Facebook page, which was taken down as well. Espen Egil Hansen, *Aftenpostnen* editor-in-chief, then published a letter to Mark Zuckerberg.[911] He called him the "*world's most powerful editor*," and urged him "*to realize that* [*he is] restricting* [*Mr. Hansen's*] *room for exercising* [*his*] *editorial responsibility*," arguing further that it was an abuse of power. For Mr. Hansen, Facebook's rules "*don't distinguish between child pornography and famous war photographs" and the way they are enforced does not provide "space for good judgement*," a situation which is further worsened as Facebook censored criticism against the decision to

---

[908] *See* Jean Zeid, #br*elfies : quand les mamans postent des selfies d'allaitement contre la censure Facebook*, France TV INFO, (Feb. 25, 2015, 12 : 05), https://www.francetvinfo.fr/replay-radio/le-17-20-numerique/brelfies-quand-les-mamans-postent-des-selfies-d-allaitement-contre-la-censure-facebook_1772801.html.

[909] This particular picture was, however, apparently taken down by mistake, and Facebook apologized. See *Facebook présente ses excuses après la censure d'une Vénus paléolithique*, LE MONDE, (MARCH 1, 2018 11 :03), https://www.lemonde.fr/pixels/article/2018/03/01/une-venus-paleolithique-censuree-sur-facebook_5264174_4408996.html. Facebook however still take down photographs of works of art featuring nudity, *see* for example, Associated Press, *Swiss Museum Laments Facebook Ban of Images of Naked Statues* (Feb. 4, 2019), https://www.courthousenews.com/swiss-museum-laments-facebook-ban-of-images-of-naked-statues.

[910] Mark Scott and Mike Isaac, *Facebook Restores Iconic Vietnam War Photo It Censored for Nudity*, THE NEW YORK TIMES, (Sept. 9, 2016), https://www.nytimes.com/2016/09/10/technology/facebook-vietnam-war-photo-nudity.html.

[911] Espen Egil Hansen, *Dear Mark. I am writing this to inform you that I shall not comply with your requirement to remove this picture,* AFTENPOSTNEN, (Sept. 8, 2016, 21:33), https://www.aftenposten.no/meninger/kommentar/i/G892Q/dear-mark-i-am-writing-this-to-inform-you-that-i-shall-not-comply-wit.

suspend the account and punished the journalist who reported the issue by censoring the journalist's post as well. *Instagram*, owned by Facebook also deletes sometimes representation of nudity, as experienced by fashion editor Grace Coddington in 2014. After she had posted a drawing of herself reclining back on a lounge chair, naked, her *Instagram* account was suspended, to be later reinstalled following online protests of fans.[912] A photograph of model Christine Teigen showing her naked from the waist up was deleted from *Instagram* after the model posted it to publicize an incoming issue of the fashion magazine *W*. Ms. Teigen then posted on Twitter: "*the nipple has been temporarily silenced but she will be back, oh yes, she will be back*."[913]

Facebook's argued that its nudity policies "*have become more nuanced over time*," as the platform "*understand*[*s*] *that nudity can be shared for a variety of reasons, including as a form of protest, to raise awareness about a cause, or for educational or medical reasons*." Facebook still "*restrict*[*s*] *some images of female breasts that include the nipple*, [*but*] *allow*[*s*] *other images, including those depicting acts of protest, women actively engaged in breast-feeding, and photos of post-mastectomy scarring*. [*It*] *also allow*[*s*] *photographs of paintings, sculptures, and other art that depicts nude figures*."[914]

One of the fist cases accepted by Facebook Oversight Board deals with a decision to take down a post, featuring eight photographs showing breast cancer symptoms and

---

[912] Isabel Wilkinson, *Grace Coddington Was Temporarily Removed From Instagram for Nudity,* THE CUT, (May 19, 2014, 2:28 pm), http://nymag.com/thecut/2014/05/grace-coddington-kicked-off-instagram-for-nudity.html.
[913] Sam Reed, *Chrissy Teigen Wages War Against Instagram's Nipple Ban*, HOLLYWOOD REPORTER, (June 30, 2015, 11:48 AM PDT), https://www.hollywoodreporter.com/news/chrissy-teigen-wages-war-instagrams-805998.
[914] *Facebook Community Standards, paragraph 14, Adult Nudity and Sexual Activity*, https://www.facebook.com/communitystandards/adult_nudity_sexual_activity (last visited Dec. 30, 2020).

explaining the symptoms underneath. Eight of these photographs showed "*visible and uncovered female nipples*," while the nipples were not visible in the other three pictures. Facebook removed the post for violating its policy on Adult Nudity and Sexual Activity. The post's title indicated they had been posted to raise awareness of signs of breast cancer .[915] Indeed, not all representations of nudity depict the body as a vessel of sexual pleasure, and "Adult Nudity" is not necessarily linked to "Sexual Activity," but may be used instead to make a political statement. Facebook apologized in 2020 for having taken down a post featuring a 19th century photograph of naked Aboriginal men with chains around their necks.[916] The image had been posted as a comment after Australia's Prime Minister Scott Morrison had claimed there had been 'no slavery in Australia.' He later apologized and acknowledged Australia's former 'hideous practices.'[917] This case is about the collusion of nudity and inaccurate information. Facebook should not have taken down the photograph of the enslaved Aboriginal men, their nudity a proof of their abject servitude. Should Facebook had taken down instead posts referring to the inaccurate assertion of the Prime Minister that there had been no slavery in Australia?

---

[915] FACEBOOK Oversight Board case 2020-004-IG-UA, see OVERSIGHT BOARD, *Announcing the Oversight Board's first cases and appointment of trustees*, (Dec. 2020), https://www.oversightboard.com/news/719406882003532-announcing-the-oversight-board-s-first-cases-and-appointment-of-trustees.

[916] Josh Taylor, *Facebook incorrectly removes picture of Aboriginal men in chains because of 'nudity'*, THE GUARDIAN, (June 12, 2020 19.11 EDT), https://www.theguardian.com/technology/2020/jun/13/facebook-incorrectly-removes-picture-of-aboriginal-men-in-chains-because-of-nudity.

[917] Katharine Murphy, *Scott Morrison sorry for 'no slavery in Australia' claim and acknowledges 'hideous practices'*, THE GUARDIAN, (June 2020 00.49 EDT), https://www.theguardian.com/australia-news/2020/jun/12/scott-morrison-sorry-for-no-slavery-in-australia-claim-and-acknowledges-hideous-practices.

b. Fake Information and the Platforms

Facebook, Google, Twitter, and Mozilla signed an EU *Code of Practice on Disinformation* in October 2018 and each presented a roadmap to implement the Code. Microsoft signed it in May 2019, and TikTok in June 2020.[918] The Code defines "disinformation" as:

> "*verifiably false or misleading information" which, cumulatively, (a) "[i]s created, presented and disseminated for economic gain or to intentionally deceive the public"; and (b) "[m]ay cause public harm", intended as "threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security*."

It does not, however, include "*misleading advertising, reporting errors, satire and parody*,[919] *or clearly identified partisan news and commentary*…"

Signatories committed to "*use commercially reasonable efforts to implement policies and processes… not to accept remuneration from, or otherwise promote accounts and websites which consistently misrepresent information about themselves*."[920] They also committed to clearly distinguish advertisements "*from editorial content, including news, whatever their form and whatever the medium used*."[921]

---

[918] *Code of Practice on Disinformation*, EUROPEAN COMMISISON, (Sep. 26, 2018), https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation.
[919] This reserve had also been made by the French Constitutional Council in its Decision no. 2018-773 DC of 20 December 2018 on the law on the fight against the manipulation of information.
[920] *Code of Practice on Disinformation*, IIA.
[921] *Code of Practice on Disinformation*, IIB2.

The issue of bots is addressed, as signatories recognized "*the importance of intensifying and demonstrating the effectiveness of efforts to close fake accounts… [and]… of establishing clear marking systems and rules for bots to ensure their activities cannot be confused with human interactions.*"[922] A Commission Staff Working Document on the *Assessment of the Code of Practice on Disinformation* (Assessment) published in September 2020, noted, however, that "malicious bots" was a term which would "*benefit from uniform definition and application*."[923] The Assessment identified four broad categories which could be improved: (1) inconsistent and incomplete application of the Code across platforms and Member States, (2) lack of uniform definitions, (3) existence of several gaps in the coverage of the Code commitments, and (4) limitations intrinsic to the self-regulatory nature of the Code. The Assessment noted, for instance, that while platforms have "*put in place policies to counter the use of manipulative techniques and tactics on their services*," the impact and relevance of these measures is difficult to evaluate because reporting on the measures "*is provided at aggregated and global level*." The Assessment called for more transparency "*about the levels of user engagement with detected disinformation campaigns*."

Facebook is engaged in the fight against fake information, which it names "*coordinated inauthentic behavior"* (CIB) and can be either "*coordinated inauthentic behavior in the context of domestic, non-state campaigns (CIB) and … coordinated inauthentic behavior on behalf of a foreign or government actor (FGI*)." CIB content warrants removal of inauthentic and authentic accounts, Pages and Groups directly involved while

[922] *Code of Practice on Disinformation*, IIC.
[923] *Commission Staff Working Document on the Assessment of the Code of Practice on Disinformation*, EUROPEAN COMMISSION, p. 13, available for download at https://ec.europa.eu/digital-single-market/en/news/assessment-code-practice-disinformation-achievements-and-areas-further-improvement.

FGI more broadly warrants the removal of "*every on-platform property connected to the operation itself and the people and organizations behind it.*"[924] A dedicated monitoring is then put in place to avoid these pages and accounts to be activated again.

Facebook's Community Standard also forbids posting "false news" on the platform. It states that there is "*a fine line between false news and satire or opinion*," and this is why the platform does not remove false news, but instead, "*significantly reduce*[*s*] *its distribution by showing it lower in the News Feed*."[925] It addressed the issue specially on its blog, noting that, while false news does not violate the platform's community standards, such speech "*often violates* [*its*] *polices* [sic] *in other categories, such as spam, hate speech or fake accounts, which [the platform] remove*.[926] False news which is considered to be spam is removed. Facebook uses the services of third-party fact-checkers in charge of reviewing and rating the accuracy of articles and posts on the platform. They are certified by the non-partisan International Fact-Checking Network, a unit of the *Poynter Institute*.[927] As wide-spread circulation of disinformation about a pandemic has the potential to be a danger to public health, Facebook announced in March 2020 that it had partnered with *The International Fact-Checking Network* (IFCN)" *to launch a $1 million grant program to increase their capacity during this time*,"[928] an international fact-checking program

---

[924] April 2020 Coordinated Inauthentic Behavior Report, FACEBOOK, (May 5, 2020), https://about.fb.com/news/2020/05/april-cib-report/.
[925] *Community Standards, 21. False News*, FACEBOOK, https://www.facebook.com/communitystandards/integrity_authenticity (last visited Dec. 30, 2020).
[926] Tessa Lyons (Facebook Product Manager), *Hard Questions: What's Facebook's Strategy for Stopping False News?* (May 23, 2018), https://about.fb.com/news/2018/05/hard-questions-false-news/.
[927] *The International Fact-Checking Network,* THE POYNTER INSTITUTE, https://www.poynter.org/ifcn/ (last visited Dec. 30, 2020).
[928] Kang-Xing Jin, *Head of Health, Keeping People Safe and Informed About the Coronavirus - Supporting Fact-Checkers and Local News Organizations*, FACEBOOK, (Update, March 17, 2020, 6:15AM PT), https://about.fb.com/news/2020/10/coronavirus/.

supporting projects aiming at fighting COVID-19 misinformation in Europe, Asia, Africa, Oceania, and the Middle East.[929] Twitter announced in April 2020 that it would increase its use of machine learning and automation "*to take a wide range of actions on potentially abusive and manipulative content.*"[930] YouTube had already announced in January 2019 that it would start reducing "*recommendations of borderline content and content that could misinform users in harmful ways—such as videos promoting a phony miracle cure for a serious illness, claiming the earth is flat, or making blatantly false claims about historic events like 9/11.*"[931]

Hashtags can be used to attempt spreading false information, such as the #DCblackout hashtag published on social media during the June 2020 protests over the death of George Floyd and was used to claim that telephone and internet communications had been cut. However, there had been no black out in Washington D.C. and Twitter banned use of the hashtag,[932] even publishing a page countering this hoax, which gathered messages from journalists on the scene assuring that they were able to use their phone and the internet to relay their information.[933] Hoaxes are particularly concerning, as they are published with the knowledge that they are false and with the intent that people would believe them. The result may be a confusing blurring in the mind of the public of what is

---

[929] Keren Goldshlager and Orlando Watson*, Launching a $1M Grant Program to Support Fact-Checkers Amid COVID-19*, FACEBOOK JOURNALISM PROJECT (April 30, 2020), https://www.facebook.com/journalismproject/coronavirus-grants-fact-checking.
[930] Matt Derella AND Vijaya Gadde, *An update on our continuity strategy during COVID-19*, TWITTER, (March 16, 2020, update April , 1, 2020), https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html.
[931] *Continuing our work to improve recommendations on YouTube*, YOUTUBE OFFICIAL BLOG, (Jan. 25, 2019), https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html.
[932] *George Floyd protests: Twitter bans over #DCBlackout hoax*, BBC NEWS, (June 2, 2020), https://www.bbc.com/news/technology-52891149.
[933] *Reporters show that Washington, DC media blackout stories are inaccurate*, TWITTER (June 1, 2020), https://twitter.com/i/events/1267466548110700544.

right and what is wrong, or the belief that a hoax is indeed right, and that its outrageousness warrants an IRL (in real life) reaction. Such was the case in "Pizzagate", where online posts had been concocted to make people believe that the Washington, D.C. Comet Ping Pong pizzeria was the front post for a child sex-abuse ring. A man, believing these false statements, had carried a loaded AR-15 assault rifle and a revolver into the restaurant and had fired the rifle into a door.[934]

Misinformation may kill in other ways. Pinterest started blocking in February 2019 the ability to search for "vaccine", "vaccination," or "anti-vax" on its platform.[935] Research for these words turns no results, thus preventing anti-vaccination users to use the site as an echo chamber for their theories, which contributes to the renewed spreading of diseases which used to be controlled by vaccination, such as measles, thus threatening global health. As noted by David Kaye, Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression to the United Nations, it "*has been evident during the COVID-19 pandemic, [that] social media and search engine companies have an enormous impact on public discourse and the rights of individuals on and off their platforms*."[936] Belief that vaccines are dangerous are widely shared on social media, and as, turn, endangers the global community: the World Health Organization (WHO) identified "vaccine hesitancy" as

---

[934] *North Carolina Man Pleads Guilty to Charges In Armed Assault at Northwest Washington Pizza Restaurant*, UNITED STATES DEPARTMENT OF JUSTICE (March 24, 2017), https://www.justice.gov/usao-dc/pr/north-carolina-man-pleads-guilty-charges-armed-assault-northwest-washington-pizza.

[935] Robert McMillan and Daniela Hernandez, *Pinterest Blocks Vaccination Searches in Move to Control the Conversation*, THE WALL STREET JOURNAL.COM, (Feb. 20, 2019 6:33 p.m). ET), https://www.wsj.com/articles/next-front-in-tech-firms-war-on-misinformation-bad-medical-advice-11550658601. *See* also Christina Caron, *Pinterest Restricts Vaccine Search Results to Curb Spread of Misinformation*, THE NEW YORK TIMES (Feb. 23, 2019), https://www.nytimes.com/2019/02/23/health/pinterest-vaccination-searches.html.

[936] *Disease pandemics and the freedom of opinion and ex*pression : report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, *A/HRC/44/49*, (April 23, 2020), paragraph 52, https://digitallibrary.un.org/record/3862160?ln=en, (last visited Dec. 30, 2020).

one of the ten threats to global threats in 2019,[937] a statement even more concerning as the COVID-19 pandemic engulfed the world in 2020, and vaccines appear to be the path to a return to normalcy. The COVID-19 crisis showed the danger of disseminating false information about the pandemic, but also showed how technology, particularly social media platforms, may be used to prevent further spreading of misinformation,[938] a phenomenon the WHO named "*infodemics*."[939] To slow down the spreading of "infodemics,", *WhatsApp* started limiting forwarding a message which has been already forwarded at least five times before to other *WhatsApp* users, to only one chat at the time, no longer to five different chats at once.[940] Unlike social media platforms, *WhatsApp* does not see the content of the messages sent, which are encrypted. However, a message becoming viral during a pandemic could be misinformative. By making it less easy to forward a message, whichever its content is, WhatsApp created a "friction" allowing users to think twice about the message they are about to forward. Indeed, misinformation about the virus is not necessarily spread maliciously, but is also spread by people forwarding in good faith a piece of information they think could be useful to their family and friends, and

[937] *Ten threats to global health in 2019*, WORLD HEALTH ORGANIZATION, https://www.who.int/emergencies/ten-threats-to-global-health-in-2019.
[938] Twitter was also a vector in misinformation during the 2014 Ebola outbreak in West Africa. See Sunday Oluwafemi Oyeyemi, Elia Gabarron, Rolf Wynn, Ebola, *Twitter, and misinformation: a dangerous combination?,* British Medical Journal, (2014), p. 349. Available at https://www.bmj.com/content/349/bmj.g6178. The authors studied tweets in English with the terms "Ebola", "prevention" or "cure" sent from Guinea, Liberia, and Nigeria the first week of September 2014 and found out that most of them contained misinformation.
[939] *Coronavirus disease 2019 (COVID-19) Situation Report –45*, WORLD HEALTH ORGANIZATION, (March 2020), https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200305-sitrep-45-covid-19.pdf?sfvrsn=ed2ba78b_4, writing that " *Infodemics are an excessive amount of information about a problem, which makes it difficult to identify a solution. Infodemics can spread misinformation, disinformation and rumors during a health emergency*."
[940] *WhatsApp to limit message forwarding to stop spread of disinformation*, ITV, (Apr. 7 2020, 1:41pm), https://www.itv.com/news/2020-04-07/whatsapp-to-limit-message-forwarding-to-stop-spread-of-disinformation/.

thus becoming a "node" of misinformation.[941] However, setting the obstacle gives users some time to reflect.

This is also why Twitter chose to temporality suspend users' ability to retweet a post without commenting it, ahead of 2020 Election Day,[942] explaining that "*though this adds some extra friction for those who simply want to Retweet, we hope it will encourage everyone to not only consider why they are amplifying a Tweet, but also increase the likelihood that people add their own thoughts, reactions and perspectives to the conversation*."[943] So the issue identified as worthy of protection against false news was the democracy. Twitter also updated its civic integrity policy in October 2020, aiming at preventing manipulation of the elections and other civic processes, such as a census or a referendum.[944]

Twitter identified four categories of misleading behavior and content:

---

[941] Digital, Culture, Media and Sport Sub-committee on *Online Harms and Disinformation*, April 30, 2020, testimony of Claire Wardle, Co-Founder and Director, First Draft News, at 10:27:06 , available at https://parliamentlive.tv/Event/Index/3c4aede5-2b89-4f33-9103-fb1c8a77a3ad. Ms. Wardle further explained that some people, while knowing that the information they forward is probably not true, forward it anyway to their loved ones, just in case it could be useful to them, and that we are all susceptible of this type of behavior, whether educated or not, at 10:28:27. and 10: 41:25. In a report for the Council of Europe she co-authored with Hossein Derakhshan, Dr.Wardle noted that "[*s*]*ocial networks are driven by the sharing of emotional content. The architecture of these sites is designed such that every time a user posts content—and it is liked, commented upon or shared further—their brain releases a tiny hit of dopamine. As social beings, we intuit the types of posts that will conform best to the prevailing attitudes of our social circle*." See Claire Wardle, PhD and Hossein Derakhshan, *Information Disorder: Toward an interdisciplinary framework for research and policy making,* Council of Europe report DGI(2017)09 (Sep. 2017), 13, available at https://shorensteincenter.org/wp-content/uploads/2017/10/PREMS-162317-GBR-2018-Report-de%CC%81sinformation.pdf.

[942] @Twitter, Twitter (Oct. 21, 2010, 9:03AM), https://twitter.com/Twitter/status/1318900658217648128.

[943] Vijaya Gadde and Kayvon Beykpour, Additional steps we're taking ahead of the 2020 US Election, TWITTER, (Oct. 9, 2020), https://blog.twitter.com/en_us/topics/company/2020/2020-election-changes.html.

[944] Civic integrity policy, TWITTER, (Oct. 2020), https://help.twitter.com/en/rules-and-policies/election-integrity-policy (last visited Dec. 30, 2020).

(1) misleading information about how to participate in an election or other civic process;

(2) suppression and intimidation, defined as "*false or misleading information intended to intimidate or dissuade people from participating in an election or other civic process*";

(3) misleading information about outcomes of an election or another civic process and;

(4) false or misleading affiliation, aiming at preventing the creation of fake accounts misrepresent their affiliation, or sharing content falsely representing its affiliation, to a candidate, elected official, political party, electoral authority, or government entity, unless the account is a parody account within the meaning of Twitter's parody, newsfeed, commentary, and fan account policy.

Making inaccurate statements about an elected or appointed official, candidate, or political party, publishing "*polarizing, biased, hyperpartisan*" content, discussing public polling information or voting and audience participation for "*competitions, game shows, or other entertainment purposes*"[945] and using the platform pseudonymously or as a parody, commentary or fan account "*to discuss elections or politics*" is however not forbidden under this policy. In September 2020, Twitter had decided to place a public interest notice on

[945] This definition strikes as being oddly worded, as it can be argued that Twitter can entirely be used for entertainment purposes.

three of President Trump's tweets for violating the platform's Civic Integrity Policy, because they had "*encourag*[*ed*] *people to potentially vote twice."*[946]

Democracy and health sometimes collude. On April 23, 2020, President Trump said during one of his COVID-19 briefings that "*ultraviolet or just very powerful light*", when "*hit*[*ing*] *the body*", could fight the virus, and also said that injecting disinfectants could have the same benefic effect.[947] This statement compelled the Emergency Management Division of the Washington Military Department to post on Twitter: "*Please don't eat tide pods or inject yourself with any kind of disinfectant."*[948] However, no social media sites made the decision to remove the original statement about the alleged virus-fighting properties of bleach made by President Trump.[949] The next day, Twitter said that the video clips of the briefing did not violate its COVID-19 misinformation policy, but, however, had blocked the hashtags "InjectDisinfectant" and "InjectingDisinfectant."[950] A tweet from the official

---

[946] @TwitterSafety, TWITTER, (Sep 3, 2020, 2:32 PM), https://twitter.com/TwitterSafety/status/1301588773864534016. The three tweets at stake, @realDonaldTrump, TWITTER, (Sep. 3, 2020, 10:32 AM), https://twitter.com/realDonaldTrump/status/1301528522582786049, read: "*Based on the massive number of Unsolicited & Solicited Ballots that will be sent to potential Voters for the upcoming 2020 Election, & in order for you to MAKE SURE YOUR VOTE COUNTS & IS COUNTED, SIGN & MAIL IN your Ballot as EARLY as possible. On Election Day, or Early Voting,.. .....go to your Polling Place to see whether or not your Mail In Vote has been Tabulated (Counted). If it has you will not be able to Vote & the Mail In System worked properly. If it has not been Counted, VOTE (which is a citizen's right to do).* ***If your Mail In Ballot arrives.... ... .after you Vote, which it should not, that Ballot will not be used or counted in that your vote has already been cast & tabulated.*** *YOU ARE NOW ASSURED THAT YOUR PRECIOUS VOTE HAS BEEN COUNTED, it hasn't been "lost, thrown out, or in any way destroyed". GOD BLESS AMERICA!!!"* (my emphasis).

[947] William J. Broad and Dan Levin, *Trump Muses About Light as Remedy, but Also Disinfectant, Which Is Dangerous*, THE NEW YORK TIMES, (Apr.24, 2020), https://www.nytimes.com/2020/04/24/health/sunlight-coronavirus-trump.html.

[948] @waEMD, Twitter, (Apr 23, 2020, 7:57 PM), https://twitter.com/waEMD/status/1253473167017865216.

[949] Sheera Frenkel and Davey Alba. *Trump's Disinfectant Talk Trips Up Sites' Vows Against Misinformation*, THE NEW YORK TIMES, Apr. 30, 2020), https://www.nytimes.com/2020/04/30/technology/trump-coronavirus-social-media.html?action=click&module=Top%20Stories&pgtype=Homepage.

[950] Reuters, *Twitter Allows Trump COVID-19 Disinfectant Videos, Blocks '#InjectDisinfectant'*, THE NEW YORK TIMES, (Apr. 24, 2020), https://www.nytimes.com/reuters/2020/04/24/technology/24reuters-health-coronavirus-trump-twitter.html.

account of the Twitter communication team, read: "*Context matters. Tweets that are clearly satirical in nature, or that discuss or report on timely issues about #COVID19 without calls to action generally do not break our rules,*"[951] adding "*We will not require every Tweet that contains incomplete or disputed information about #COVID19 to be removed. As an open service, this is not scalable and limits active discussion. However, when content does break our rules, we're taking action*."[952] A few hours earlier, the President had publicly presented his speech about the virtues of bleach to cure the COVID-19 virus as being "*sarcastic*,"[953] appearing thus to imply as his comments about the curing virtues of bleach were not meant to be taken at face value.

Twitter had informed its users on March 23, 2020 that "*COVID-19 is affecting* [*its*] *content moderation capacities in unique ways*" and that the platform is "*adjusting to meet the challenge…focus*[*ing*] *on content that has the highest potential of directly causing physical harm.*"[954] In a March 27, 2020 "*update on our content moderation work*," Twitter explained that it had broadened its definition of harm "*to address content that goes directly against guidance from authoritative sources of global and local public health information*."[955] It would now require people to remove messages including inaccurate guidance such as "social distancing is not effective" or messages encouraging not to respecting social

[951] @TwitterComms, (Apr. 24, 2020, 3:30 PM), https://twitter.com/TwitterComms/status/1253768351723085824.
[952] @TwitterComms, (Apr. 24, 2020, 3:30 PM), https://twitter.com/TwitterComms/status/1253768352478031872.
[953] Poppy Noor, *Was Trump being 'sarcastic' with his disinfectant comments? You decide*, THE GUARDIAN, (Apr, 24 Apr 2020, 15.51 EDT), https://www.theguardian.com/world/2020/apr/24/trump-disinfectant-bleach-sarcastic.
[954] @TwitterSafety, Twitter (March. 26, 2020, 6:46 PM), https://twitter.com/TwitterSafety/status/1242221130326241280.
[955] *An update on our content moderation work*, TWITTER BLOG (March 27, 2020), https://blog.twitter.com/en_us/topics/company/2020/covid-19.html#moderation.

distances recommendation "*in areas known to be impacted by COVID-19 where such measures have been recommended by the relevant authorities*." It would also ask users to remove messages describing COVID-19 cures, even if they "*are not immediately harmful but are known to be ineffective*," messages *"shared with the intent to mislead others, even if made in jest, such as "coronavirus is not heat-resistant — walking outside is enough to disinfect you*" or "*use aromatherapy and essential oils to cure COVID-19*, " description of ineffective treatments, or denials of established scientific facts about how the virus is being transmitted. Even a parody account could be removed if used to post "*specific and unverified claims made by people impersonating a government or health official or organization*," giving as example "*a parody account of an Italian health official stating that the country's quarantine is over*." Twitter posted few days later that it would "*prioritize removing content when it has a clear call to action that could directly pose a risk to people's health or well-being, but we want to make* it *clear that we will not be able to take enforcement action on every Tweet that contains incomplete or disputed information about COVID-19*."[956]

While the new guidelines gave as example of posts warranting deletion those denying established scientific facts about transmission of the virus, such as "*COVID-19 does not infect children because we haven't seen any cases of children being sick,*" Twitter nevertheless chose not to delete one of Elon Musk's tweets stating that "*Kids are essentially immune, but elderly with existing conditions are vulnerable. Family gatherings with close*

[956] *Coronavirus: Staying safe and informed on Twitter*, TWITTER BLOG, (Apr. 3, 2020), https://blog.twitter.com/en_us/topics/company/2020/covid-19.html#definition.

contact between kids & grandparents probably most risky."[957] Tweets from Brazilian President Jair Bolsonaro were however deleted, as was one, posted by former New York mayor and President Trump personal attorney Rudy Giuliani, stating that "*Hydroxychloroquine has been shown to have a 100% effective rate treating COVID-19*."[958] Giuliani's account was even temporarily locked, showing how serious the threat to global health constitute such statements.

Twitter company announced on April 22, 2020, that, since the introduction of the policy March 18, the company had removed "*over 2,230 Tweets containing misleading and potentially harmful content. Our automated systems have challenged more than 3.4 million accounts targeting manipulative discussions around COVID-19*."[959] Twitter posted on April 1 that "[*s*]*ince introducing these policies*, [it had] *removed more than 1,100 tweets containing misleading and potentially harmful content from Twitter. [Its] automated systems have also challenged more than 1.5 million accounts which were targeting manipulative discussions around COVID-19*,"[960] and updated these figures three weeks later to announce that over 2,230 Tweets containing misleading and potentially harmful content had been removed so

[957] @elonmusk, Twitter (March 19, 2020, 5:55 PM), https://twitter.com/elonmusk/status/1240758710646878208. See also Ina Fried, *Twitter lets Musk's coronavirus misinformation stand*, AXios (March 20, 2020), https://www.axios.com/twitter-lets-musks-coronavirus-misinformation-stand-0f05b1fa-d1e7-4d9b-91f7-95c7311748d1.html.
[958] A screenshot of the March 27, 2020 tweet can be seen at Tommy Christopher, *Twitter DELETES Rudy Giuliani Tweet Featuring Coronavirus Misinformation and False Attack on Gov. Whitmer* MEDIAITE, March 28, 2020, 9:08 am), https://www.mediaite.com/news/twitter-deletes-rudy-giuliani-tweet-featuring-coronavirus-misinformation-and-false-attack-on-gov-whitmer/.
[959] @TwitterSafety, Twitter, (Apr 22, 2020, 3:35 PM), https://twitter.com/TwitterSafety/status/1253044734416711680.
[960] @TwitterSafety, Twitter, (Apr 1, 2020, 2:20 PM), https://twitter.com/TwitterSafety/status/1245415840440143873.

far, and that Twitter's automated systems have challenged more than 3.4 million accounts targeting manipulative discussions around COVID-19.[961]

Twitter announced on May 11, 2020, that it will start labeling tweets *"containing potentially harmful, misleading information related to COVID-19*."[962] Three categories of information are targeted by this initiative:

> (1) misleading information, which Twitter defines as "*statements or assertions that have been confirmed to be false or misleading by subject-matter experts, such as public health authorities;*"
>
> (2) disputed claims, defined as "*statements or assertions in which the accuracy, truthfulness, or credibility of the claim is contested or unknown,*" and;
>
> (3) unverified claims, which are "*information (which could be true or false) that is unconfirmed at the time it is shared.*"

Misleading information are tagged and removed, disputed claims are tagged, and a warning is while, while no action is taken on unverified claims. As such, Twitter is cataloguing the vast "marketplace of ideas" published on its platform every day and deems false statements unworthy of its corporate protection. The benchmark for making such a decision is, in this case, public health.

---

[961] @TwitterSafety, Twitter, (Apr 22, 2020, 3:35PM), https://twitter.com/TwitterSafety/status/1253044734416711680

[962] Yoel Roth and Nick Pickles, *Updating our Approach to Misleading Information*, TWITTLER BLOG, (May 11, 2020), https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html.

Facebook announced on April 16, 2020, that it had removed "*hundreds of thousands of pieces of misinformation that could lead to imminent physical harm…*[*including*] *harmful claims like drinking bleach cures the virus and theories like physical distancing is ineffective in preventing the disease from spreading.*" The social media company also announced that it fights the spread of misinformation and harmful content about the virus on its apps by working with more than sixty fact-checking organizations reviewing and rating content in more than fifty languages around the world.[963] If fact-checkers rate a particular content as being false, Facebook then reduces its distribution and shows "*warning labels with more context*." Forty million posts related to COVID-19 displayed such labels in March 2020 alone, leading to 95% of the users deciding not to view the content. On August 5, 2020, Facebook announced it had deleted a post published by Donald Trump because it contained "*harmful Covid misinformation.*" The post was a clip from an interview the President had given to the *Fox & Friends* television show, where he claimed that children are "*almost immune*" to COVID-19.[964] Facebook spokesperson explained that the decision had been made because "*the video includes false claims that a group of people is immune from COVID-19 which is a violation of our policies around harmful COVID misinformation*."[965] Twitter announced the same day it had temporarily blocked the Trump election campaign @TeamTrump Twitter account which had posted the same clip. A Twitter spokesperson

---

[963] Guy Rosen, VP Integrity, *An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19*, FACEBOOK, (Apr. 16, 2020), https://about.fb.com/news/2020/04/covid-19-misinfo-update.
[964] *Facebook and Twitter restrict Trump accounts over 'harmful' virus claim*, BBC NEWS (Aug. 6, 2020), https://www.bbc.com/news/election-us-2020-53673797#, *Facebook, Twitter, YouTube Pull Trump Posts Over Coronavirus Misinformation*, THE NEW YORK TIMES (Aug. 6, 2020), https://www.nytimes.com/reuters/2020/08/06/technology/06reuters-facebook-trump.html?searchResultPosition=1.
[965] Shannon Bond, *Twitter, Facebook Remove Trump Post Over False Claim About Children And COVID-19,* NPR, (Aug. 5, 2020), https://www.npr.org/2020/08/05/899558311/facebook-removes-trump-post-over-false-claim-about-children-and-covid-19.

explained that the tweet had violated the Twitter Rules on COVID-19 misinformation, and that "[*t*]*he account owner will be required to remove the Tweet before they can Tweet again*."

We discussed so far only misinformation about the COVID-19 pandemic published by individuals. However, bots are often at the origin of misinformation. A team of researchers from the Carnegie Mellon University studied a sample of 200 million tweets discussing coronavirus or COVID-19 and found out that 82% of top fifty influential retweeters are bots, while 62% of top 1,000 influential retweeters are bots.[966] If one is ready to accept abuse and hate speech on social media for the sake of protecting a vibrant "marketplace of ideas," this is disheartening news, as bots are saturating the marketplace using artificial intelligence (AI). Indeed, a "bot" is defined by the Merriam-Webster dictionary as "*a computer program that performs automatic repetitive tasks*,"[967] and are thus efficient if programmed to repeatedly post messages on social media. Should bots be protected by the First Amendment?[968] Senator Dianne Feinstein [D-CA] introduced in July 2019 S.2125, the Bot Disclosure and Accountability Act of 2019.[969] It aimed at "*mitigating the deceptiveness of social media bots*" by requiring social media providers to establish policies and procedures requiring that users of the platforms operating bots to publish

---

[966] Virginia Alvino Young, *Nearly Half Of The Twitter Accounts Discussing 'Reopening America' May Be Bots*, CARNEGIE MELLON UNIVERSITY SCHOOL OF COMPUTER SCIENCE, (May 20, 2020), https://www.cs.cmu.edu/news/nearly-half-twitter-accounts-discussing-%E2%80%98reopening-america%E2%80%99-may-be-bots?mod=article_inline (last visited Dec. 30, 2020).
[967] *Bot*, MERRIAM-WEBSTER DICTIONARY, https://www.merriam-webster.com/dictionary/bot, (last visited Dec. 30, 2020).
[968] See Jared Schroeder, *Are bots entitled to free speech?*, COLUMBIA JOURNALISM REVIEW, (May 24, 2018), https://www.cjr.org/innovations/are-bots-entitled-to-free-speech.php, arguing that if courts would focus on whether a bot is publisher, they would not be protected as journalists, but if the courts would focus on the content published, they may be protected, "*particularly if their content can be seen as a public good*." See also Jamie Lee Williams, *Cavalier Bot Regulation and the First Amendment's Threat Model*, KNIGHT FIRST AMENDMENT INSTITUTE AT COLUMBIA UNIVERSITY, (Aug. 21 2019), https://knightcolumbia.org/content/cavalier-bot-regulation-and-the-first-amendments-threat-model,
[969] S.2125, Bot Disclosure and Accountability Act of 2019, 116th Cong. (2019).

speech on the platform disclose accounts which are bots, defined as "*any automated software program or process intended to impersonate or replicate human activity online*."[970] The FTC would be in charge of promulgating such regulations under its rulemaking authority. Such notice should be "*clear and conspicuous*… [*and*] *in clear and plain language.*"[971] The FTC would also be in charge of enforcing the rule under its enforcement authority. However, if one looks at how difficult it is for the FTC to generally enforce its rules on use of endorsement in social media, where material connection between an endorser and a seller must be "*clearly and conspicuously*" disclosed,[972] one wonders how the FTC could have enforced the bot disclosing rule, at least without having been provided significant resources. We will return to the issue of the right of bots to speak freely later in this article.

Spreading fake information about disease is also sometimes a way to blame a particular group for it, such as Jewish people. Former British soccer player and conspiracy theorist David Icke blamed the 5G mobile phone network for spreading the COVID-19 virus, but also claimed in a *YouTube* video that the Rothschilds had been involved in planning the coronavirus outbreak. The video was viewed 5.9 million times, making it the twenty seventh most viewed video about coronavirus on the platform at the time.[973] Twitter, however, did not follow suit and did not block Icke's Twitter account. However, false

---

[970] Sec. 4 (a)(1)(A).
[971] Sec. 4 (c)(1).
[972] CFR §255.5.
[973] *Coronavirus: David Icke kicked off Facebook*, THE BBC, (May 1, 2020), https://www.bbc.com/news/technology-52501453. The Non-profit organization Center for Countering Digital Hate (CCDH) described him in a report a "professional conspiracy theorist" and called for his removal from the social media platforms, using the hashtag #DeplatformIcke. *See #DeplatformIcke How Big Tech powers and profits from David Icke's lies and hate, and why it must stop*, p. 3 and p. 4, https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9_db8ff469f6914534ac02309bb488f948.pdf.

information and hate speech sometimes collude. What about information which is obviously false? The Holocaust did occur, it killed 6 million Jews,[974] it is well documented, and there was still, in 2020, living individuals who have either been deported to the camps,[975] or had visited them shortly after the camp's liberation by Allied troops. In a July 2018 interview, Mark Zuckerberg said he found Holocaust-denying "*deeply offensive* but added: *"… at the end of the day, I don't believe that our platform should take that down because I think there are things that different people get wrong. I don't think that they're intentionally getting it wrong*…"[976] Facebook's CEO must have later changed his mind as, on October 12, 2020, Facebook announced that it had updated its hate speech policy to prohibit any content denying or distorting the Holocaust, citing, in particular, a recent survey of U.S. adults aged 18 to 39 years old, which revealed that almost a quarter of them believed the Holocaust was a myth, that it had been exaggerated or weren't sure. It also noted that "[*i*]*nstitutions focused on Holocaust research and remembrance, such as Yad Vashem, have noted that Holocaust education is also a key component in combatting anti-Semitism*."[977] Holocaust denial is not only false, but also, arguably, hate speech. How do social media platforms deal with hate speech?

---

[974] *Documenting Numbers of Victims of the Holocaust and Nazi Persecution*, UNITED STATES HOLOCAUST MEMORIAL MUSEUM, https://encyclopedia.ushmm.org/content/en/article/documenting-numbers-of-victims-of-the-holocaust-and-nazi-persecution, (last visited Dec. 30, 2020).
[975] *First Person Podcast Series* , UNITED STATES HOLOCAUST MEMORIAL MUSEUM , https://www.ushmm.org/remember/holocaust-survivors/first-person-conversations-with-survivors/first-person (last visited Dec. 30, 2020).
[976] Kara Swisher, *Full transcript: Facebook CEO Mark Zuckerberg on Recode Decode*, VOX, (Jul 18, 2018, 11:01am EDT), https://www.vox.com/2018/7/18/17575158/mark-zuckerberg-facebook-interview-full-transcript-kara-swisher.
[977] Monika Bickert, VP of Content Policy, *Removing Holocaust Denial Content*, FACEBOOK, (Oct. 12, 2020), https://about.fb.com/news/2020/10/removing-holocaust-denial-content.

c. Hate Speech and the Platforms

What is hate speech? The European Commission noted in October 2015 its conclusions paper to its first colloquium on hate speech that:

> "[*t*]*he role of online intermediaries/platforms (e.g. Google, Facebook and Twitter) in removing hate speech is … seen as central. Clearer procedures are needed for the effective prosecution and taking down of hate speech on the internet. Dialogue should be developed at EU level with IT companies on how to address hate speech online more efficiently"* and vowed to "*initiate dialogue at EU level with IT companies and bring together businesses, national authorities and civil society to combat hate speech online, including by making it easier for users to report illegal content to companies*."[978]

Hate speech can also be spread by videos, and all social medias platforms allow videos to be posted by users on their pages. A Recommendation of the Committee of Ministers provides a definition, at least for EU Member States, "*as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin*."[979]

i. Hate Speech on Social Media

---

[978] *European Commission, Joining forces against anti-Semitic and anti-Muslim hatred in the EU: outcomes of the first Annual Colloquium on Fundamental Rights*, EUROPEAN COMMISSION, (Oct.9, 2015), p.5. http://ec.europa.eu/justice/events/colloquium-fundamental-rights-2015/index_en.htm.

[979] Recommendation No. R (97) 20 of the Committee of Ministers to member states on "hate speech" (Oct. 30, 1997), available at https://rm.coe.int/1680505d5b.

While some values, such are punishing child abuse, are almost universally acknowledged the same way around the world, others, such as morale, or ethics, which are "soft laws," are more fluid, and the private law of the social media platforms may not align with the laws of a particular country. Chief Justice Burger had written in *Miller v. California*, that "[*i*]*t is neither realistic nor constitutionally sound to read the First Amendment as requiring that the people of Maine or Mississippi accept public depiction of conduct found tolerable in Las Vegas, or New York City.*[980] Chief Justice Burger further noted that, in *Jacobellis v. Ohio*,[981] Justice Brennan and Justice Goldberg had:

> "*argued that application of "local" community standards would run the risk of preventing dissemination of materials in some places because sellers would be unwilling to risk criminal conviction by testing variations in standards from place to place… The use of "national" standards, however, necessarily implies that materials found tolerable in some places, but not under the "national" criteria, will nevertheless be unavailable where they are acceptable. Thus, in terms of danger to free expression, the potential for suppression seems at least as great in the application of a single nationwide standard as in allowing distribution in accordance with local taste*…"

One of the positive aspects of social media is that users may discover the culture of people around the world. Seeing our own cultures in the eyes of others may allow us to understand our own bias. For instance, we now know that 'blackface' is offensive. Twitch updated its 'Hateful Conduct and Harassment' Guidelines in December 2020.[982] The new

[980] Miller v. California, 413 U.S. 15, 32 (1973).
[981] Jacobellis v. Ohio, 378 U. S. 184 (1964).
[982] *Hateful Conduct and Harassment* [NEW], TWITCH, https://www.twitch.tv/p/legal/community-guidelines/harassment/20210122/, (last visited Dec. 30, 2020).

Guidelines, who went into effect on January 22, 2021, forbids the posting of "[*s*]*peech, imagery, or emote combinations that dehumanize or perpetuate negative stereotypes and/or memes*", which includes "[*b*]*lack/brown/yellow/redface*." When Facebook updated its hate speech policy in August 2020 to include "*content depicting blackface, or stereotypes about Jewish people controlling the world*" as hate speech,[983] the Belgian far right party Vlaams Belang denounced the move as "*censorship*" and an attack on Belgian traditions.[984] Indeed, blackface is used in Belgium to portray "*Le Sauvage*"[985] and "*Zwarte Piet* (Black Peter), Sinterklaas (Santa Claus)'s helper,[986] who is also famous in The Netherlands.[987] Vlaaams Belang announced it would publish a resolution with other political parties in the European Parliament" *to condemn arbitrary censorship on social media and will urge the Commission to take action*." However, such customs, which may reflect the "local taste" of a local community, within the meaning of *Jacobellis*, may now be shunned by the "local taste" of the global social media community.

What constitute "hate speech" on social media is not, however, always easy to ascertain, especially as social media users develop their own hate speech codes, symbols and shortcuts to communicate hate even more efficiently online. The Anti-Defamation

---

[983] Guy Rosen, *Community Standards Enforcement Report*, August 2020, FACEBOOK, (Aug. 11, 2020), https://about.fb.com/news/2020/08/community-standards-enforcement-report-aug-2020.
[984] *Facebook en Instagram censureren Zwarte Piet: Vlaams Belang veroordeelt censuur en aanval op onze tradities*, VLAAMS BELANG, (Aug, 12, 2020), https://www.vlaamsbelang.org/facebook-en-instagram-censureren-zwarte-piet-vlaams-belang-veroordeelt-censuur-en-aanval-op-onze-tradities, (last visited Dec. 30, 2020).
[985] Lindsey Johnstone, *Watch: Blackface in Belgium back in the spotlight after controversial parade*, EURONEWS, (Aug. 27, 2019), https://www.euronews.com/2019/08/26/watch-blackface-in-belgium-back-in-the-spotlight-after-controversial-parade.
[986] *What's the issue with Zwarte Piet ?*, THE BRUSSELS TIMES, (Dec. 4, 2019), https://www.brusselstimes.com/all-news/belgium-all-news/81413/sinterklaas-who-is-zwarte-piet-belgium-the-netherlands-december-blackface/.
[987] Philip Huff, *The False Innocence of Black Pete* (Dec. 5, 2019), THE PARIS REVIEW, https://www.theparisreview.org/blog/2019/12/05/the-false-innocence-of-black-pete/.

League added in June 2016 the (((Echo))) symbol, used to call attention to Jewish social media users, to its database of white supremacist symbols.[988] It was first used in 2014 in an anti-Semitic podcast, which applied a specific sound effect to Jewish names to make them echo.[989] Jon Weisman, Deputy Washington Editor for the *New York Times*, wrote an article in May 2016 explaining how he first discovered the use of the echo symbol when it was associated with his name by a Twitter user.[990] The tweet was "*Hello ((Weisman)).*" Mr. Weisman explained that *"@CyberTrump was responding to my recent tweet of an essay by Robert Kagan on the emergence of fascism in the United States*." When Mr. Weisman asked for explanation about the symbol, the (anonymous) Twitter user answered: "*What, ho, the vaunted Ashkenazi intelligence, hahaha!*", adding "*It's a dog whistle, fool. Belling the cat for my fellow goyim*." This led to a multitude of anti-Semitic Twitter messages from various users, for the most from users specifying in their bio that they support Donald Trump.[991] Mr. Weisman left Twitter on June 8, 2016, vowing to move to Facebook "*where at least people need to use their real names and can't hide behind fakery to spread their hate*."[992] He later explained that his departure was triggered by Twitter's failure to enforce its own

---

[988] Press Release, Anti-Defamation League, *ADL to Add (((Echo))) Symbol, Used by Anti-Semites on Twitter, to Online Hate Symbols Database, Joins Swastika, Wolfsangel, and Blood Drop Cross*. (June 6, 2016), http://www.adl.org/press-center/press-releases/anti-semitism-usa/adl-to-add-echo-symbol-used-by-anti-semites-on-twitter-to-online-hate-symbol-database.html. ADL added the 'OK' sign to the list in 2019, as it is sometimes used as a "*sincere expression of white supremacy*." See *OK hand sign added to list of hate symbols*, BBC, (Sep. 27, 2019), https://www.bbc.com/news/newsbeat-49837898.

[989] Amanda Hess, *They Punctuate Their Messages With Subtle Symbols of Hatred,* THE NEW YORK TIMES, (June 10, 2016), http://www.nytimes.com/2016/06/11/arts/for-the-alt-right-the-message-is-in-the-punctuation.html?_r=0.

[990] Jon Weisman, *The Nazi Tweets of 'Trump God Emperor'*, THE NEW YORK TIMES, (May 26, 2016), http://www.nytimes.com/2016/05/29/opinion/sunday/the-nazi-tweets-of-trump-god-emperor.html?_r=0.

[991] For example, one of these Twitter users sent to Mr. Weisman an image of the gates of the Auschwitz concentration camp, where "Arbeit Macht Frei" was replaced by "Machen Amerika Great."

[992] (((Jon Weisman))) (@JonathanWeisman), Twitter (June 8, 2016, 9:34 AM), https://twitter.com/jonathanweisman/status/740537561211228160.

Terms of Service governing hateful conduct and harassment, even pornography.[993] Twitter finally deleted some of the accounts which Mr. Weisman had reported, but he noted that some other accounts, similarly hateful, had been allowed to stay. Failure to delete all the accounts used to tweet hateful messages has been reported by other victims of such messages.[994] Mr. Weisman further reported that after he announced he left his Twitter account, he opened another one, choosing the handle @Jew_Hater, to make the point that the social media site would allow such blatant anti-Semitic handle to be chosen and used. On June 18, 2020, Facebook removed Trump campaign ads stating: *"Dangerous MOBS of far-left groups are running through our streets and causing absolute mayhem. They are DESTROYING our cities and rioting – it's absolute madness … Please add your name IMMEDIATELY to stand with your President and his decision to declare ANTIFA a Terrorist Organization."* The ads featured an inverted red triangle bordered in black, a symbol used by Nazis to designate political prisoners during World War II. The Twitter account of non-profit organization *Bend the Arc* published a photograph of the uniform of concentration camp prisoner bearing such symbol sewn on a sleeve and explained its significance.[995] The Trump campaign denied the ads referred to this symbol, calling it instead an emoji.[996] Facebook removed the ad as the image violated Facebook policy against organized hate, as

---

[993] Jonathan Weisman, *Why I Quit Twitter — and Left Behind 35,000 Followers*, THE NEW YORK TIMES, (June 10, 2016), http://www.nytimes.com/2016/06/10/insider/why-i-quit-twitter-and-left-behind-35000-followers.html?_r=0. Mr. Weisman wrote that a Twitter user even sent him an explicit GIF.
[994] See for example Sara Ashley O'Brien, *Muslim woman deluged by 'hate tweets' after helping Homeland Security panel*, CNN MONEY, (June 19, 2016: 7:54 PM ET), http://money.cnn.com/2016/06/19/technology/laila-alawa-trolling/,
[995] @jewishaction, Twitter, (June 18, 2020, 1:07AM), https://twitter.com/jewishaction/status/1273482511918616578.
[996] Annie Karni, *Facebook Removes Trump Ads Displaying Symbol Used by Nazis*, THE NEW YORK TIMES (June 18, 2020, Updated June 19, 2020, 1:00 a.m. ET), https://www.nytimes.com/2020/06/18/us/politics/facebook-trump-ads-antifa-red-triangle.html

its policy "*prohibits using a banned hate group's symbol to identify political prisoners without the context that condemns or discusses the symbol*."[997]

Facebook used to have a humor exception to its hate speech policy. The second progress report on the platform's civil rights audit, published in June 2019, recommended "*remov*[*ing*] *humor as an exception to the hate speech policy (and ensure that any future humor-related carve-outs are limited and precisely and objectively defined*), because the policy did not provide and objective, or clearly defined standard of what is "humor," noting that "*what one user* (*or content reviewer*) *may find humorous may be perceived as a hateful, personal attack by another*." As such, the humor exception to Facebook's hate speech policy "*likely contributes to enforcement inconsistencies and runs the risk of allowing the exception to swallow the rule*."[998] The hate speech rules no longer carve a humor exception.

ii. How Social Media Platforms Address Hate Speech

Hate speech may be prevented, by changing the private law of the platforms, just as a government might do so. For instance, Facebook Community Standards on Dangerous Individuals and Organizations bar from the site "*any organizations or individuals that proclaim a violent mission or are engaged in violence, from having a presence on Facebook*" including "*organizations or individuals involved in… [t]errorist activity; [o]rganized hate; [m]ass or serial murder; [h]uman trafficking [and] [o]rganized violence or criminal*

---

[997] Julia Carrie Wong, *Facebook removes Trump re-election ads that feature a Nazi symbol,* THE GUARDIAN, (June 18 2020 18.29 EDT), https://www.theguardian.com/technology/2020/jun/18/facebook-removes-trump-re-election-ads-that-feature-a-nazi-symbol, Bobby Allyn, *Facebook Removes Trump Ads With Symbol Used By Nazis. Campaign Calls It An 'Emoji'*, NPR, (June 18, 2020 2:58 PM ET), https://www.npr.org/2020/06/18/880377872/facebook-removes-trump-political-ads-with-nazi-symbol-campaign-calls-it-an-emoji.

[998] Facebook's Civil Rights Audit – Progress Report (June 30, 2019), p.12, https://about.fb.com/wp-content/uploads/2019/06/civilrightaudit_final.pdf.

activity."[999] Facebook "*also remove content that expresses support or praise for groups, leaders, or individuals involved in these activities*."[1000] The former Article 3 of Facebook's Statement of Rights and Responsibilities about "Safety," and the former article 3.6 forbade Facebook users to post "content *that: is hate speech, threatening, or pornographic; incites violence; or contains nudity or graphic or gratuitous violence*."[1001]

Hate speech is defined:

> "*as a direct attack on people based on… protected characteristics — race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability*." Facebook also "*protect*(*s*) *against attacks on the basis of age when age is paired with another protected characteristic, and also provide certain protections for immigration status. We define attack as violent or dehumanizing speech, harmful stereotypes, statements of inferiority, or calls for exclusion or segregation*."[1002]

The Russian social network VK forbids using the platform to "*propagate*[*e*] *and/or incit*[*e*] *racial, religious, or ethnic hatred or hostility, including hatred or hostility towards a specific gender, orientation, or any other individual attributes or characteristics of a person (including those concerning a person's health).*"[1003] TikTok's Community Guidelines state

---

[999] FACEBOOK COMMUNITY STANDARDS, *2. Dangerous Individuals and Organizations*, https://www.facebook.com/communitystandards/dangerous_individuals_organizations, (last visited Dec. 30, 2020).
[1000] Id.
[1001] *Statement of Rights and Responsibilities*, FACEBOOK, https://www.facebook.com/legal/terms/previous (last visited Dec. 30, 2020).
[1002] Facebook Community Standard, Hate Speech, https://www.facebook.com/communitystandards/hate_speech (last visited Dec. 30, 2020).
[1003] *VK Terms of Service*, paragraph 6.3.4 (e) VK, https://vk.com/terms (last visited Dec. 30, 2020).

that the Chinese-owned platform "*do*[*es*] *not tolerate content that attacks or incites violence against an individual or a group of individuals on the basis of protected attributes.*[1004] *We do not allow content that includes hate speech, and we remove it from our platform. We also suspend or ban accounts that have multiple hate speech violations."*[1005] Speech which excludes a particular user appears to be "hate speech" based on an aspect of his or her or them personality and identity rights, a particular of an intimate and important part of themselves. Twitter updated its policy against hateful conduct in December 2020, adding to the list of prohibited language dehumanizing people because of their race, ethnicity, or national origin.[1006]

The social media platforms identify hate speech in two ways, though reports of its users and by proactive detection by technology.[1007] Transgressing these rules lead to takedown of the speech, and in some cases, even banning from the platform.[1008] Twitter announced in September 2018 that it had permanently banned Alex Jones and his *InfoWars*

[1004] They are defined as: race; ethnicity; national origin; religion; caste; sexual orientation; sex; gender; gender identity; serious disease or disability and immigration status.
[1005] *Community Guidelines*, TIKTOK, https://www.tiktok.com/community-guidelines?lang=en.
[1006] Twitter Safety, *Updating our rules against hateful conduct*, TWITTER BLOG, (Dec. 2, 2020), https://blog.twitter.com/en_us/topics/company/2019/hatefulconductupdate.html. The rules against hateful conduct had already been expanded in July 2019 to include language that dehumanizes others on the basis of religion or caste, and again in March 2020, to include language that dehumanizes on the basis of age, disability, or disease.
[1007] *See* Facebook's Civil Rights Audit, p. 42. Facebook reported, as of March 2019, that 65% of the hate speech removed had been proactively detected, and that thanks to new technologies, including artificial intelligence, this percentage rose to 89% of removals as of March 2020. The report explained that "*Facebook relies on human reviewers to assess context (e.g., is the user using hate speech for purposes of condemning it) and also to assess usage nuances in ways that artificial intelligence cannot*."
[1008] This recalls banishment, used in the Middle Ages, less frequently in modern times, to punish. In France, banishment was used to punish an individual without having to kill. While local courts had the power to ban an individual withing their jurisdictions, the royal courts had the power to ban outside the kingdom of France. Banishment could be temporary or permanent. *See* JEAN-MARIE CARBASSE, HISTOIRE DU DROIT PÉNAL ET DE LA JUSTICE CRIMINELLE, 289-290, (Presses Universitaires de France ed. (2d ed. 2009).

website from Twitter and Periscope.[1009] Jones had repeatedly stated, including on social media, that Sandy Hook massacre of schoolchildren was a hoax. Twitter had banned Milo Yiannopoulos in 2016, following a hateful campaign against comedian Leslie Jones, set to star in a new "*Ghostbusters*" movie.[1010] Facebook announced in March 2019 that it will ban "*praise, support and representation of white nationalism and white separatism on Facebook and Instagram*," and vowed to be "*better and faster at finding and removing hate from* [*its*] *platform.*[1011] It also announced that it would now direct users searching "*for terms associated with white supremacy to resources focused on helping people leave behind hate groups. People searching for these terms will be directed to Life After Hate, an organization founded by former violent extremists that provides crisis intervention, education, support groups and outreach*."[1012] A few weeks after this announcement, Facebook banned from its platform Alex Jones and Louis Farrakhan,[1013] Milo Yiannopolous, and Laura Loomer,[1014] all considered to have posted hate speech on the platforms. More recently, Facebook

---

[1009] @TwitterSafety, Twitter (Sep 6, 2018, 4:47 PM), https://twitter.com/TwitterSafety/status/1037804427992686593, see also Avie Schneider, *Twitter Bans Alex Jones And InfoWars; Cites Abusive Behavior*, NPR, (Sept. 6, 2018, 5:34 PM ET), https://www.npr.org/2018/09/06/645352618/twitter-bans-alex-jones-and-infowars-cites-abusive-behavior.

[1010] Mike Isaac, *Twitter Bars Milo Yiannopoulos in Wake of Leslie Jones's Reports of Abuse*, THE NEW YOK TIMES, (July 20, 2016), https://www.nytimes.com/2016/07/20/technology/twitter-bars-milo-yiannopoulos-in-crackdown-on-abusive-comments.html.

[1011] *Standing Against Hate*, FACEBOOK NEWSROOM, (March 27, 2019), https://newsroom.fb.com/news/2019/03/standing-against-hate/.

[1012] Id.

[1013] Mike Isaac and Kevin Roose, *Facebook Bars Alex Jones, Louis Farrakhan and Others From Its Services*, THE NEW YORK TIMES (May 2, 2019), https://www.nytimes.com/2019/05/02/technology/facebook-alex-jones-louis-farrakhan-ban.html.

[1014] Dave Lee, *Facebook bans 'dangerous individuals'* BBC, (May 3, 2019), https://www.bbc.com/news/technology-48142098. Laura Loomer had also been permanently banned by Twitter in November 218, following tweets against Minnesotan Democratic Representative Ilhan Omar, who had just been elected, leading to her decision to handcuff herself to the company's New York headquarters in protest, *see Far-right activist Laura Loomer handcuffed herself to Twitter's NYC headquarters*, THE VERGE (Nov 29, 2018, 4:55pm EST), https://www.theverge.com/2018/11/29/18118529/far-right-activist-laura-loomer-twitter-protest-handcuff-nyc-hq.

announced in August 2020 that it had expanded its "*Dangerous Individuals and Organizations policy to address organizations and movements that have demonstrated significant risks to public safety but do not meet the rigorous criteria to be designated as a dangerous organization and banned from having any presence on our platform*."[1015] In December 2020, Steven Bannon was banned for having called on the platform for the for the beheading of Dr Anthony Fauci and FBI director Christopher Wray.[1016] Having to use the words "banning" and "beheading" in the same phrase, about a technology which was not used twenty years ago, is chilling. However, taking down speech may be bad for a platform's public image as such actions may be interpreted by the public as an unfair and aggressive act. Sometimes, users even file lawsuits. In one instance,[1017] plaintiff, an "original Christian ministry music" artist, had posted one of her songs on her YouTube account, which was taken down by the platform, allegedly because plaintiff had used a robot/spider to make appear that the video had been accessed multiple times, which is prohibited under YouTube's Terms of Service. The video was replaced by a statement that the video had violated the website's terms of service and a link to a page providing examples of reasons why videos might be removed by YouTube, including "Sex and Nudity," "Hate Speech," "Shocking and Disgusting," "Dangerous Illegal Acts," "Children," "Copyright," "Privacy," "Harassment," "Impersonation," and "Threats." The musician claimed that this constituted libel per quod under California law, Civ. Code, § 45a, which is

[1015] *An Update to How We Address Movements and Organizations Tied to Violence,* FACEBOOK (Aug. 19, 2020), https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence.
[1016] Peter Beaumont, *Steve Bannon banned by Twitter for calling for Fauci beheading*, THE GUARDIAN, (Nov. 6, 2020, 16.21 EST, first published Nov.6 2020 06.30 EST), https://www.theguardian.com/us-news/2020/nov/06/steve-bannon-banned-by-twitter-for-calling-for-fauci-beheading.
[1017] Bartholomew v. YouTube, LLC, 17 Cal. App. 5th 1217, 225 Cal. Rptr. 3d 917, 2017 Cal. App. LEXIS 1070, 2017 WL 5986446.

defamatory language not libelous on its face but having caused special damage to the Plaintiff, defined by the law[1018] as damages "*that plaintiff alleges and proves that he or she has suffered in respect to his or her property, business, trade, profession, or occupation, including the amounts of money the plaintiff alleges and proves he or she has expended as a result of the alleged libel, and no other.*" In a similar case,[1019] Google had removed for the same reason the song music video, "*Luv ya Luv ya Luv ya*" that plaintiffs had posted on YouTube, had reposted it to another, private, without view count, "likes," or comments, and had replaced it by a statement that the video's "content" violated the platform's Terms of Service. Plaintiffs sued, *inter alia*, for libel, but the court dismissed this claim as YouTube's statement was not libelous on its face and thus libel per se, but instead libel per quod. As plaintiffs had not pleaded special damages, their libel claims were dismissed.

Sometimes, the decision to take down is viewed as a political choice. When President Trump tweeted on May 29, 2020: "*These THUGS are dishonoring the memory of George Floyd, and I won't let that happen. Just spoke to Governor Tim Walz and told him that the Military is with him all the way. Any difficulty and we will assume control but, when the looting starts, the shooting starts. Thank you!*"[1020] Twitter flagged it as breaching its glorification of violence policy but allowed it to stay published under its public-interest exception, as it may be the best interest of the public to be able to see this tweet.[1021] The

[1018] Cal. Civ. Code § 48a(d)(2).
[1019] Song Fi Inc. v. Google, Inc., 108 F. Supp. 3d 876, 879, 2015 U.S. Dist. LEXIS 75272.
[1020] @realDonaldTrump, Twitter (May 29, 2020, 12:53 AM), https://twitter.com/realDonaldTrump/status/1266231100780744704.
[1021] User have to click to read the President's tweet, and first see instead "*This Tweet violated the Twitter Rules about glorifying violence. However, Twitter has determined that it may be in the public's interest for the Tweet to remain accessible*." The warning featuring a link toward a page explaining the platform's public-interest exceptions on Twitter, see https://help.twitter.com/en/rules-and-policies/public-interest.

same tweet published on the White House's Twitter account[1022] met the same fate. The President posted a longer version of the message on Facebook.[1023] Facebook, however, decided not to delete the message, a move explained by Marc Zuckerberg on a Facebook post.[1024] He acknowledged he knew that:

> "*many people are upset that* [*Facebook has*] *left the President's posts up, but* [*Facebook's*] *position is that* [*it*] *should enable as much expression as possible unless it will cause imminent risk of specific harms or dangers spelled out in clear policies*, " adding that, "[*u*]*nlike Twitter, we do not have a policy of putting a warning in front of posts that may incite violence because we believe that if a post incites violence, it should be removed regardless of whether it is newsworthy, even if it comes from a politician*."

The authors of Facebook's Civil Rights Audit noted in the Final Report, published on July 8, 2020, that the decision not to remove these tweets, and the one posted in May 2020, calling into doubt the legality of mail-in ballots, "*ha*[*d*] *caused considerable alarm for the Auditors and the civil rights community*,"[1025] adding that this decision "*exposed a major hole in Facebook's understanding and application of civil rights*."

[1022] @WhiteHouse, Twitter (May 29, 2020, 8:17 AM), @realDonaldTrump, Twitter (May 29, 2020, 12:53 AM).
[1023] The post read: *" I can't stand back & watch this happen to a great American City, Minneapolis. A total lack of leadership. Either the very weak Radical Left Mayor, Jacob Frey, get his act together and bring the City under control, or I will send in the National Guard & get the job done right. These THUGS are dishonoring the memory of George Floyd, and I won't let that happen. Just spoke to Governor Tim Walz and told him that the Military is with him all the way. Any difficulty and we will assume control but, when the looting starts, the shooting starts. Thank you!",* Donald Trump, FACEBOOK (May 28, 2020), https://www.facebook.com/DonaldTrump/posts/10164767134275725.
[1024] Marc Zuckerberg, FACEBOOK, (May 29, 2020), https://www.facebook.com/zuck/posts/this-has-been-an-incredibly-tough-week-after-a-string-of-tough-weeks-the-killing/10111961824369871/.
[1025] *Facebook's Civil Rights Audit – Final Report* (June 30, 2019), p.10, https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf.

iii. The Hate Speech Economy

The Facebook's Civil Rights Audit Progress Report of June 2019 noted that Facebook defines an "attack" for hate speech purposes "*as violent or dehumanizing speech, statements of inferiority, expressions of contempt or disgust, or calls for exclusion or segregation.*"[1026] Indeed, Facebook rules and practices were audited for two years, from 2018 to 2020, a process suggested by Congress and civil rights organizations, accepted by Facebook. It was undertaken by Laura W. Murphy, a civil rights and civil liberties leader, working with a team from the Relman Colfax civil rights law firm, led by firm partner Megan Cacace.[1027] A final report of the audit was published in July 2020. Its chapter 3 addressed the issue of content moderation and enforcement and stated that even though Facebook's Community Standards prohibit posting hate speech, harassment, and attempts to incite violence on its platform:

> "*civil rights advocates contend that not only do Facebook's policies not go far enough in capturing hateful and harmful content, they also assert that Facebook unevenly enforces or fails to enforce its own policies against prohibited content*."[1028]

Hate speech proliferates on social media. Facebook noted in one of its hate speech transparency reports that instances of such speech, which the social media giant defines as "*violent or dehumanizing speech, statements of inferiority, calls for exclusion or segregation*

[1026] *Facebook's Civil Rights Audit – Progress Report* (June 30, 2019), p. 7, https://about.fb.com/wp-content/uploads/2019/06/civilrightaudit_final.pdf (last visited Dec. 30, 2020).
[1027] *Facebook's Civil Rights Audit – Final Report*, FACEBOOK, (July 8, 2020), available at https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf (last visited Dec. 30, 2020).
[1028] *Facebook's Civil Rights Audit,* p. 42.

based on protected characteristics, or slurs,"[1029] are augmenting.[1030] The platform has been, however, accused by human rights groups to have been used in 2018 as a propaganda tool for ethnic cleansing by the Myanmar military, posing as celebrities, by inciting murders, rapes and forced human migration,[1031] before Facebook blocked them from using the platform.[1032]

The COVID-19 pandemic forced the world to confine and the web became for many the only way to communicate with the outside world, leading to an increase in online abuse. In France, three French non-profits, L'*Union des Étudiants Juifs de France*, *Touche Pas À Mon Pote* and *SOS Homophobie*, dedicated to fighting, respectively, anti-Semitism, racism, and homophobia, claimed after the confinement due to COVID-19 that the number of racist tweets rose 40% during confinement, the number of anti-Semitic tweets rose 20%, and the number of homophobic tweets rose 48%.[1033] The report also claimed that 84% of the racist tweets which had been reported by the authors of the report had not been taken down by Twitter, a percentage which rose to 86.9% for anti-Semitic tweets, and 95.6% for homophobic tweets. When dealing with hate speech, platforms need to first define what is

[1029] Facebook specifies that "[*t*]*hese characteristics include race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disability or disease", Community Standards Enforcement Report* , FACEBOOK, https://transparency.facebook.com/community-standards-enforcement#hate-speech.
[1030] For instance, it noted in April 2020 that such content increased "*from 5.7 million pieces of content in Q4 2019 to 9.6 million in Q1 2020*.", ibid.
[1031] Paul Mozur, *A Genocide Incited on Facebook, With Posts From Myanmar's Military*, THE NEW YORK TIMES, (Oct. 15, 2018), https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.
[1032] *Removing Myanmar Military Officials From Facebook* (Aug. 28, 2018), FACEBOOK, https://about.fb.com/news/2018/08/removing-myanmar-officials : *"Today, we removed 425 Facebook Pages, 17 Facebook Groups, 135 Facebook accounts and 15 Instagram accounts in Myanmar for engaging in coordinated inauthentic behavior on Facebook.*"
[1033] *La Haine en ligne se propage pendant le confinement,* UNION DES ÉTUDIANTS DE France, https://uejf.org/wp-content/uploads/2020/05/dossier-Haine-en-Ligne.pdf. The claim was based on their own study, made by analyzing tweets sent from March 17, 2020, to April 27, 2020.

"hate speech," then identify it, and filter it or take it down. This endeavor may or may not be financially benefitting to them, as social media platforms benefit from publishing any type of content, especially viral content, even if the message is a menace for society.[1034]

Dealing efficiently with hate speech is also an economical issue for platforms which need to maintain the luster of the corporate brand. The web site *The Verge* reported in February 2015 about an internal memo that then Twitter CEO Dick Costolo sent to employees on February 2, 2015, where he took full responsibility for how poorly Twitter has responded to the issue of rampant trolling and abuse on the microblogging site. He wrote: "*We suck at dealing with abuse and trolls on the platform and we've sucked at it for years. It's no secret and the rest of the world talks about it every day. We lose core user after core user by not addressing simple trolling issues that they face every day."*[1035] Mr. Costolo sent that report just a few days before the announcement of Twitter's fourth-quarter earnings for 2014, which, incidentally, failed to impress investors, although there were no reports of a link between the lack of adequate response to abuse and lack-luster financial results.

There is a hate speech economy. The Center for Countering Digital Hate noted in its report about U.K. conspirationist David Icke that "[*s*]*ocial media platforms profit from hosting Icke too. His audience of over two million followers could be worth up to $23.8million*

[1034] Artificial Intelligence (AI) technologies are likely to play a growing role in regulating speech over the next years. See Sashaank Pejathaya Murali, *Detecting Cyber Bullies on Twitter Using Machine Learning Techniques*, 6 INT'l J. INFO. Sec. & CYBERCRIME 63 (2017), presenting a process to detect bullying tweets using machine learning algorithms.
[1035] Nitasha Tiku: *Twitter CEO: 'We suck at dealing with abuse',* THE VERGE, Feb. 4, 2015, 9:25PM. http://www.theverge.com/2015/2/4/7982099/twitter-ceo-sent-memo-taking-personal-responsibility-for-the

in annual revenue, primarily generated by advertisers."[1036] The report also noted that platforms remove content in a piecemeal fashion, making these decisions one by one. In October 2019, a Report of the Special Rapporteur on the promotion of the right to freedom of opinion and expression to the General Assembly of the United Nations addressed the issue of moderation of hate speech by the platforms and noted that hate full speech spreads online "*seemingly spurred on by a business model that values attention and virality*."[1037] Special Rapporteur David Kaye noted in this Report that "[c]*ompanies do not have the same obligations of Governments, but their impact is of a sort that requires them to assess the same kind of questions about protecting their users 'right to freedom of expression*."[1038] It is often mentioned that Facebook's active users are more than 2 billion (2.7 billion as of the second quarter of 2020),[1039] which is more than the world's largest country, China, which "only" has a little less than 1.4 billion.[1040] The power of Facebook over its users is immense, and Facebook is a privately owned-company, quite profitable one. As such, some, including Facebook employees,[1041] are arguing that it is financially profiting from hate speech.

[1036] Center for Countering Digital Hate, *#DeplatformIcke How Big Tech powers and profits from David Icke's lies and hate, and why it must stop*, p. 3, https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9_db8ff469f6914534ac02309bb488f948.pdf.

[1037] David Kaye, P*romotion and protection of the right to freedom of opinion and expression, A/74/486*, 16. (Oct. 9, 2019), citing TIM WU, THE ATTENTION MERCHANTS; THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS (Vintage Books ed.) (2016).

[1038] Ibid.

[1039] J. Clement, *Number of monthly active Facebook users worldwide as of 2nd quarter 2020*, STATISTA, (Aug. 10, 2020), https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide.

[1040] *Population, total, China*, THE WORLD BANK, https://data.worldbank.org/indicator/SP.POP.TOTL?locations=CN.

[1041] See Craig Timberg and Elizabeth Dwoskin, *Another Facebook worker quits in disgust, saying the company 'is on the wrong side of history'*, THE WASHINGTON POST, (Sep. 8, 2020 at 6:36 p.m. EDT) and *Resignation letter from Facebook engineer*, THE WASHINGTON POST, (Sep. 8, 2020 at 10:52 AM), https://www.washingtonpost.com/context/resignation-letter-from-facebook-engineer/0538edee-7487-4822-956a-e880c2024324/?itid=lk_inline_manual_3 . Facebook software engineer Ashok Chandwaney resigned from his position, writing "***I'm quitting because I can no longer stomach contributing to an*

iv. Hate Speech, Social Media and Crime

The General Assembly of the United Nations noted in its Resolution adopted on June 26, 2018:

> "*that terrorists may craft distorted narratives that are based on the misinterpretation and misrepresentation of religion to justify violence, which are utilized to recruit supporters and foreign terrorist fighters, mobilize resources and garner support from sympathizers, in particular by exploiting information and communications technologies, including through the Internet and social media*."[1042]

Facebook, Microsoft, Twitter and YouTube announced in December 2016 that they were partnering to "*curb the spread of terrorist content online*" to create a shared industry database of "hashes," described as "*unique digital "fingerprints… for violent terrorist imagery or terrorist recruitment videos or images*" which were removed by the platforms over the years.[1043] These "hashes," which are so named because images or videos are "hashed" in their original form and not linked to a source (either a platform or user data),[1044] are numerical representation of the original content, which means it cannot be easily reverse engineered to create the image and/or video. They are shared "*to help*

---

organization that is profiting off hate in the US and globally.***"See also *Facebook 'profits from hate' claims engineer who quit*, BBC NEWS, (Dec. 9, 2020), https://www.bbc.com/news/technology-54086598.

[1042] United Nations General Assembly, *Resolution adopted by the General Assembly on 26 June2018, A/RES/72/284*, (June 26, 2018), Paragraph 22, UN SECURITY COUNCIL, http://www.securitycouncilreport.org/atf/cf/%7B65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7D/a_res_72_284.pdf.

[1043] *Partnering to Help Curb Spread of Online Terrorist Content*, FACEBOOK (Dec. 5, 2016), https://about.fb.com/news/2016/12/partnering-to-help-curb-spread-of-online-terrorist-content.

[1044] See *Transparency Report*, Global Internet Forum to Counter Terrorism, (July 2020), https://www.gifct.org/transparency/, (last visited Dec. 30, 2020).

identify potential terrorist content" on the platforms and 200,000 hashes have been gathered in a database as of December 2019.[1045] This "Hash Sharing Consortium" is supported by the Global Internet Forum to Counter Terrorism (GIFCT), launched by Facebook, Microsoft, Twitter and YouTube on August 1, 2017, and later reorganized as an independent Non-Governmental Organization, which mission is to prevent terrorists and violent extremists from exploiting digital platforms.[1046] As of July 2020, 0.1% of these hashes were imminent credible threats, 16,9% were graphic violence against defenseless people, 72% were glorification of terrorist acts and 2.1% of them are about radicalization, recruitment, and instruction.

The platforms participating to this database are determining themselves what constitute terrorist content, as "*each company will independently determine what image and video hashes to contribute to the shared database*."[1047] Is it always that obvious to know what is terrorist speech? While not alluding directly to the database, Professor Fionnuala Ní Aoláin, the United Nations Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, wrote to Marc Zuckerberg in July 2018, expressing concern over Facebook's broad definition of terrorism as "[*a*]*ny non-governmental organization that engages in premeditated acts of violence against persons or property to intimidate a civilian population, government, or international*

[1045] *Fighting Terrorism Online: EU Internet Forum committed to an EU-wide Crisis Protocol*, EUROEPAN COMMISSION, (Oct. 7, 2019), https://ec.europa.eu/commission/presscorner/detail/en/IP_19_6009.
[1046] *Transparency Report*, Global Internet Forum to Counter Terrorism, (July 2020), https://www.gifct.org/transparency/, (last visited Dec. 30, 2020).
[1047] *Partnering to Help Curb Spread of Online Terrorist Content*, FACEBOOK, (Dec. 5, 2016), https://about.fb.com/news/2016/12/partnering-to-help-curb-spread-of-online-terrorist-content.

organization in order to achieve a political, religious, or ideological aim."[1048] The Special Rapporteur was concerned by "*the overly broad definition of terrorism and terrorist organizations used by Facebook as well as the seeming lack of a human rights approach to content moderation policies*," which "*equates all non-state groups that use violence in pursuit of any goals or ends to terrorist entities*." The hashes database is an extreme example of the private laws of the platforms superseding the law of the countries, even in an area as sensitive and crucial for global security as terrorism.

Hate speech is also sometimes terrorist or speech advocating violence. In a sinister twist, Facebook Live had been used by a gunman to live-stream the New Zealand Christchurch terrorist attack of March 15, 2019, which killed fifty-one persons and injured fifty. The victims were targeted because of their Islamic faith. The video had also been live streamed on Amazon-owned Twitch.[1049] Facebook removed one million and a half videos within 24 hours of the attacks, as the video was reproduced and posted again by yet another user.[1050] Facebook's chief operating officer Sheryl Sandberg wrote then in a letter to the NZ Herald that Facebook vowed "*strengthening the rules for using Facebook Live*" and "*build[ing] better technology to quickly identify edited versions of violent videos and images*

[1048] The letter is available at https://www.ohchr.org/Documents/Issues/Terrorism/OL_OTH_46_2018.pdf. (last visited Dec. 30, 2020).
[1049] Kevin Roose, A Mass Murder of, and for, the Internet, THE NEW YORK TIMES, (March 15, 2019), https://www.nytimes.com/2019/03/15/technology/facebook-youtube-christchurch-shooting.html?module=inline.
[1050] Shibani Mahtani, *Facebook removed 1.5 million videos of the Christchurch attacks within 24 hours — and there were still many more*, THE WASHINGTON POST, (March 17, 2019, 6:05 AM EDT), https://www.washingtonpost.com/world/facebook-removed-15-million-videos-of-the-christchurch-attacks-within-24-hours--and-there-were-still-many-more/2019/03/17/fe3124b2-4898-11e9-b871-978e5c757325_story.html. 1.2 million of these videos were not even published and were blocked as users uploaded them.

and prevent people from re-sharing these versions."[1051] Indeed, the video was shared multiple times on social media, thus propagating quickly these images of violence around the globe. Sandberg wrote that prior Community Standard violations would warrant restricting who can use Facebook Live. Facebook then announced on that "*people who have broken certain rules on Facebook — including our Dangerous Organizations and Individuals policy — will be restricted from using Facebook Live*."[1052]

New Zealand Prime Minister Jacinda Ardern and French President Emmanuel Macron launched the day after the Christchurch attack the *Christchurch Call to Eliminate Terrorist & Violent Content Online,*[1053] a group of governments, major tech companies[1054] and non-profit organizations to adopt a pledge and cooperate to stop and prevent terrorists' attacks from being broadcast and spread online.[1055] Tech companies agreed "*to a set of commitments and ongoing collaboration to make the internet safer.*" Jacinda Ardern stated that "*the March 15 attack was shocking in its use of social media as a tool in the act of terror and with the Christchurch Call we have taken a unique approach to solving this problem.*"[1056] The U.S. did not join the pledge.[1057] A May 15, 2019 statement from the White House's Office of Science and Technology Policy explained that, while the United States:

[1051] *Facebook Chief Operating Officer Sheryl Sandberg's letter to New Zealand*, NZHERALD.CO.NZ, (March 30, 2019, 1:00PM), https://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=12217454.
[1052] Guy Rosen, VP Integrity, *Protecting Facebook Live From Abuse and Investing in Manipulated Media Research*, FACEBOOK, (May 14, 2019), https://about.fb.com/news/2019/05/protecting-live-from-abuse.
[1053] The complete text of the Call is available at: https://www.christchurchcall.com/call.html.
[1054] Facebook, Twitter, Google, Qwant, Microsoft, YouTube, DailyMotion and other companies signed the pledge.
[1055] Rt Hon Jacinda Ardern, *Christchurch Call to eliminate terrorist and violent extremist online content adopted,* BEEHIVE. GOVT.NZ (The official website of the New Zealand Government), (May 16, 2019), https://www.beehive.govt.nz/release/christchurch-call-eliminate-terrorist-and-violent-extremist-online-content-adopted.
[1056] Id.
[1057] *US says it will not join Christchurch Call against online terror*, BBC NEWS, (May 15, 2019), https://www.bbc.com/news/technology-48288353. Along with founders New Zealand and France,Argentina,

"*stands with the international community in condemning terrorist and violent extremist content online in the strongest terms*," and that they "*agree with the overarching message of the Christchurch Call for Action*", they are "*not currently in a position to join the endorsement.*"[1058]

The White House explained further that it chose not to join the pledge because:

"*the best tool to defeat terrorist speech is productive speech, and thus we emphasize the importance of promoting credible, alternative narratives as the primary means by which we can defeat terrorist messaging*."

Should the marketplace of ideas be left alone to fight terrorism? The Call took care to specify that "[*a*]ll *action on this issue must be consistent with principles of a free, open and secure internet, without compromising human rights and fundamental freedoms, including freedom of expression. It must also recognise the internet's ability to act as a force for good, including by promoting innovation and economic development and fostering inclusive societies."* The video of the Christchurch massacre had been shared widely, sometimes leading to criminal charges.[1059] Another terrorist attack, in Halle, Germany, in October

Austria, Australia, Belgium, Bulgaria, Canada, Chile, Colombia, Costa Rica, Cyprus, Denmark, the European Commission, Finland, Georgia, Germany, Ghana, Greece, Hungary , Iceland, Indonesia, India, Ireland, Italy, Ivory Coast, Japan, Jordan, Kenya , Latvia, Lithuania, Luxembourg, Maldives, Malta, Mexico, Mongolia, the Netherlands, Norway, Poland, Portugal, Romania, Senegal, South Korea, Spain, Slovenia, Sri Lanka, Sweden, Switzerland, UNESCO, the United Kingdom and the Council of Europe have all joined the pledge.

[1058] The White House, Office of Science and Technology, *Policy Statement on Christchurch Call for Action*, (May 15, 2019), UNITED STATES OF AMERICA, DEPARTMENT OF STATE, U.S. EMBASSY & CONSULATE IN NEW ZEALAND, COOK ISLANDS AND NIUE, https://nz.usembassy.gov/statement-on-christchurch-call-for-action.

[1059] Charlotte Graham-McLay, *New Zealand Man Gets 21 Months for Sharing Video of Christchurch Attacks*, THE NEW YORK TIMES, (June 18, 2019), https://www.nytimes.com/2019/06/18/world/asia/new-zealand-video.html. A New Zealand man was sentenced to nearly two years in jail for having distributed objectionable content by sharing a video of the Christchurch attack.

2019, was also live streamed, on Twitch.[1060] The video remained available for thirty minutes and was viewed by some two hundred persons before being flagged and taken down.[1061] It can be argued that these attacks were live streamed to incite more violence and are thus "fighting words" unprotected by the First Amendment. Witnessing live the violent death of human beings may also appeal to "prurient interest," as defined by *Miller*, as deviant personalities may enjoy sexual pleasure by witnessing it.[1062] However, banning the video did not prevent some social media users to post conspiracy theories claiming that the attack had been staged and played by actors.[1063]

Social media platforms can also be used to plot common crimes. In October 2020, the Federal Bureau of Investigation (FBI) arrested several individuals who had allegedly attempted to kidnap Michigan Governor Gretchen Whitmer. An FBI agent explained in the criminal complaint that the FBI had become aware in early 2020 "*through social media that a group of individuals were discussing the violent overthrow of certain government and law-enforcement components*." The FBI had monitored the group which allegedly plotted to kidnap the governor on social media, including on private Facebook groups.[1064] Social media platforms can also be used to commit crimes, including sexual assault. A French law

---

[1060] Tiffany Hsu, *2,200 Viewed Germany Shooting Before Twitch Removed Post*, THE NEW YORK TIMES, (Oct. 9, 2019), https://www.nytimes.com/2019/10/09/business/twitch-germany-shooting.html.
[1061] Twitch (@Twitch), Twitter (Oct 9, 2019, 4:53 PM), https://twitter.com/Twitch/status/1182036268202381313.
[1062] R. Douglas Fields Ph.D., The Explosive Mix of Sex and Violence, PSYCHOLOGY TODAY, (Jan. 26, 2016), https://www.psychologytoday.com/us/blog/the-new-brain/201601/the-explosive-mix-sex-and-violence. The author notes that "[*b*]*iologically, sex and violence share a number of common brain states and functions. Both behaviors evoke intense arousal--indeed, the most intense states of arousal possible*."
[1063] Taylor Lorenz, *Instagram Is the Internet's New Home for* Hate, THE ATLANTIC, (March 21, 2019), https://www.theatlantic.com/technology/archive/2019/03/instagram-is-the-internets-new-home-for-hate/585382/.
[1064] *Six Arrested on Federal Charge of Conspiracy to Kidnap the Governor of Michigan*, U.S. DEPARTMENT OF JUSTICE, (Oct. 8, 2020), https://www.justice.gov/opa/pr/six-arrested-federal-charge-conspiracy-kidnap-governor-michigan (last visited Dec. 30, 2020).

enacted on July 30, 2020,[1065] which primary aim is to protect victims of domestic violence, introduced in the French criminal Code a new crime, "*making offers or promises to a person or offering him any gifts, presents or advantages so that he commits rape, including outside the national territory*." The crime is punishable, even if the rape has not been committed, by ten years in jail and a 150,000 Euros fine.[1066] If the offer or promise has been made to incite a third party to commit a sexual assault, and the sexual assault has not been committed, then the crime is publishable by seven years in jail and a 100,000 Euros fine.[1067] The law was enacted as a response to several instances where French nationals used foreign darknet platforms to order sexual abuses to be performed abroad and watched them while there were live streamed. As this sordid crime of rape or sexual assault is actually committed in a foreign country, the July 30, 2020 law modified the French criminal Code to extend jurisdiction of France of these crimes.[1068] A French professor was killed, in October 22020, because he had shown caricatures of the Prophet Mohammad to his students. Third parties allegedly shared information online allowing the killer to identify his victim. French Prime Minister Jean Castex then announced in November 2020 that the government would propose a bill which would make it a crime to "*revea*[*l*], *disseminat*[*e*] *or transmi*[*t*] *by any means whatsoever, information relating to the private, family or professional life of a person, making it possible to identify or locate him, with the aim of placing* [*the individual or members of his or her family], at an immediate risk of harm to life or physical or mental*

[1065] Loi 2020-936 du 30 juillet 2020 visant à protéger les victimes de violences conjugales [Law 2020-936 of July 30, 2020 to protect victims of domestic violence], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], July 31, 2020, available at https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000042176652. Its article 24 creates article Art. 222-26-1 and article 222-30-2 of the French criminal Code.
[1066] Article 222-26-1 of the French criminal Code
[1067] Article 222-30-2 of the French criminal Code.
[1068] Article 113-5 of the French criminal Code, as modified by article 24 of the July 30, 2020 law.

integrity, or property." The crime would be punishable by three years' imprisonment and 45,000 euro fine.[1069]

However, social media platforms may be used to fight crime. For instance, in 2014, Twitter users helped the police to identify a group of people who had allegedly beaten up a couple of gay men.[1070] The Philadelphia police posted on its blog images from a surveillance camera of a group of people walking together in the area where the assault had taken place, taken minutes before the crime took place, and the link to the blog post was posted on Twitter.[1071] Twitter users were then able to identify them, as the group later went together to a restaurant.[1072] Twitter users were able to identify the restaurant, using Facebook Graph Search to identify who had checked in at this restaurant on Facebook the night of the crime. A detective for the Philadelphia police later tweeted: "*This is how Twitter is supposed to work for cops. I will take a couple thousand Twitter detectives over any one real detective any day*."[1073] Social media posts may also be used the police to find out that mandatory COVID-19 quarantine has been violated.[1074]

[1069] Marc Rees, *Le futur délit de « mise en danger de la vie d'autrui » par diffusion de données personnelles*, NEXTINPACT (Nov. 18, 2020, 06 :39), https://www.nextinpact.com/article/44336/le-futur-delit-mise-en-danger-vie-dautrui-par-diffusion-donnees-personnelles.
[1070] Rheana Murray, *Twitter Sleuths Lead Cops to Suspects in Pennsylvania Hate Crime*, ABC NEWS, (Sept. 17, 2014, 2:42 PM ET), http://abcnews.go.com/US/twitter-sleuths-lead-cops-suspects-pennsylvania-hate-crime/story?id=25562144.
[1071] @FanSince09, Twitter (Sept. 16, 2014, 4:14 PM), https://twitter.com/FanSince09/status/511971313385103360?ref_src=twsrc%5Etfw.
[1072] @GreggyBennet, Twitter (Sept. 16, 2014, 8:41 PM), https://twitter.com/GreggyBennett/status/512038676918845440/photo/1?ref_src=twsrc%5Etfw.
[1073] @PPDJoeMurray, Twitter, (Sept. 2014, 10:02 PM), https://twitter.com/PPDJoeMurray/status/512059079507083264?ref_src=twsrc%5Etfw.
[1074] *Man Arrested for Quarantine Violation Based on Social Media Posts*, BIG ISLAND NOW, (May 17, 2020, 10:58 AM HST). A tourist was arrested by the Hawaii police after he allegedly posted on social media that he was at the airport and would be leaving the island in a few hours, thus violating the mandatory 14-day self-quarantine rule for all incoming visitors and returning residents imposed at the time.

If social media platforms have their own laws, should have they have their own courts as well?

III. The Private Courts of the Platforms

An April 2018 Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression to the General Assembly of the United Nations stated that online companies engaged in moderating user-generated content "*must embark on radically different approaches to transparency at all stages of their operations, from rule-making to implementation and development of "case law" framing the interpretation of private rules*."[1075] Should social media platforms have their own, private, judicial system, maybe even creating their own case-law?

Platforms decide what is illegal speech warranting takedown, with no judicial oversight. Such decisions must be made every day and is carried out by algorithms, by employees, and by outside contractors. This is a difficult and high-stress job, as these employees are subjected every day to hate speech, extreme violence, such as beheadings, and to graphic or violent speech, to such an extent that some suffer from post-traumatic stress disorder (PTSD). In a preliminary settlement filed in the Superior Court of California for the County of San Mateo, Facebook agreed in May 2020 to pay fifty-two million dollars to current and former moderators. All will receive at least one thousand dollars, but more if they are diagnosed with PTSD. Facebook agreed in the settlement to change its content moderation tools, so that the impact of viewing harmful images and videos would be

[1075] *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression A/HRC/38/35*, (April 8, 2018), paragraph 71, available at https://digitallibrary.un.org/record/1631686?ln=en.

reduced, such as offering the option to mute the audio by default or to view the videos in black and white when evaluating content against Community Standards relating to graphic violence, murder, sexual abuse and exploitation, child sexual exploitation, and physical abuse.[1076] Moderators viewing graphic and objectionable content on a regular basis have now access to weekly, one-on-one coaching sessions with a licensed mental health professional.[1077]

Yet, moderation must be provided, as quickly as possible, day after day. As penalties for failing to remove content are high, the risk of over-deleting content is also high and may lead to private censorship, as noted by the non-profit organization Human Rights Watch: "*[f]aced with short review periods and the risk of steep fines, companies have little incentive to err on the side of free expression*.)"[1078] Users may provide valuable help to platforms signaling content as being illegal. However, the procedure is not strictly a notice and take down procedure, as is, for instance, the procedure provided by Section 512 of the Digital Millennium Copyright Act.[1079] Social media platforms provide users tools to report messages they deemed to be illegal. However, not every notice is acted upon by taking down the content, particularly when a content is flagged by users as violating the rules of a platform but found by moderators as not to be in violation. Social media platforms rely more and more on "trusted flaggers," who are expert in their fields. The European Commission encouraged social media platforms to use them as they "*can be expected to bring their expertise and work with high quality standards*" and their notices "*should be able*

[1076] Facebook's Civil Rights Audit, (July 8, 2020), p.44
[1077] Ibid.
[1078] *Germany: Flawed Social Media Law- NetzDG is Wrong Response to Online Abuse*, HUMAN RIGHTS WATCH, (Feb.14, 2018 12:01AM EST), https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law.
[1079] 17 U.S.C. § 512.

to be fast-tacked by the platforms."[1080] The Commission further noted that "[*a*]*s a general rule, removal* [*of content*] *deriving from trusted flaggers notices should be addressed more quickly, due to the quality and accuracy of the information provided in the notice and the trusted status of the flaggers*."[1081]

Facebook CEO Mark Zuckerberg explained in a 2018 interview[1082] that his goal was "*to create a governance structure around the content and the community that reflects more what people in the community want than what short-term-oriented shareholders might want*." He stated that such scheme should have "*some sort of independent appeal process*" after a particular speech has been taken down by Facebook, adding: "*I think in any kind of good-functioning democratic system, there needs to be a way to appeal*." He explained further that "*over the long term*," he would like to create an independent appeal. Facebook would first decide to take down, or to not take down, a particular content based on the site's community standards, a sort of private law, as we saw. Users could then get a second opinion, explaining:

> "*You can imagine some sort of structure, almost like a Supreme Court, that is made up of independent folks who don't work for Facebook, who ultimately make the final judgment call on what should be acceptable speech in a community that reflects the social norms and values of people all around the world*."

[1080] *Tackling Illegal Content Online –Towards an enhanced responsibility of online platforms*, Communication COM (2017) 555 of September 2017 from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, p. 8-9, giving as an example of such trusted flaggers Europol's Internet Referral Unit which "*has the necessary experience to assess whether a given content constitutes terrorist and violent extremist online content*."
[1081] Communication COM (2017) 555, p. 14.
[1082] Ezra Klein, *Mark Zuckerberg on Facebook's hardest year, and what comes next*, VOX, (Apr 2, 2018, 6:00am EDT), https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge.

The June 2019 progress report on Facebook's Civil Rights Audit Report announced that Facebook would create an Oversight Board (the Oversight Board) to allow its users to appeal content decisions to Board.[1083] The progress report explained that Facebook's goal was "*to establish a body with independent judgment whose decisions are transparent and binding to review some of the most challenging content decisions that Facebook makes*." Facebook announced in July 2020 the creation of an independent Oversight Board, composed of forty members, which would review content removal decisions and decide if the content should stay published or not. The decision of the Oversight Board will be binding.[1084] Facebook selected four co-chairs of the Board, who selected, together with Facebook, the other sixteen Board members. These initial twenty members then selected, in partnership with Facebook, twenty additional members.

This move may very well lead to the establishment of a private court of law, with its own rules and caselaw. The Oversight Board 's purpose is, according to its Charter, "*to protect free expression by making principled, independent decisions about important pieces of content and by issuing policy advisory opinions on Facebook's content policies*."[1085] The Board oversees content both on Facebook and on Instagram, which is owned by Facebook. Users can ask the board to review a particular content, if they disagree with a decision made by Facebook, which can also refer a particular case to the Oversight Board (article 2 of the Charter). The Oversight Board's resolution of each case is binding and is "*promptly*"

[1083] See *Facebook's Civil Rights Audit – Progress Report* (June 30, 2019),p.14, available at https://about.fb.com/wp-content/uploads/2019/06/civilrightaudit_final.pdf, (last visited Dec. 30, 2020). For a detailed explanation on how the Oversight Board was created, see Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L. J. 2418 (2020).
[1084] See *Facebook's Civil Rights Audit – Final Report* (July 8, 2020), p.46, available at https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf (last visited Dec. 30, 2020).
[1085] Oversight Board Charter, available at OVERSIGHT BOARD, https://www.oversightboard.com/governance/ (last visited Dec. 30, 2020).

implemented by Facebook," *unless implementation of a resolution could violate the law*" (article 4 of the Charter).

The Oversight Board serves thus as a sort of private court of appeals and review decision taken by Facebook about content published on Facebook and Instagram.[1086] It will review information to make its decision, including information provided by the user: "[*t*]*he panel will deliberate the case and make a decision based on all the information provided by the person who submitted the appeal, by Facebook and by any experts called upon to provide further context*."[1087] The Oversight Board will then "*produce a written explanation of its decision, which will be available for the public to read on this website*."[1088] As such, there is a real possibility that the board will develop its own caselaw, which may possibly influence the thinking of judges and legislative bodies around the world, as an example of a private, yet universal, set of social media and freedom of expression cases.

There will be some rules to follow to lodge an appeal, although they are, at least at the time this article is written, rather simple to be called "procedure." They are four main rules.

First, one must have standing, that is, one must be the holder of an active account:

[1086] *Appealing Content Decisions on Facebook or Instagram,* THE OVERSIGHT BOARD, https://www.oversightboard.com/appeals-process, (last visited Dec. 30, 2020): "*The Oversight Board appeals process gives people a way to challenge content decisions on Facebook or Instagram. If you have already requested that Facebook or Instagram review one of its content decisions and you disagree with the final decision, you can appeal to the board*."

[1087] Ibid.: "*Once a case is selected, a panel of members will be assigned to it and will receive information to help with their in-depth review. This includes information shared by the person who submitted the appeal as well as contextual information from the Facebook company, in compliance with applicable legal and privacy restrictions*."

[1088] Ibid.

"[*t*]*he person appealing a content decision must have an active account on the service where the content was posted. This means the account cannot be deleted or disabled and the person must be able to log into it*."

This means that, if a person has been banned from the platform for having posted content violating Facebook Rules, she or he cannot appeal to the Oversight Board to review whether the content indeed violated the Rules, thus indirectly overturn the banning decision.

The second rule states that the party must have exhausted all remedies, that is:

"*Facebook must have already reviewed its initial decision. The person submitting the appeal must have already requested that Facebook or Instagram review its content decision and received a final decision."*

This is a mere procedure rule, showing the way a complaint must proceed on the path leading to the review by the Oversight Board.

The third rule is more difficult to apprehend, as it states that:

"[*c*]*ontent decisions must be eligible for appeal*," explaining further that "[*t*]*he board is committed to keeping users safe and abides by country-specific laws. Because of this, not all content decisions are eligible for appeal. When a decision is eligible, the update from Facebook or Instagram will include a reference ID that can be used to submit an appeal*."

The Oversight Board thus appears to be able to grant certiorari, but its decision to take or not the case will be taken by "*abid*[*ing*] *by country-specific laws*. "It is thus likely that

the Oversight Board will not take a case of a user complaining about a post being taken down when the post blatantly violates the law of the user's country. Such would be the case if a post made by a Thai user criticizes the Thai king. This rule may lead to the Oversight Board to regard its jurisdiction as limited by country-specific laws, which may, in turn, induce legislators to enact laws precisely forbidding a certain type of speech, to weaken the scope of the Oversight Board, possibly leading to an inflation of laws limiting speech.

The fourth rule is another procedure rule, a sort of statute of limitations, as it states that "*appeals must be submitted within 15 days*." Interestingly, the public is provided an opportunity, within this time frame, to comment on the cases the Oversight Board has accepted to review. The public may forward comments about the cases within a week from the publication of the cases, in English or in the language of the original post.

The entire Oversight Board does not review all the cases. Instead, "*a subset of the board deliberates and issues a draft decision*."[1089] However, "*the entire board will have the opportunity to review the draft decision before it becomes final. The board will publish a written statement about its decision, which may include a policy recommendation for Facebook*."[1090] The decisions will be made public and are likely to create a corpus of caselaw, which, after a few years or less, will be important enough to show general trends. The policy recommendations are likely to announce upcoming changes in Facebook policy.

Will we see instances where individuals and common interest groups will try to create a controversy by posting a message designed to be taken down, then follow the

[1089] Ibid.
[1090] Ibid.

procedural rules of the Oversight Board to have the decision reviewed or not, to be able to argue bias?[1091] Some already see in the Oversight Board a partisan body: FCC Commissioner Brendan Carr said in a May 28, 2020 interview, shortly after the publication of the Executive Order on social media, that the Board had been *"stacked … with people who are emotionally, viscerally against the President."*[1092]

The Oversight Board accepted its first cases on October 22, 2020.[1093] Will its established influence the removal practice of Facebook, who may, perhaps, take down more content now, knowing that a heavy-handed decision may be reversed by the Oversight Board? This would also shift the responsibility from Facebook to the Oversight Board.[1094] The Oversight Board selected its first six cases in December 2020.[1095] It will be interesting to watch how the first private freedom of expression caselaw will influence, or not, the states' jurisprudence and the legislators' policies, around the world. Will it lead to more division or to more inclusion?

[1091] The U.S. Republicans may use it to further their claim that the social media platforms are biased against them, *see* for instance James Clayton, *Social media: Is it really biased against US Republicans?* BBC (Oct. 27, 2020), https://www.bbc.com/news/technology-54698186.

[1092] A segment of the video has been published on Twitter by the White House Twitter account, see @WhiteHouse, Twitter (May 28, 2020, 7:21 PM), https://twitter.com/WhiteHouse/status/1266147550140272641.

[1093] Shirin Ghaffary, *Facebook's independent oversight board is finally up and running*, VOX, (Oct. 22, 2020, 2:14pm EDT), https://www.vox.com/recode/2020/10/22/21528859/facebook-oversight-board-mark-zuckerberg.

[1094] See Kate Klonick, The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression, 129 YALE L. J. 2418, 248 (2020).

[1095] *Announcing the Oversight Board's first cases and appointment of trustees*, THE OVERSIGHT BOARD, https://www.oversightboard.com/news/719406882003532-announcing-the-oversight-board-s-first-cases-and-appointment-of-trustees.

Even if a particular speech may be posted on social media without breaking a law or the private rules of the platform, it may nevertheless breach yet another private rule, the one crafted and enforced by an employer, or a professional body.

IV. The Private Censorship of Speech

A. Employees and Social Media

Employees are regularly fired in the U.S. over a message posted on social media, even if the message is about a private matter and is posted from a private account. We need to distinguish, however, whether the employee works for a private employer or for a public one. If the employer is a public entity, the employee may be considered to have exercised her or his free speech right, as protected by the First Amendment. In *Jamie Marquardt v. Nicole Carlton*, the Court of appeals for the Sixth Circuit addressed the issue of whether an employee of the city of Cleveland had been fired in retaliation for private Facebook posts.[1096] In this case, a Cleveland Emergency Medical Services "*allegedly made incendiary comments on in private Facebook page*" about the death of Tamir Rice, who had been shot by the police in a Cleveland park. The posts were only visible to his Facebook "friends," and had been removed within hours. However, they quickly became the topic of the employee's colleagues. The employee was dismissed by his employer, for violation of the City of Cleveland's social media policies. He sued the City of Cleveland, alleging a violation of his First Amendment rights. The Sixth Circuit examines three questions to assess whether a public employer impermissibly retaliated against an employee for her speech: (1) did the employee engage in protected speech; (2) would the dismissal discourage an individual of

[1096] Jamie Marquardt v. Nicole Carlton, 971 F.3d 546 (2020).

"ordinary firmness" from engaging in the activity which led to disciplining the employee and (3) was the employee's protected speech "a motivating factor" for the adverse action. The court uses a two-part test to find out whether the employee's speech was protected: (1) was the speech on a "*matter of public concern*" and (2), if it was, did the "*employee's free speech interests outweigh the efficiency interests of the government as an employer*."[1097] The Sixth Circuit examined only part one of the two-part test, whether the speech was a matter of public concern, and found that it was, as the death of Tamir Rice had "*set off a fierce public debate over whether the officer's actions were justified*."[1098] For the Court, "*expressions of opinion, even distasteful ones, do not become matters of personal interest simply because they are phrased in the first person or reflect a personal desire*."[1099] Also, even though the posts were only visible to the employee's Facebook's "friends," the Court explained that "*speech need not be communicated to the general public to be on a matter of public concern*."[1100]

The legal issues are different, however, if the employer is a private entity. For instance, a Texas law firm fired one of its employees in May 2020 following a social media post allegedly threatening a food business requiring wearing mask during the Covid-19 pandemic .[1101] It allegedly read: *"Do I have to show the lame security guard outside of a ghetto store my CV19 test results? I will show him my Glock 21 shooting results… I have more power then they do… they just don't know it yet*." The law firm announced the firing on

[1097] Marquardt at 549, citing Roerr v. City of Stow, 743 F. 3d1025, 1047 (6th Cir. 2014).
[1098] Ibid.
[1099] Marquardt at 550.
[1100] Marquardt at 551.
[1101] Debra Cassens Weiss, *Firm fires staffer for 'no more masks' social media post that referred to Glock pistol*, ABA JOURNAL, (May 11, 2020, 2:16 pm CDT), https://www.abajournal.com/news/article/firm-fires-staffer-for-no-more-masks-post-that-referred-to-glock-pistol.

Facebook,[1102] which led to several complimentary comments about the move from Facebook users not appearing to be clients. As such announcing the dismissal could be viewed generating goodwill. While the Facebook's post announcing the firing stated that the employee social media activities had violated the "values" of the firm, these values must be written in the employee handbook and the social media policy to protect a company against lawsuit for wrongful termination. Can private companies fire their employee at will over a social media post?

While employment is at-will in the U.S., and thus employees can be dismissed, even without cause, such dismissal can nevertheless not be decided because an employee had engaged in concerted activities, as defined by the National Labor Relations Act (NLRA). The National Labor Relations Board (NLRB) enforces the NLRA, which paragraph 7 provides that private employees, whether unionized or not, have the right to "*to engage in other concerted activities for the purpose of collective bargaining or other mutual aid or protection*."[1103] The Supreme Court held in 1975, in *NLRB v. J. Weingarten, Inc*. that:

> "[*t*]*he action of an employee in seeking to have the assistance of his union representative at a confrontation with his employer clearly falls within the literal wording of § 7 that "[e]mployees shall have the right . . . to engage in . . . concerted*

[1102] The law firm posted on Facebook: "This afternoon we learned that an administrative employee of the Firm issued a threatening and offensive *post on a personal social media account related to COVID-19 mask protections. This post is a complete violation of the values of our Firm, including our commitment to the health and safety of the communities we serve. We have terminated this individual's employment and notified the proper authorities about the post as a precaution. We are deeply sorry for this situation. This type of post is not and never will be tolerated by our Firm."* Thompson & Knight LLP , Facebook (May 8, 2020), https://www.facebook.com/ThompsonKnightLLP/photos/a.10150249018926027/10157146850846027/?type=3&theater (last visited Dec. 30, 2020).
[1103] 29 U.S. Code § 157.

activities for the purpose of . . . mutual aid or protection, " a right referred since by the NLRB as "Weingarten rights."[1104]

Employers violates Section 8(a)(1) of the NLRA provides that it is unlawful for a labor organization or its agents "*to restrain or coerce employees in the exercise of the rights guaranteed them in Section 7 of the Act*" and violation of the Weingarten rights also violates Section 8(a)(1) of the NLRA.[1105]

In 2010, the NLRB found in *Am. Med. Response of Conn., Inc., N.L.R.B. Gen. Counsel Advice Mem*.[1106] that a Facebook post criticizing a supervisor in choice terms was concerted activity. In this case, a union-member paramedic worked for American Medical Response of Connecticut (AMR). After the employee had been confronted by her supervisor after an incident and had refused to file a report without a union representative being present, she commented about the incident with colleagues on her Facebook page.[1107] As noted by the NLRB, "[*m*]*any of* [*union-member paramedic*]*'s coworkers and supervisors have access to her Facebook page and regularly use the site to communicate, including to criticize management*." The employer employee was then discharged for violation of her employer's blogging policy and for having refused to write the report. The NLRB held that the employee was engaged in protected activity and exercising her Weingarten rights "*by discussing supervisory actions with coworkers in her Facebook post*." The NLRB applied the four factors it uses to determine if an employee engaged in a protected, concerted activity

[1104] NLRB v. J. Weingarten, Inc., 420 US 251, 260 (1975).
[1105] T.N.T. RedStarExpress, Inc., 299 NLRB 894, 894 (1990).
[1106] Am. Med. Response of Conn., Inc., N.L.R.B. Gen. Counsel Advice Mem., 2010 WL 7368090 (Oct. 5, 2010).
[1107] She wrote, for instance: "*Looks like I' m getting some time off. Love how the company allows a 17 [AMR code for a psychiatric patient] to be a supervisor*."

lost the protection of the NLRA "*by opprobrious conduct*,"[1108] (1) the place of the discussion; (2) the subject matter of the discussion; (3) the nature of the employee's outburst; and (4) whether the outburst was, in any way, provoked by an employer's unfair labor practice. The Board concluded that the employee's conduct was not so opprobrious as to lose protection of the NLRA, even if she had referred to her supervisor as a "dick" and a "scumbag" on Facebook. The social media posts had not interrupted employees work, as "*they occurred outside the workplace and during the non-working time of both [the employee] and her coworkers*" (first factor). The comments had been made "*during an online employee discussion of supervisory action, which is… protected activity*" (second factor). The way the employee had referred to her supervisor "*was not accompanied by any verbal or physical threats*," and the NLRB noted further that it has found that "*more egregious name-calling protected*"[1109] (third factor). Finally, the NLRB noted that the fourth factor "*strongly favor*[*ed*] *a finding that the conduct was protected*," as the Facebook postings had been "*provoked by* [*the supervisor*]*'s unlawful refusal to provide* [*the employee*] *with a Union representative for the completion of the incident report and by his unlawful threat to discipline her*."

The case is particularly interesting because the NLRB also addressed the issue of the employer's blogging and internet policy which provided that:

> "*Employees are prohibited from posting pictures of themselves in any media, including but not limited to the Internet, which depicts the Company in any way, including but*

[1108] Atlantic Steel Co., 245 NLRB 814, 814 (1979).
[1109] Citing Stanford Hotel, 344 NLRB 558, 558-559 (2005), a case where calling a supervisor a "liar and a bitch" and a "fucking son of a bitch" was found not to be opprobrious.

not limited to a Company uniform, corporate logo or an ambulance, unless the employee receives written approval from the EMSC Vice President of Corporate Communications in advance of the posting;

Employees are prohibited from making disparaging, discriminatory or defamatory comments when discussing the Company or the employee' s superiors, co-workers and/or competitors."

The NLRB found that prohibiting employees to post pictures of themselves depicting the company in any way restricted employees' Section 7 rights in violation of Section 8 (a)(1), as it would prohibit an employee to engage in protected activity, giving as example a picture of an employee wearing a tee-shirt with the company's logo taken during a protest about the terms and conditions of employment.[1110] Prohibiting making disparaging comments when discussing the company or its employees was also found to be unlawful, especially as the policy was overbroad and "*contain*[*ed*] *no limiting language to inform employees that does not apply to Section 7 activity*."

This case was the first of a series of NLRB case law about social media use by employees. It should be noted that, if an employee is fired because of a social media posting, he or she will not be protected by the NLRA if the employee's activity is "*wholly distinct from activity that feels within the ambit of Section 7*."[1111] The NLRB General Counsel identified in a April 2011Memorandum "[*c*]*ases involving employer rules prohibiting, or*

[1110] Citing Boise Cascade Corp., 300 NLRB 89, 86 (1990), where the NLRB held that wearing a tee-shirt bearing the employer's logo in connection with a protest of terms and conditions of employment is protected.
[1111] Walmart, No. 11-CA-067171 (May 230, 2012), cited by Jessica A. Magaldi, Jonathan S. Sales & Elizabeth A. Cameron, *How the NLRB's Decisions in Cases Involving Social Media Are Narrowing the Definition of ConcertedActivity - Whether Employees like It or Not*, 49 U. TOL. L. REV. 233,250 (2018).

discipline of employees for engaging in, protected concerted activity using social media" as "*requiring a decision by the General Counsel because of the absence of precedent or because they involve identifies policy priorities*. "[1112] The Acting General Counsel issued three more reports about the issue in 2011 and 2012.[1113] As social media use became more and more mainstream, including among U.S. workers, the NLRB had the opportunity to develop its "social media and free speech" case law. For instance, it held in December 2012 that discharging employees because they had posted comments on Facebook in response to a coworker's criticisms of their job performance is a violation of Section 8(a)(1) of the NLRA.[1114] The NLBR uses four factors[1115] to determine if the discipline or discharge of an employee violates Section 8(a)(1) of the NLRA, two of which were found to be relevant in this case: (1) whether the activity in which the employee was engaged was "concerted" within the meaning of Section 7 of the NLRA ; (2) whether the employee know of the concerted nature of the employee's activity; (3) whether the activity was protected by the NLRA and (4) whether the discipline or discharge was motivated by the employee concerted activity. Only the first and the third factors were in dispute in this case. The

[1112] NLRB Gen. Counsel Mem. 11-11, Mandatory Submissions to Advice, 2011 WL 3348287 (Apr. 12, 2011). In this case, an employee had been fired after having posted on his private Facebook account*: "The government needs to step in and set a limit on how many kids people are allowed to have based on their income. If you can't afford to feed them you shouldn't be allowed to have them. Our population needs to be controlled! In my neck of the woods when the whitetail deer get to be too numerous we thin them out! ... Just go to your nearest big box store and start picking them off. ... We cater too much to the handicapped nowadays! Hell, if you can't walk, why don't you stay the f**k home!!!!*"It can be argued that this post would be considered hate speech in the EU. For more comment on the case, *see* Anna Hickman, *Lawful Social Media Policy: NLRB Finally Provides an Example*, THE FEDERAL LAWYER (Dec. 2012), available at https://www.fedbar.org/wp-content/uploads/2012/12/le-dec12-pdf-1.pdf.

[1113] NLRB Gen. Counsel Opers. Mem. 11-74, Report of the Acting Gen. Counsel Concerning Social Media Cases, 2011 WL 11718018 (Aug. 18, 2011); NLRB Gen. Counsel Opers. Mem. 12-31, Report of the Acting Gen. Counsel Concerning Social Media Cases, 2012 WL 10739276 (Jan. 24, 2012); NLRB Gen. Counsel Opers. Mem. 12-59, Report of the Acting Gen. Counsel Concerning Social Media Cases, 2012 WL 10739277 (May 30, 2012).

[1114] Hispanics United of Buffalo, Inc., Case 03-CA-027872 (Dec. 14, 2012).

[1115] *See* Meyers Industries, 268 NLRB 493 (1983) (Meyers I).

NLRB found that the employees had engaged in concerted activity when posting comments on Facebook, as they had done so to discuss the fact that a colleague was planning to tell their supervisor that their job performance was no satisfactory. Interestingly, these Facebook posts wad been made on a non-working day, using the employees' own computers, from their home.[1116] What was relevant for the purpose of the first factor was that the employees, even on their day off, even from home, had been engaged in a concerted activity for the "*purpose of mutual aid or protection"* under Section 7 of the NLRB, discussing together and for their common cause the issue of the upcoming discussion about their performance. The NLRB also found that the employee's concerted activity was protected by the NLRA, as required by the third factor, as the Facebook comments "*plainly centered*" on the topic of job performance.

In *Novelis Corp. v. NLRB*,[1117] an employee who had actively conducted organizing activities, but had failed to have the company unionized, was fired for having posted on his Facebook account a message using vulgar words, complaining about his salary and criticizing colleagues who had voted against unionizing.[1118] The message had been "liked" by eleven employees. The administrative law judge found that this Facebook post was concerted activities protected under section 7 of the NLRA, reasoning that eleven employees had "liked" or commented on the post. The Second Circuit agreed, as "*employee's speech is "concerted*" if "*it is engaged in with the object of initiating or inducing*

[1116] Member of the NLRB Hayes dissented, arguing that "*the mere fact that the subject of discussion involved an aspect of employment… is not enough to find concerted activity for mutual aid and protection*."
[1117] Novelis Corp. v. NLRB, 885 F. 3d 100 (2nd Circ. 2018).
[1118] The post read: "*As I look at my pay stub for the 36 hour [sic] check we get twice a month, One worse than the other. I would just like to thank all the F*#KTARDS out there that voted `NO' and that they wanted to give them another chance...! The chance they gave them was to screw us more and not get back the things we lost....! Eat $hit `NO' voters.....*"

group action, " and concluded that the "[*NLRB*]*'s finding was grounded in substantial evidence, including* [*the dismissed employee*]'s *testimony and screenshots of his Facebook post.*"[1119] In *Three D LLC v. National Labor Relations Board*, an employee had "liked" another employee's status update ("*Maybe someone should do the owners of Triple Play a favor and buy it from them. They can't even do the tax paperwork correctly!!! Now I OWE money...Wtf!!!*!") and another employee had commented: "*I owe too. Such an asshole*."[1120] The NLRB had considered it to be "*ongoing dialogue among employees about tax withholding.*" The NLRB had agreed with the administrative law judge that the Facebook posts were concerted activity under *Meyer*, as it "*involved four current employees and was "part of an ongoing sequence of discussions that began in the workplace about [Triple Play's] calculation of employees' tax withholding*." The employer argued on appeals that the employee should lose the protection of the NLRA because the post had used obscene language, which had been viewed by customers. The Second Circuit was not convinced, as considering that that online posting by employees is made *""in the presence of customers" could lead to the undesirable result of chilling virtually all employee speech online*."[1121]

Employers often address the issue of social media and blog posts in their Employee Handbook. Such was the case for a car dealership in Illinois, which addressed "Bad Attitude"[1122] and "Courtesy."[1123] The car dealership, which had several facilities, had fired

[1119] Novelis Corp. at 108, citing NLRB v. Caval Tool Div., 262 F.3d 184, 190 (2d Cir. 2001).
[1120] Three D, LLC v. National Labor Relations Board (Oct. 21, 2015) (2nd Circ. 2015).
[1121] Under NLRB v. Starbucks Corp., 679 F. 3d 70,79 (2nd Circ. 2012), finding that the employer has a "*legitimate concern*" not to tolerate *"employee outbursts containing obscenities in the presence of customers*."
[1122] It read: "*Employees should display a positive attitude toward their job. A bad attitude creates a difficult working environment and prevents the Dealership from providing quality service to our customers*."
[1123] It read: *"Courtesy is the responsibility of every employee. Everyone is expected to be courteous, polite and friendly to our customers, vendors and suppliers, as well as to their fellow employees. No one should be disrespectful or use profanity or any language which injures the image or reputation of the Dealership*."

one car salesman who had criticized on his Facebook wall the choice of food (hot dogs and chips) to be offered to customers during a major sales event organized to introduce a redesigned luxury automobile. The employee had deemed these offerings to be inadequate for potential buyers of luxury cars and had shared his concerns with colleagues and management prior to the event. The employee had also posted pictures of an accident which had occurred at one of his employer's dealerships adjacent to the one where he was working, where a vehicle had ended into a pond. The Facebook comment about the accident read: "*This is your car; This is your car on drug*." The salesperson had been fired because of these two postings. The administrative law judge found that the posting about the food served at the sales event was protected concerted activity, as the inadequacy of the food offered at the sales event "*could have had an effect upon* [*the employee's*] *compensation*," since it was based on commission on sales, sales volume, and a customer satisfaction index, based upon surveys sent to customers who bought a car.[1124] The administrative judge, found, however, that the post about the accident was not protected nor concerted activity, as it "*had no connection to any of the employees' terms and conditions of employment*." Addressing the issue of the Employee Handbook, the administrative law judge found that the dealership had the right to ask its employees to be courteous towards the customers, but that the rule on "Courtesy" violated Section 8(a) of the NLRA as "*employees could reasonably interpret it as curtailing their Section 7 rights*." The NLRB agreed with the judge, because the "Courtesy" rule could reasonably be construed by employees "*as encompassing Section 7 activity, such as employees' protected statements…*

[1124] N.L.R.B. Div. of Judges (Sept. 28, 2011), WL 4499437.

that object to their working conditions and seek the support of others in improving them."[1125] This case is considered to have been superseded by *The Boeing Company* NLRB case, decided in December 2017, which established the balancing "Boeing test" to be used when assessing whether an employer rule is lawful.[1126] A rule established by the employer does no longer violate Section (8)(a)(1) of the NLRA if the employees "*would reasonably construe*" it to restrict protected activity. Instead, the NLRB now considers (1) the rule's potential impact on protected concerted activity, as protected by Section 7 of the NLRA and (2) the employee's legitimate business considerations for the rule. These business considerations include, for instance, maintaining discipline and ensuring productivity.

Following the *Boeing* case, the NLRB General Counsel issued a Memorandum in June 2018 commenting on the new *Boeing* balancing test.[1127] The Memorandum examined the legality of three categories of employees' rules: (1) "*rules that are generally lawful to maintain*"; (2) "*rules warranting individualized scrutiny*" and (3) "*rules that are unlawful to maintain*." Among "*rules that are generally lawful to maintain"* are "*civility rules*," such as rules prohibiting rude, disparaging, or condescending language. The NLRB General Counsel noted that "*the vast majority of conduct covered by such as rule, including name-calling, gossip, and rudeness, does not implicate Section 7 at all*," adding that *"while protected activity may involve criticism of fellow employees or supervisors, the requirement that such criticism remain civil does not unduly burden the core right to criticize*." However, it can be argued that an employee feeling slight or shortchanged may very well use strong language, in the

[1125] 358 NLRB No. 164 (N.L.R.B.)
[1126] 365 NLRB No. 154, slip op.
[1127] NLRB General Counsel Memorandum 18-04, 2018 WL 2761555 (June 6, 2018).

spur of the moment, especially if doing so on his or her own social media account, considered to be a private space.

Among the rules "*generally lawful to maintain*" are also no-photography rules. The NLRB General Counsel notes that, in *Boeing*, the NLRB had found these rules to "*have little impact on NLRA-protected rights, since photography is not central to protected concerted activity*."[1128] However, photography, and its progeny, video, are technologies now readily available in smart phones, and are prominently used during protests and other events, including events protected by the NLRA, and pictures of these events may be posted simultaneously or after the event on employee's social media accounts. It thus can be argued that photography is indeed "*central to protected activity*." The "*rules protecting confidential, proprietary, and customer information or documents"* can however be broken by employees on social media, if sharing information either obviously confidential, such as photographing secret projects or documents, or less obviously confidential, as, for instance, sharing regular check-ins at a competitor's location, which may provide clues of negotiations for a takeover. This proprietary information may or may not be considered trade secrets. In a recent case, a journalist had resigned from his sportswriter position at the *Virginian-Pilot* newspaper and had kept the Twitter account he had used while working there. He had not created the account, but, rather, was provided its login information when starting to work for the newspaper. The account had originally been created by the former sportswriter and had been used to post about Virginia sports events of interest to the readers. When the journalist left his position, he kept the Twitter account, which had then

[1128] Boing Co., 365 NLRB No.154, slip op, at 19.

some twenty-seven thousand followers. His former employee sued him, *inter alia*, for breach of the journal's social media policy and misappropriation of trade secrets under the Defend Trade Secrets Act, 18 U.S.C. § 1836, et seq., and the Virginia Uniform Trade Secrets Act, Va. Code § 59.1-336, arguing that the list of followers, which:

> "*provide*[*d*] *direct, unfettered, and instant access to a unique group of individuals and entities that have affirmatively indicated an interest in the products of* [*Plaintiff*], … *a unique Twitter Feed visible only to those with access rights that displays various tweets and retweets of any individual or entity that the followers of the account lows, which provides invaluable insight into the interests of those individuals and entities and… the exclusive right to direct message or "DM" Twitter followers*"

were ancillary information not publicly available to or readily ascertainable from outside sources, and, as such, were trade secrets.[1129] Plaintiff moved to dismiss for failure to state a claim upon which relief may be granted and then countersued for defamation.[1130] The case eventually settled. Many, if not all journalists have a social media account which they use in their professional capacity to inform the public and to gather news. As they are using them in their capacity as journalists, they are of interests to their employers.

[1129] BH Media Group, Inc. v. Andy Bitter, N° 7:18-cv-388 (Aug. 6, 2018, W.D. Va.), Complaint, at 16. See Anthony C. Adornato & Andrew S. Horsfall, *Failed Strategy: Using Trade Secret Laws to Assert Ownership of Employees' Social Media Accounts in the Journalism Industry*, 9 NYU J. INTELL. PROP. & ENT. L. 62 (2019).
[1130] Michael Phillips, *Former Roanoke Times reporter countersues for more than $150,000 in Twitter case*, RICHMOND TIMES-DISPATCH, (Aug. 30, 2018), https://richmond.com/sports/college/schools/virginia-tech/former-roanoke-times-reporter-countersues-for-more-than-in-twitter/article_0404201e-ea6b-52e3-b8c5-eac46230c69c.html.

B. Journalists and Social Media

The British Broadcasting Channel (BBC) issued a new social media guide for its employees in October 2020, which applies whether employees post on social media in their professional or in their personal capacity.[1131] They are asked not to "*express a personal opinion on matters of public policy, politics, or controversial subjects."* The rules specify that disclaimers such as "My views, not the BBC's" in social media profiles and biographies are not enough to overcome the requirements of the rules. The BBC states that there will be "*tougher guidelines for some staff in news, current affairs, factual journalism, senior leadership, and a small number of presenters who have a significant public profile*."

Sport's cable network ESPN commentator Keith Olberman was suspended in February 2015 by ESPN over a series of tweets mocking Penn State University (PSU) students. A Penn State had tweeted him a link to the students' THON, the student's charity raising money for pediatric cancer, preceded by the words "*We are!"* Mr. Olbermann answered: "… *Pitiful*". This led to heated exchanges with other Twitter users, students and alumni of PSU. ESPN later suspended Mr. Olbermann for a week. His contract was not renewed, although it is not clear whether the two events are connected. In September 2015, former Red Sox and All-Star pitcher Curt Schilling was suspended by ESPN for the rest of baseball season from his baseball analyst position on Sunday night over a "meme" picture posted on his personal Twitter account, which appeared to compare Muslims with Nazis. ESPN issued a statement which read that:

[1131] *BBC issues staff with new social media guidance*, (Oct. 29, 2020), BBC NEWS, https://www.bbc.com/news/entertainment-arts-54723282.

"*At all times during the course of their engagement with us, our commentators are directly linked to ESPN and are the face of our brand. We are a sports media company. Curt's actions have not been consistent with his contractual obligations nor have they been professionally handled; they have obviously not reflected well on the company. As a result, he will not appear on ESPN through the remainder of the regular season and our Wild Card playoff game.*"[1132]

Shilling returned to ESPN for the 2016 season, albeit in the somewhat less prestigious "*Monday Night Baseball*" slot. However, ESPN fired him in April 2016 following comments he had posted on his Facebook page about a recently passed North Carolina bill which barred transgender persons to use bathrooms and locker rooms that do not correspond to their birth genders. The former player had shared on Facebook a picture of a man wearing a wig and revealing woman's clothing.[1133] This "meme" commented: "*LET HIM IN! to the restroom with your daughter or else you're a narrow-minded, judgmental, unloving racist bigot who needs to die*." Curt Schilling added this comment: "*A man is a man no matter what they call themselves. I don't care what they are, who they sleep with, men's room was designated for the penis, women's not so much. Now you need laws telling us differently? Pathetic*."[1134] Some social media users lamented that ESPN had trampled on the First Amendment by preventing Curt Shilling to exercise his right to free speech[1135] and

---

[1132] *ESPN Statement on Curt Schilling*, ESPN, (Sept. 3, 2015), http://espnmediazone.com/us/press-releases/2015/09/espn-statement-on-curt-schilling-2/ (last visited Dec. 30, 2020).
[1133] The image can be seen here: Justin Wn. Moyer, *The radicalization of Curt Schilling*, THE WASHINGTON POST, (April 21, 2016), https://www.washingtonpost.com/news/morning-mix/wp/2016/04/21/the-radicalization-of-curt-schilling/.
[1134] Richard Sandomir, *Curt Schilling, ESPN Analyst, Is Fired Over Offensive Social Media Post*, THE NEW YORK TIMES (April 20, 2016), http://www.nytimes.com/2016/04/21/sports/baseball/curt-schilling-is-fired-by-espn.html?_r=0.
[1135] *See* for example: @varepall, Twitter (April 20, 2016, 8:56 PM), https://twitter.com/varepall/status/722952303481880577.

some conservative members of the media also criticized the firing, which they saw as free speech censorship.[1136] ESPN hires journalists and former players alike to act as commenters for the sports game it features. 2016 was a Presidential election year, and the race to the White House announced itself early to be a long and contested one, as it turned out to be. Therefore, it is not surprising that ESPN issued guidelines to its employees, detailing the corporate policy for political comment on the air. The guidelines stated, inter alia, that ESPN employees:

> "*should refrain from political editorializing, personal attacks or "drive-by" comments regarding the candidates and their campaigns (including but not limited to on platforms such as Twitter or other social media). Approved commentaries on sports-specific issues, or seeking responses from candidates on relevant news issues, are appropriate. However perceived endorsements should be avoided*."[1137]

C. Athletes and Social Media

Do athletes have a right to speak their mind on social media? After George Zimmerman was found not guilty, on July 13, 2013, of the second-degree murder and manslaughter of Trayvon Martin, New York Giants wide receiver Victor Cruz tweeted: "*Thoroughly confused. Zimmerman doesn't last a year before the hood catches up to him*." He quickly deleted the tweet and later apologized for it.[1138] Atlanta Falcons wide receiver

---

[1136] Brendan Karet, *Right-Wing Media Rush To Defend Curt Schilling And Attack ESPN Following Firing For Anti-Trans Facebook Post*, MEDIA MATTERS, (April 21, 2016, 12:24AM EDT), http://mediamatters.org/research/2016/04/21/right-wing-media-rush-defend-curt-schilling-and-attack-espn-following-firing-anti-trans-facebook/210013.

[1137] Greg Rajan, *ESPN looking into Curt Schilling's comments on Hillary Clinton*, SEATTLE PI, (March 3, 2016 Updated: March 3, 2016 1:06 pm), https://www.seattlepi.com/sports/article/ESPN-looking-into-Curt-Schilling-s-comments-on-6868592.php.

[1138] Rich Cimini, *Victor Cruz: Tweet was 'wrong',* ESPN, (July 15, 2015), http://espn.go.com/new-york/nfl/story/_/id/9479415/victor-cruz-new-york-giants-says-zimmerman-tweet-wrong.

Roddy White tweeted on the same topic that "*All them jurors should go home tonight and kill themselves for letting a grown man get away with killing a kid*."[1139] He later apologized for the tweet. In Europe, Croatian soccer player Marko Livaja posted a few choice[1140] words on Facebook to fans of the Atlanta soccer team, in Bergamo, Italy, who had used his Facebook page to express their hopes that he would leave the team and return to Croatia.[1141] The player later apologized, also on Facebook, explaining that the fans had insulted his mother and called him a "*gipsy*." He added: "*I hope not to have anymore such reactions, but I hope that I will be only criticized for my performances at the club*."[1142] Sometimes, it is not the content of the message published on social media which gets the athlete in trouble, but rather what it shows. Such was the case when Formula 1 driver Nikita Mazepin posted a video on his Instagram account where he allegedly appeared to touch a woman inappropriately, without her consent.[1143] Even retired players may find out that they cannot post controversial speech on social media without consequences. In February 2020, the San Francisco Giants decided not to invite former player Aubrey Huff to a reunion of the 2010 team, which had won the 2010 World Series, citing "unacceptable" comments posed by the player on social media.[1144] One of the tweets found to be offensive

---

[1139] Nick Schwartz, *Roddy White tells jurors in Zimmerman trial to 'kill themselves'* http://ftw.usatoday.com/2013/07/roddy-white-tells-jurors-in-zimmerman-trial-to-kill-themselves.
[1140] Reportedly, the player posted: "Venite in Croazia con me, italiani bastardi" (Come with me in Croatia, Italian bastards."
[1141] *Atalanta: Livaja, insulti shock su Facebook: "Italiani bastard*", LA GAZETTA DELLO SPORT, (April 20, 2014), http://www.gazzetta.it/Calcio/Serie-A/Atalanta/20-04-2014/atalanta-livaja-insulti-shock-facebook-italiani-bastardi-80488803498.shtml.
[1142] "*Mi auguro di non avere più queste reazioni, ma spero anche di essere criticato solo per le mie prestazioni sul campo*."
[1143] PA Media, *Haas condemn F1 driver Nikita Mazepin's 'abhorrent' Instagram video*, THE GUARDIAN, (Dec. 9, 2020 06.37 EST), https://www.theguardian.com/sport/2020/dec/09/haas-condemn-formula-one-driver-nikita-mazepin-abhorrent-instagram-video.
[1144] *Giants to exclude Aubrey Huff from 2010 World Series reunion, citing 'unacceptable' tweets*, ESPN, (Feb. 17, 2020), https://www.espn.com/mlb/story/_/id/28727115/giants-exclude-aubrey-huff-2010-world-series-reunion-citing-unacceptable-tweets.

was one posted in November 2019, during the Republican primaries, where Huff posted: "*Getting my boys trained up on how to use a gun in the unlikely event @BernieSanders beats @realDonaldTrump in 2020. In which case knowing how to effectively use a gun under socialism will be a must. By the way most the head shots were theirs. @NRA @WatchChad #2ndAmendment*."[1145]

Players may be in trouble over a simple "like", on an uncontroversial topic, if use of social media during game is forbidden by their teams. Third baseman Pablo Sandoval was benched in June 2015 by the Boston Red Sox for one game for having used his *Instagram* account to "like" two pictures during a game. He had gone to the bathroom at the seventh inning and had used his phone to check out his *Instagram* feed, which is against the policies of both the Red Sox and Major League Baseball (MLB).[1146] The MLB has its own social media policy, first published in 2012,[1147] which prohibits, inter alia, displaying or transmitting speech condoning use of illegal substances, or content "*that reasonably could be viewed as discriminatory, bullying and/or harassing based on race, color, ancestry, sex, sexual orientation, national origin, age, disability, religion, or other categories protected by law and/or which would not be permitted in the workplace, including, but not limited to, [c]ontent that could contribute to a hostile work environment (e.g., slurs, obscenities, stereotypes) or reasonably could be viewed as retaliatory*. "An MLB player violating the

[1145] Aubrey Huff, @aubrey_huff, Twitter, (Nov. 26 2019, 12: 39 AM), https://twitter.com/aubrey_huff/status/1199200986658553857.
[1146] Steve Silva, *Pablo Sandoval admits to and apologizes for Instagram use during Red Sox game*, BOSTON.COM, (June 18, 2015, 12:53PM), http://www.boston.com/sports/baseball/redsox/2015/06/18/was-pablo-sandoval-instagram-during-wednesday-red-sox-game/kjgeTVKMq73NgoHsxW9jBM/story.html.
[1147] Craig Calcaterra, *Major League Baseball releases its social media policy — and it's pretty good*, NBC SPORTS, (Mar 14, 2012, 4:00 PM EDT), https://mlb.nbcsports.com/2012/03/14/major-league-baseball-releases-its-social-media-policy-and-its-pretty-good/. The MLB Social Media Policy is available at http://content.mlb.com/documents/1/0/2/296982102/Social_Media_Policy.pdf.

policy may be disciplined by his club or by the MLB Commissioner under Article XII of the Basic Agreement, which states that a player may be subjected to disciplinary action for just cause. The National Hockey League (NHL) published its *Social Media Policy for League and Club Personnel* on September 14, 2011, which had been collectively bargained with the National Hockey League Players Association (NHLPA). Players are prevented to use social media on game days during a "blackout period" starting two hours prior to opening face-off until the completion of them post-game media obligations. The policy suggests that blackout period for hockey operations staff should start at 11 a.m. on game days.[1148] Professional sports teams have their own social media accounts, which they use to share information, connect with their fans, and even live-tweet a game. In that case, the account is turned over to an employee, often the social media manager of the organization, who becomes the live spokesman of the organization, during games, which may last late into the night, when emotions can fire up quickly and typing to express them is easy. The digital communication media manager of the Houston Rockets basketball team, Chad Shanks was fired from his position in April 2016 over a tweet he posted during a Rockets game against the Dallas Maverick, which included the emojis of a horse and a gun, pointing out at the horse, and read: "*Shhhhh. Just close your eyes. It will all be over soon*."[1149] Mr. Shanks explained that he "*meant it to just be a play on taking an old horse out to pasture that would get our fans even more pumped up and agitate Mavs fans*."[1150] His employer quickly tweeted

[1148] *NHL institutes new social media policy*, NHL.COM, (Sept. 15, 2011, 12:00PM), http://www.nhl.com/ice/news.htm?id=588534.
[1149] Adi Joseph, *Exclusive Q&A: Rockets' fired social media manager explains ill-fated tweet*, SPORTING NEWS (April 29, 2016), http://www.sportingnews.com/nba-news/4642958-rockets-tweet-mavericks-social-media-manager-fired-chad-shanks.
[1150] *Ibid.*

an apology to the fans and the opposing team,[1151] and later fired Mr. Shanks, who said: "*I had an impulse toward the end of the game last night and should've resisted it*." This story is indicative of the dangers of expressing oneself using emojis. Images, more than words, may easily be misinterpreted.[1152] The International Olympic Committee (IOC) had published the *IOC Social Media, Blogging and Internet Guidelines for participants and other accredited persons at the London 2012 Olympic Games* prior to the 2012 Olympic Games in London.[1153] The guidelines addressed the general tone of these messages as its target public was advised that "[*p*]*ostings, blogs and tweets should at all times conform to the Olympic spirit and fundamental principles of Olympism as contained in the Olympic Charter, be dignified and in good taste, and not contain vulgar or obscene words or images.*" The guidelines also addressed the content of these messages, which "*must not report on competition or comment on the activities of other participants or accredited persons, or disclose any information which is confidential or private in relation to any other person or organization*." It also addressed possible privacy and right to publicity issues, as "*photos of the athletes themselves or other accredited persons in the Olympic Village can be posted, but if any other persons appear in the photo, their prior permission must be obtained by the person posting such photo*." The IOC policy also addressed trademark issues, instructing its target public not to use "*the Olympic Symbol – i.e. the five interlaced rings, which is the property of the IOC*

[1151] "*Our Tweet earlier was in very poor taste & not indicative of the respect we have for the @dallasmavs & their fans. We sincerely apologize.*",@HoustonRockets, Twitter (April 28, 2016, 11:17 PM), https://twitter.com/HoustonRockets/status/593252813712404480?ref_src=twsrc%5Etfw

[1152] For instance, the emoji of an eggplant may have a crude signification, and naively using one, say, to express one's love of ratatouille, could be misinterpreted.

[1153] The International Olympic Committee, *IOC Social Media, Blogging and Internet Guidelines for participants and other accredited persons at the London 2012 Olympic Games*, available at http://www.olympic.org/Documents/Games_London_2012/IOC_Social_Media_Blogging_and_Internet_Guidelines-London.pdf .

– on their postings, blogs or tweets on any social media platforms or on any websites. Participants and other accredited persons may use the word "Olympic" and other Olympic-related words on their postings, blogs or tweets on any social media platforms or on their websites, as a factual reference, provided that the word "Olympic" and other Olympic-related words are not associated with any third party or any third party's products or services."

Failure to abide by social media policies may lead to dismissal of the athletes, even at a young age.[1154] A teenager who had been selected to her high school's cheerleading team was dismissed from it after having posted a *SnapChat* story showing her and four other teenagers, who had also been selected for the team, singing to lyrics of Big Sean's song "I.D.F.W.U."[1155] They all wore the cheerleader tee-shirts that had just been given to them after having been selected to the team. The story was sent to thirty to forty of her Snapchat contacts, but deleted after half an hour or so. The school's "Cheer and Stunt Squad Constitution" stated that "[*m*]*embers will be dismissed for improper social media usage*."[1156] The teenager had been informed during the cheerleading squad tryouts, that the purpose of having such a policy was to avoid inappropriate social media usage by the school's cheerleaders which may exacerbate "*the sometimes violent rivalry* [*the school*] *had with its neighboring high school*… "Also, after making the squad, the teenager had been instructed by the school's administrators not to post about making the team on social media, before

---

[1154] *See* also Dan Levin, *A Racial Slur, a Viral Video, and a Reckoning*, THE NEW YORK TIMES, (Dec. 26, 2020), https://www.nytimes.com/2020/12/26/us/mimi-groves-jimmy-galligan-racial-slurs.html, which relates how a high school student who had used a racist slur in a Snapchat video saw her acceptance to the cheerleading team of her University of choice removed, followed by the University asking the student to withdraw.

[1155] The teenagers sung: "*I don't fuck with you, you little stupid ass bitch, I ain't fucking with you*." Johnson v. Cache Cty. Sch. Dist., 323 F. Supp. 3d 1301, 1309 (D. Utah 2018).

[1156] Johnson v. Cache Cty. Sch. Dist., 323 F. Supp. 3d 1301, (D. Utah 2018).

the official announcement the next day. The next day, the school administrators were informed about the video by a former member of the cheer team, who had been sent the video by other students.[1157] The administrators interpreted the video as boasting to other students that the teenagers had been selected to the squad. The teenager was then dismissed from the team.[1158] She sued the school, claiming that her First Amendment's right to free speech had been violated, arguing that "*schools cannot punish students for private, out-of-school speech that does not cause substantial, material disruption to school activities*."[1159] The Court was not convinced, noting that "*the student handbook, and cheer constitution reinforce the importance of courtesy, character, honor, and humility*."[1160] The court also noted that the Supreme Court recognizes that schools may regulate student speech for four main reasons. They may do so (1) if it materially and substantially disrupts the work and discipline of the school,[1161] or if (2) speech made on campus is vulgar or offensive speech, as schools have a mission to "*inculcate the habits and manner of civility*" and to "*teach students the boundaries of socially appropriate behavior*."[1162] Schools may also regulate student (3) "*expressive activities that students, parents, and members of the public might reasonably perceive to bear the imprimatur of the school… so long as* [*the school's*] *actions are reasonably related to legitimate pedagogical concerns*."[1163] Schools can also (4)

[1157] The Court noted that "*SnapChat stories can be saved and screen recorded by anyone in the SnapChat story and sent to others*," Johnson, at 1309.
[1158] While the other four girls in the video stated their regret, and were provided an opportunity to join the team, if apologizing and serving fifty hours of community services, Plaintiff had not been provided this opportunity because she "*was unrepentant and insistent that the post was accidental and unintentional*," Johnson, at 1310.
[1159] Johnson, at 1318.
[1160] *Ibid*.
[1161] Tinker v. Des Moines Independent Community School Dist., 393 U. S. 513 (1969).
[1162] Bethel School District No. 403 et al. v. Fraser, a Minor, et al., 478 U.S. 675, 683,685 (1986).
[1163] Hazelwood School Dist. v. Kuhlmeier,484 U. S. 271, 273 (1988).

punish speech advocating illegal drug use.[1164] Plaintiff argued that *Tinker* was the controlling case, but the court was not convinced. In Tinker, the armbands worn by the students was political speech, expressing their opinion against the Vietnam war. Instead, the Snapchat Story video was "*not political commentary or related to a public issue. In fact, it could be viewed as more vulgar than the offensive speech in Fraser. It is thus subject to* lesser *protection than the "nondisruptive, passive expression of a political viewpoint in Tinker*,"[1165] and denied Plaintiff's motion for temporary restraining order and preliminary injunction.

In a somewhat similar case, however, the Third Circuit Court of appeals held in June 2020[1166] in favor of a student who had posted a vulgar message[1167] on Snapchat to express her frustration at not having been selected to her high school varsity cheerleading team and had been dismissed from the junior varsity team for having violated the team rules, which required cheerleaders to:

> "*have respect for [their] school, coaches, ... [and] other cheerleaders"; avoid "foul language and inappropriate gestures"; and refrain from sharing "negative information regarding cheerleading, cheerleaders, or coaches ... on the internet*," and the school rules, which required student athletes to "*conduct[ ] themselves in such a way that the image of the Mahanoy School District would not be tarnished in any manner*."

[1164] Deborah Morse et al., v. Joseph Frederick, 551 U.S.393 (2007).
[1165] Johnson, at 1319.
[1166] B.L. v. Mahanoy Area School District, 964 F.3d 170 (2020).
[1167] The message read "*Fuck school fuck softball fuck cheer fuck everything*" and featured the student "flipping the bird."

The student sued the School District, claiming, inter alia, that the rules were overbroad, viewpoint discriminatory, and unconstitutionally vague, and that her dismissal had violated the First Amendment. The District Court granted her summary judgment, and the Court of appeals affirmed. The Third Circuit reasoned that the social media post had been "*off-campus,*" cautioning, however, that "*the schoolyard's physical boundaries are not necessarily coextensive with the "school context."*[1168] The Snapchat message had not taken place in a *"school-sponsored" forum*,[1169] or in a context that "*bear[s] the imprimatur of the school*,"[1170] and nor had the message be sent on a platform owned or operated by the school. While the message mentioned the school and reached the school's students and officials, "*those few points of contact are not enough*,"[1171] and a school cannot "*seeks to control student speech using even modest measures.*"[1172] The School District had argued that the student First Amendment rights had not been violated because if fell within the *Tinker* narrow exception carved out by the Supreme Court in *Tinker* "*in light of the special characteristics of the school environment*… [*as some form of speech can] "interfere[ ] ... with the rights of other students to be secure and to be let alone*."[1173] While the Supreme Court recognized in *Tinker* the right of schools to regulate speech that "*would materially and substantially interfere with the requirements of appropriate discipline in the operation of the school*,"[1174] such is the case in the Third Circuit only if the school shows *"a specific and*

[1168] B.L., at 178 , citing JS Ex Rel. Snyder v. Blue Mountain School Dist., 650 F. 3d 915. 932 (3d Circ. 2011) (en banc).
[1169] Citing Bethel School Dist. No. 403 v. Fraser, 478 US 675, 677 (1986).
[1170] Citing Hazelwood School Dist. v. Kuhlmeier, 484 US 260, 271 (1988).
[1171] B.L. at 181.
[1172] B.L. at 183.
[1173] Tinker v. Des Moines Independent Community School Dist., 393 US 503, 506 (1969).
[1174] Tinker, at 509.

significant fear of disruption."[1175] The Third Circuit held that "*Tinker does not apply to off-campus speech—that is, speech that is outside school-owned, -operated, or -supervised channels and that is not reasonably interpreted as bearing the school's imprimatur*, "[1176] and that the student had not waived her First Amendment Rights when agreeing to the school and the team rules. The cheerleading team's "Respect Rule" asked cheerleaders to respect their school, coaches, teachers, and other cheerleaders and teams, and urged them to remember that they "*are representing [their] school when at games, fundraisers, and other events… [and that] [g]ood sportsmanship will be enforced[;] this includes foul language and inappropriate gestures*." The Court found that this rule did not "*cover a weekend post to Snapchat unconnected with any game or school event and before the cheerleading season had even begun*."[1177] The "Negative Information Rule" informed students that the school did not tolerate "*any negative information regarding cheerleading, cheerleaders, or coaches placed on the internet*." The Court found it also inapplicable to the speech at stake, even though it covered off-campus speech, because it was a rule about information, not expressions of opinion or emotion. As for the school's "Personal Conduct Rule," which informed athletes that they must "conduct[ ] themselves in such a way that the image of the… School District would not be tarnished in any manner," the Court found it not to be applicable as well as it only applied "*during the sports season*" and the Snapchat message had been posted before the season had begun. The School District filed a petition for certiorari to the Supreme Court in August 2020, asking the Supreme Court to grant it so that it can review whether

[1175] J.S. ex rel. Snyder v. Blue Mountain Sch. Dist, at 926.
[1176] B.L. at 189.
[1177] B.L. at 193.

Tinker applies to student speech that occurs off campus.[1178] This would considerably extend the power of schools to police their student's speech, including their social media posts posted off-campus, on the weekend, and on topics not related to school.

D. Judges, Attorneys, and Social Media

Attorneys may use social media accounts to advertise their practice, to help them practicing. It appears, however, that they mostly used for personal purposes.[1179] Regardless of the reason, they can get in trouble for using social media, even getting disbarred. The Supreme Court of Louisiana disbarred on June 30, 2015 a Louisiana attorney who had used her Twitter account to publicize a contentious children custody case, which included allegations of sexual abuse. The attorney represented the mother who claimed that her former husband had abused their two minor daughters. The attorney had started an online petition on < change.org>, asking signatories to urge the two judges involved in the case, one in Louisiana, the other in Mississippi, to look at the evidence which had allegedly been presented to them, which would prove that the two minors had been abused. The petition, which the attorney had signed, had been faxed to the judges. The attorney also used her Twitter account to publicize her campaign. For instance, on a particular day, she had tweeted thirty messages, writing, *inter alia*: "*GIMME GIMME GIMME Evidence! Want some? I got it. Think u can convince a judge to look at it? Sign this petition*" and provided a link to the

---

[1178] Brief for Petitioner, Mahanoy Area School District v. B.L., No. 20-255, 2020, available at https://www.supremecourt.gov/DocketPDF/20/20-255/151619/20200828144703420_Mahanoy%20Cert%20Petition%20-%20Final.pdf.

[1179] *See 2019 Websites & Marketing* , TECHREPORT 2019, AMERICAN BAR ASSOCIATION : "*The most commonly reported use of Twitter was for social or personal reasons (76%), followed by education and awareness (49%), and career development/networking (24%)".* https://www.americanbar.org/groups/law_practice/publications/techreport/abatechreport2019/websitesmarketing2019.

petition. Some of the tweets linked to an audio recording of the minor children discussing alleged sexual abuse with their mother. The attorney even asked a Hollywood actor for help:" "*I am SO going 2 have 2 change jobs after this @russellcrowe come on! I'm risking sanctions by the LA Supreme Court; u could be a HUGE help*." That prophecy turned out to be true as she was disbarred by the Supreme Court of Louisiana.

The Louisiana Attorney Disciplinary Board had only recommended a suspension for one year and a day,[1180] finding that the attorney's conduct had violated Rule 3.5(a), Rule 3.5(b) and Rule 8.4(a) of the Louisiana Rule of Professional Conduct,[1181] when she "*used the internet, an online petition and social media to spread information, some of which was false, misleading and inflammatory, about [both Judges'] handling and rulings in pending litigation*." Rule 3.5(a), prohibits attorneys to "*seek to influence a judge, juror, prospective juror or other official by means prohibited by law.*" Rule 3.5(b) prohibits attorneys from having *ex parte* communications with a judge during proceedings, and Rule 8.4(a) prohibits attorneys to knowingly assist or induce another to do so or doing so through the acts of another. For the Disciplinary Board, "[*t*]*he clear intent of Respondent's online campaign was an attempt to influence the judges' future rulings in the respective cases, and to do so through improper ex parte communication directed at the judges*." The Board had also found that the attorney had violated Rule 8.4(c),which prohibits attorneys from engaging in conduct involving dishonesty, fraud, deceit or misrepresentation, because she had "*disseminated false, misleading and inflammatory information on the internet and through social media about* [*both Judges*] *and their handling of these pending domestic proceeding*" when she had

[1180] Louisiana Attorney Disciplinary Board, *In Re Joyce Nanime McCool*, number 13-DB-059.
[1181] La. R. Prof. Conduct (1992).

posted an article online claiming that both judges had refused to look at evidence that the two minors had been sexually abused by their father. However, this evidence had not been presented to the judge. The Board had also found that the attorney had violated Rule 8.4(d), which prohibits attorneys from engaging in conduct prejudicial to the administration of justice, because she:

> "*used the internet and social media in an effort to influence* [*both Judges'*] *future rulings in pending litigation. Respondent's conduct threatened the independence and integrity of the court and was clearly prejudicial to the administration of justice*" and "*Respondent also used her Twitter account to publish multiple tweets linking the audio recordings of the minor children discussing alleged sexual abuse; to publish false, misleading and inflammatory information about* [*both Judges*]*, and to promote the online petition, all of which was designed to intimidate and influence the judges' future rulings in the underlying proceedings.*"

As such, the attorney had "*knowingly, if not intentionally, spearheaded a social media blitz in an attempt to influence the judiciary*."

The attorney then argued unsuccessfully in front of the Louisiana Supreme Court that such actions were protected by the First Amendment. The Supreme Court of Louisiana found that the attorney had indeed violated Rules 3.5(a) and (b) and Rule 8.4(a) of the Rules of Professional Conduct, as the telephone calls some people made to the Judges as a result of the attorney's campaign, the email, and the faxed petitions "*constitute prohibited ex parte communication induced and/or encouraged by respondent. Coupled with her social*

media postings, [*The Supreme Court*] *further conclude[d] respondent 's online activity amounted to a viral campaign to influence and intimidate the judiciary*."[1182]

In Iowa, a judge ruled to delay a civil trial by several months, following a posting of Facebook of the attorney representing the victim of an alleged assault.[1183] The post read:

> *"I am going to trial on Monday in a case where [X] slept with his client and then beat her up. … It is one of those cases where I feel duty bound to take it because I have an obligation to my own profession. The bar (association) Grievance Commission recommended that his license be suspended for four years. The Supreme Court affirmed their findings but reduced the penalty to 18 months. Yet another example of why an all-white, all-male court really needs a woman. I hope a jury will be a little harder on him!"*

The opposing counsels had argued that the post would unduly influence jurors. Indeed, social media can be used unethically by attorneys, especially trial attorneys. The Commercial and Federal Litigation Section of the New York State Bar Association (NYSBA) first published in 2014 its *Social Media Ethics Guidelines,* which were updated in 2017 and again in 2019.[1184] These Guidelines are best practice rules only for a number of situations where an attorney may use social media in his or her practice. For instance, Guideline No.

[1182] *In Re Joyce Nanine McCool*, 172 So.3d 1058, 1073 (2015). *See* Debra Cassens Weiss*, Lawyer is disbarred for 'social media blitz' intended to influence custody case and top state court*, ABA JOURNAL (July 8, 2015, 5:45 am CDT), https://www.abajournal.com/news/article/lawyer_is_disbarred_for_social_media_blitz_intended_to_influence_custody.

[1183] Debra Cassens Weiss, *Judge delays civil trial because of lawyer's Facebook post*, THE ABA JOURNAL, (July 14, 2015, 10:37 AM CDT), http://www.abajournal.com/news/article/judge_delays_civil_trial_because_of_lawyers_facebook_post.

[1184] *Social Media Ethics Guidelines of the Commercial and Federal Litigation Section of the New York State Bar Association*, THE NEW YORK STATE BAR ASSOCIATION, (June 20, 2019), https://nysba.org/app/uploads/2020/02/NYSBA-Social-Media-Ethics-Guidelines-Final-6-20-19.pdf.

4.A instructs attorneys that they can "*view the public portion of a person's social media profile or view public posts even if such person is represented by another lawyer*" and Guideline No. 4.B instructs them that they can join a social media network for the purpose of obtaining information about a witness. In France, article 10.5 the National Regulations of the lawyers' profession (*Règlement intérieur national de la profession d'avocat*) states that « [*t]he lawyer participating in a blog or an online social network must respect the essential principles of the profession*." As French attorneys swear to exercise their functions "*with dignity, conscience, independence, probity and humanity*, " their social media posts must respect theses values, including the "catch-all" requirement to show "humanity," which surely encompasses politeness and respect for other's people's opinions and point of views. As such, posts denigrating minorities would surely be in breach of ethics. In the U.S., a Tennessee attorney, who had been suspended and whose case was to be reviewed by the Tennessee Board of Professional Responsibility, filed a motion to disqualify a disciplinary counsel, member of the Tennessee Board of Professional Responsibility, alleging that his virulent ant-Muslims opinions, as stated on his Twitter account,[1185] showed that his "*extreme prejudice*" disqualified him from the case.[1186] The suspended attorney is married to a Muslim, and that couple has two children raised in a Muslim household.[1187] The disciplinary counsel deleted his entire Twitter account shortly thereafter, a move called by Plaintiff a subsequent motion, as "*knowingly and immediately destroy*[*ing*] e*xtensive*

[1185] Screenshots of the tweets part of the exhibits can be seen at https://tennessee.tylerhost.net/ViewDocuments.aspx?FID=929ff9c3-e22e-4ee7-87a0-4ace07eb7a69.
[1186] The motion argued that "[*t*]*he Constitution… forbids prosecutions that are influenced by bigotry*," citing *United States v. Gist*, 382 F. App'x 181, 183 (3d Cir. 2010): "*Although prosecutors enjoy wide discretion, they may not prosecute based on a defendant's 'race, religion, or other arbitrary classification*.'"
[1187] See Debras Cassens Weiss, *Disciplinary counsel resigns after filing alleges he is a 'proud anti-Muslim bigot'*, ABA JOURNAL, (Dec. 14, 2020, 12:25 pm CST), https://www.abajournal.com/news/article/disciplinary-counsel-resigns-after-filing-alleges-he-is-a-proud-anti-muslim-bigot.

evidence regarding and relating to it—including all of the original statements constituting [*Plaintiff*]'s *exhibits and thousands of additional statements that* [*Plaintiff*] *specifically identified as potentially relevant to this matter*."[1188] The disciplinary counsel resigned from the Board a few days later.

Judges are required to follow their own ethics rules, the Code of Judicial Conduct, and even seasoned judges active on social media may get in trouble. A Minnesota judge who had been appointed to the bench in 1983, had retired in 2006 and had been assigned to serve the State as a senior judge, and who had no disciplinary history, was publicly reprimanded in November 2015 for having posted comments on his Facebook page about cases to which he was assigned.[1189] The judge stated to the Minnesota Board on Judicial Standards that he believed his comments to be private, not public, as the posts were available to about eighty people, among them family members and members of his church, but were in fact public.[1190] The Board found published these comments had violated several provisions of the Code of Judicial Conduct:

"*Rule 1.2, requiring a judge to promote confidence in the independence, integrity, and impartiality of the judiciary;*

[1188] Petitioner's Response to the Board of Professional Responsibility's Motion to Allow [Jerry Morgan] to Withdraw as Attorney of Record in this Proceeding, Brian P. Manookian v. Board of Professional Responsibility of the Supreme Court of Tennessee, Case No.: 20-0833-I (Dec. 3, 2020),(Chancery Court for Davidson County, Tennessee), available at https://cdn.baseplatform.io/files/base/scomm/nvs/document/2020/12/Response_to_Motion_to_Withdraw.5fd1550ab7633.pdf (last visited Dec. 30, 2020).

[1189] Martha Neil, *Senior judge reprimanded for Facebook posts about his trials*, ABA JOURNAL, (Nov. 20 2015, 04:10 PM CST), http://www.abajournal.com/news/article/senior_judge_reprimanded_for_facebook_posts_about_his_trials

[1190] MINNESOTA BOARD ON JUDICIAL STANDARDS, In the Matter of Senior Judge Edward W. Bearse PUBLIC REPRIMAND, File No. 15-17http://www.bjs.state.mn.us/file/public-discipline/1517-public-reprimand.pdf .

Rule 2.1, requiring that the duties of judicial office take precedence over a judge's personal activities;

Rule 2.8(B), requiring a judge to be dignified and courteous with litigants;

Rule 2.10(A), prohibiting a judge from making a public statement that might reasonably be expected to affect the outcome or impair the fairness of a matter pending or impending in any court; and

Rule 3.1(A) and (C), prohibiting a judge from participating in activities that interfere with the proper performance of the judge's judicial duties or that would appear to a reasonable person to undermine the judge's independence, integrity, or impartiality."

Judges may not "friend" litigants on Facebook: a judge who had send in 2014 a Facebook friend request to a litigant appearing in her court was removed from the case.[1191] The litigant moved to disqualify the judge, which was denied by the trial court, but the Court of appeals quashed the order denying the motion to disqualify.[1192] Indeed, such a request can lead to ex parte communications, and being "Facebook friend" with a litigant may very well "*convey the impression that any person or organization is in a position to influence the judge*," as stated by Rule 2.4(C) of the American Bar Association (ABA) Model Code of Judicial Conduct. The American Bar Association published its Formal Opinion 462,

[1191] Stephanie Francis Ward, Judge removed from divorce case after sending one party a Facebook friend request, ABA JOURNAL, (January 29, 2014, 11:40 am CST), https://www.abajournal.com/news/article/judge_removed_from_divorce_case_after_party_rebuffs_facebook_friend_request.

[1192] Chace v. Loisel, 170 So. 3d 802 (Fla. Dist. Court of Appeals, 5th Dist. 2014).

Judge's Use of Electronic Social Networking Media, on February 21, 2013,[1193] which stated that:

> "[*a*] *judge must also take care to avoid comments and interactions that may be interpreted as ex parte communications concerning pending or impending matters in violation of Rule 2.9(A)* [*on ex parte communication*] *and avoid using any [electronic social media] site to obtain information regarding a matter before the judge in violation of Rule 2.9(C).*"[1194]

The issue of whether a judge can be "Facebook friend" with an attorney has been debated in courts. In one case, a criminal defendant had moved to disqualify the trial judge because he was "Facebook friend" with the prosecutor handling the case, arguing that the judge could not "*be fair and impartial*."[1195] The Court of appeals quashed the order denying disqualification of the trial judge, citing an opinion of the Florida Judicial Ethics Advisory Committee[1196] which had found that "*the Florida Code of Judicial Conduct precludes a judge from both adding lawyers who appear before the judge as "friends" on a social networking site and allowing such lawyers to add the judge as their "friend*." For the Committee, an attorney appearing as a "friend" on the judge's social networking page "[*t*]*o the extent that such identification is available for any other person to view*" would violate Florida Code of

---

[1193] American Bar Association, *Formal Opinion 462, Judge's Use of Electronic Social Networking Media*, (Feb. 21, 2013), available at https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/formal_opinion_462.authcheckdam.pdf.

[1194] Rule 2.9 (C) states that "*A judge shall not investigate facts in a matter independently, and shall consider only the evidence presented and any facts that may properly be judicially noticed*."

[1195] Domville v. State, 103 So. 3d 184, 185 (Fla.Dist. Court of Appeals, 4th Dist. 2012).

[1196] Fla. JEAC Op.2009-20 (Nov. 17, 2009).

Judicial Conduct Canon 2B, which states that "[*a*] *judge shall not... convey or permit others to convey the impression that they are in a special position to influence the judge*."

However, the Florida Supreme Court held in 2018, in another case, that the fact that "*a trial judge is a Facebook "friend" with an attorney appearing before the judge, standing alone, does not constitute a legally sufficient basis for disqualification.*"[1197] The Supreme Court of Florida noted that, in real life, "*the mere existence of a friendship between a judge and an attorney appearing before the judge, without more, does not reasonably convey to others the impression of an inherently close or intimate relationship*."[1198] When this friendship in online, relations may be even less close or intimate, as "[*t*]*he establishment of a Facebook "friendship" does not objectively signal the existence of the affection and esteem involved in a traditional "friendship*,"[1199] and "*it is regularly the case that Facebook "friendships" are more casual and less permanent than traditional friendships*."[1200] The Court noted that the Florida Judicial Ethics Advisory Committee "*was one of the first to advise that judges were prohibited from adding attorneys who appear before them as "friends" on their Facebook page or from allowing attorneys who appear before them to add them as "friends" on the attorneys*,"[1201] but called it an "*overarching concern*," arguing further that the Florida Judicial Ethics Advisory Committee had "missed the intrinsic nature of Facebook "friendship."[1202]

[1197] Law Office of Herssein v. US Auto, 271 So. 3d 889, 891 (Fla. Sup. Ct. 2018).
[1198] Law Office of Herssein, at 894.
[1199] Law Office of Herssein, at 896.
[1200] Law Office of Herssein, at 896.
[1201] Law Office of Herssein, at 898.
[1202] Law Office of Herssein, at 898.

In France,[1203] the Compendium of the Judiciary's Ethical Obligations states that "[*m*]*embers of the judiciary, who are not Internet users like any others, must be vigilant in their use of social networks, particularly when they express opinions under their identity and in the capacity of a judiciary member.*"[1204] Indeed, some judges post on social media under their real names, as Grenoble District Attorney Éric Vaillant (@egajvpr on Twitter) or the President of the Rennes Court of appeals Xavier Ronsin (@xavierRonsin on Twitter). Others posts, while divulging they are judges, post anonymously. What about social media use by judges inside a court room? Two judges, one trial judge, one prosecuting judge, were disciplined by the Supreme Judicial Council (*Conseil supérieur de la magistature*, thereafter CSM) for having communicated together by tweets in 2012 during a criminal trial. The CSM noted in 2014, in its Annual Report, that "*these messages* [*had*] *been widely relayed by the press and* [*had*] *publicized the relationship and connivance between these two magistrates*."[1205] The Annual Report quoted the disciplinary body for the trial judges (*CSM Siège*), for which:

> "*the use of social networks, including under the guise of pseudonyms, cannot free the magistrate from the duties of his or her state, in particular from her or his obligation of secrecy, a pledge for litigants of his or her impartiality and neutrality, particularly during the conduct of the trial; that this use is all the more inappropriate as the messages*

[1203] *See* Fabrice Defferrard, *Réseaux sociaux et professionnels du droit : le risque pénal*, Dalloz IP/IT, 471, (2019) ; Geoffray Brunaux, *Réseaux sociaux et professionnels du droit : le risque disciplinaire*, Dalloz IP/IT, 476, (2019) ; Jacques Martinon, *Le réseau social, faux-ami du magistrat judiciaire ?* , Dalloz IP/IT, 480, (2019) ;
[1204] Article 13 of Chapter II of the Compendium, available in English at http://www.conseil-superieur-magistrature.fr/sites/default/files/atoms/files/gb_compendium.pdf.
[1205] CONSEIL SUPÉRIEUR DE LA MAGISTRATURE, RAPPORT D'ACTIVITÉ 2014, at 179, (2015) available for download at http://www.conseil-superieur-magistrature.fr/publications/rapports-annuels-dactivite?page=1.

exchanged can be read in real time by people outside the judicial institution and they identify both their authors and the circumstances of their transmission."[1206]

The disciplinary body for the prosecuting judges, the *CSM Parquet*, stated that:

"*the alleged anonymity that some social networks would bring cannot free the magistrate from the duties of his or her state, in particular from his or her obligation of secrecy, a pledge for litigants of his or her impartiality and neutrality, especially during the course of the trial*."[1207]

The 2018 annual report of the CSM addressed social media usage by magistrates in an annex.[1208] It noted that, while judges must be prudent when communicating on social media:

"[*t*]*he degree of prudence is assessed differently depending on whether the magistrate expresses herself or himself on social networks without stating his or her capacity to deal with subjects having nothing to do with his activity or, on the contrary, if she or he states this quality to comment on judicial or legal news."*

If a magistrate uses her or his real name on social media, then she or he must "*tak*[*ing*] *care, particularly in the creation of his profile … and in his or her messages, not to cast doubt on her or his impartiality in the disputes she or he deals with*." If the judge chooses not to use her or his name, he or she must be as prudent, as "*the magistrate is no less identifiable by cross-referencing*." In both instances, however, the judge "*must keep in mind that he can be*

[1206] Ibid.
[1207] Ibid.
[1208] CONSEIL SUPÉRIEUR DE LA MAGISTRATURE, RAPPORT D'ACTIVITÉ 2018, 157-159, (2019) available for download at http://www.conseil-superieur-magistrature.fr/publications/rapports-annuels-dactivite.

identified." The report also remined judges that their social media posting may reach a wider audience than the one they had intended to reach, for instance through screen captures, and "[*a*]*ny message disseminated on social networks immediately escapes its author and can be disseminated widely, without his or her authorization, including if she or he has deleted it*."

Some attorneys are politicians, but their bar ethics apply even then. The day before Michael Cohen, former attorney to President Donald Trump, testified before the House Oversight Committee, Florida attorney and Representative Matt Gaetz tweeted: "*Hey @MichaelCohen212 – Do you wife & father-in-law know about your girlfriends? Maybe tonight would be a good time for that chat. I wonder if she'll remain faithful when you're in prison. She's about to learn a lot*…"[1209] The same day, Speaker Nancy Pelosi posted on Twitter that she "*encourage*[*d*] *all Members to be mindful that comments made on social media or in the press can adversely affect the ability of House Committees to obtain the truthful and complete information necessary to fulfill their duties*."[1210] Matt Gaetz later apologized for that tweet, on Twitter,[1211] but the Florida Bar opened nevertheless an investigation.[1212] The grievance committee later dismissed the complaint, but wrote, in the

---

[1209] Maggie Haberman and Nicholas Fandos, *On Eve of Michael Cohen's Testimony, Republican Threatens to Reveal Compromising Information*, THE NEW YORK TIMES, (Feb. 26, 2019), https://www.nytimes.com/2019/02/26/us/politics/michael-cohen-testimony.html?module=inline.
[1210] https://twitter.com/SpeakerPelosi/status/1100545912697511938. This tweet referenced an earlier tweet sent by Speaker Pelosi which read: "*Michael Cohen will come before the @OversightDems & @HouseIntel Committees next week. Congress has an independent duty under the Constitution to conduct oversight of the Executive Branch, and any efforts to intimidate family members or pressure witnesses will not be tolerated*." https://twitter.com/SpeakerPelosi/status/1098378159685337090.
[1211] @mattgaetz, Twitter, (Feb 27, 2019, 9:18 PM), https://twitter.com/mattgaetz/status/1100943221390225408.
[1212] Nicholas Fandos, *Florida Bar Will Investigate Matt Gaetz's Threat Against Cohen*, THE NEW YORK TIMES, (Feb. 27, 2019), https://www.nytimes.com/2019/02/27/us/politics/matt-gaetz-cohen.html?smtyp=cur&smid=tw-nytimes.

letter addressed to Representative Gaetz, that his conduct "*was not consistent with the high standards of our profession, and ...[his] actions do not reflect favorably on [him] as a member of the Florida Bar*."[1213] Politicians, even if they are not attorneys, have their own social media to respect.

E. Politicians and Social Media

Politicians are now using social media platforms to communicate with their constituents and the public.[1214] Many legislatures now have a social media policy.[1215] For instance, the Alaska Legislature Social Media Guideline advises to "[*w*]*rite what you know. Make sure you write and post about your areas of expertise*" and to "[*a*]*lways stop and pause, thinking before posting*."[1216] The Social Media Comment Policy of the California Senate, which applies to '[*a*]*ny social media page that is established or maintained by a Senator or Senate staff using legislative resources*,"[1217] prohibits, inter alia, posting" [*c*]*ampaign content, including content urging or opposing the nomination or election of a candidate or the qualification or passage of a ballot measure"* and "[*c*]*ontent that is unrelated to the topic being discussed*." This concern is shared by the Utah House of Representatives, which Social

---

[1213] A copy of the letter is available at Jim Rosica, *Florida Bar panel calls Matt Gaetz 'unprofessional, reckless, insensitive'*, FLORIDA POLITICS, (Aug. 17, 2019), https://floridapolitics.com/archives/303569-florida-bar-panel-matt-gaetz.

[1214] *See Congress Soars to New Heights on Social Media*, Patrick van Kessel, Regina Widjaya, Sono Shah, Aaron Smith and Adam Hughes, THE PEW RESEARCH CENTER, https://www.pewresearch.org/internet/2020/07/16/congress-soars-to-new-heights-on-social-media/, which found that "*the typical member of Congress .. tweets nearly twice as often (81% more), has nearly three times as many followers and receives more than six times as many retweets on their average post* " as four years ago.

[1215] *See* National Conference of State Legislatures, Legislative Social Media Policies and Resources, https://www.ncsl.org/research/about-state-legislatures/policies-related-to-legislative-use-social-media.aspx (last visited Dec. 30, 2020).

[1216] Alaska Legislature, *Social Media Guidelines*, https://www.ncsl.org/documents/nalit/AKSocialMedia.pdf, (last visited Dec. 30, 2020).

[1217] California Senate, *Social Media Comment Policy*, https://www.senate.ca.gov/sites/senate.ca.gov/files/2017_senate_social_media_comment_policy_website.pdf, (last visited Dec. 30, 2020).

media Policy prohibits comments which are not part of "*a discussion of legitimate concerns related to legislation and state government*," and also prohibits posting campaign information.[1218]

New German parliamentary speaker Wolfgang Schäuble sent a memo in November 2017 to the seven hundred and nine Bundestag members to let them know that tweeting inside the Assembly about the session was inappropriate.[1219] However, social media used inside an assembly can sometimes be used to promote democracy. On June 22, 2016, and June 23, 2016, Democrats members of the U.S. Congress held a sit-in the House of Representatives asking House Speaker Paul Ryan to bring two gun-control bills to the floor for voting. The sit-in started ten days after the horrific mass-shooting at the Pulse nightclub in Orlando, Florida, which had made forty-nine victims. As soon as the sit-in started in the morning of June 22, the speaker of the House called for a recess, which consequently turned off the official cameras for the House, which film only when the House is in session. Representatives started using Periscope. Twitter's live streaming app, and Facebook Live to film the sit-in. C-SPAN started carrying these feeds as it could no longer carry the official House of Representatives feed, and journalists are preventing from filming themselves, as House Rules for electronic media coverage forbids shooting live or recorded video images

[1218] Utah House of Representatives, *Social Media Policy*, https://house.utah.gov/social-media-policies/(last visited Dec. 30, 2020).
[1219] David Martin, *Wolfgang Schäuble's Bundestag Twitter ban met with backlash by parliamentarians*, DW.COM (Nov. 23, 2017), https://www.dw.com/en/wolfgang-sch%C3%A4ubles-bundestag-twitter-ban-met-with-backlash-by-parliamentarians/a-41508022.

in the House and Senate chambers.[1220] C-SPAN's communications director Howard Mortman explained in an interview why C-SPAN made this decision:

> "*It's a misconception that C-SPAN controls the cameras. It becomes a busier day for us, but it's also a heck of a great teachable moment and educational opportunity for people to know where the video comes from, and that the government controls the video."*[1221]

Indeed, social media may greatly enhance the reach of members of Congress. A Congressional Research Service report on *[t]he Impact of Electronic Media on Member Communications*, published in May 2016, noted that some argue that "*the rise of mass media, particularly television, has given the President a comparative advantage over Congress, [but that] Members of Congress lack the institutional resources to compete with the President, and Congress as a whole lacks a unity of message*," adding that "[*t]he rise of electronic communications has arguably allowed Congress, as a sum of its Members, to have a more influential voice in public political debates*.[1222]

V. The Private Users

The Congressional Research Service report was written before Donald Trump became President and kept using his private Twitter account as a Presidential platform, commenting on policies, announcing policies, even firing members of his cabinet through this medium. Did this make the President's private account a public forum? We will explore

---

[1220] *Rules for Electronic Media Coverage*, UNITED STATES HOUSE OF REPRESENTATIVES RADIO-TELEVISION CORRESPONDENTS'GALLERY, https://radiotv.house.gov/for-gallery-members/rules-for-electronic-media-coverage-of-congress (last visited Dec. 30, 2020).
[1221] Benjamin Freed, *Here's How C-SPAN Put House Democrats' Sit-in on Television*, WASHINGTONIAN (June 22, 2016), https://www.washingtonian.com/2016/06/22/heres-c-span-put-house-democrats-sit-television/.
[1222] Social Media in Congress: The Impact of Electronic Media on Member Communications, at 18, (May 26, 2016).

the issue (D), after having first seen how private rules may control social media speech (A), how social media shaming may be a way to control speech (B), a prevalent practice which led many to choose not to speak entirely, or at least, without having weighted the appropriateness of a post (C).

A. Using Private Law to Control Online Speech

We saw that the social media platforms, which are all private companies, are using their private rules to control and limit their users 'speech. However, even private individuals may control speech on social media, either by through their own professional Terms of Use (a) or by contract or copyright laws. (b)

a. Controlling Speech Through Terms of Use

In France, a court ruled in an emergency procedure (*référé*) in June 2014 that a blogger who had published a harsh critic of a restaurant on her blog had to pay damages to its owner for tortious disparagement.[1223] The same year, a New York state hotel, the Union Street Guest House, fined $500 wedding couples who had posted negative reviews about its services online, or couples whose guests or attendants who had posted online negative review.[1224] The hotel took directly the $500 from the deposit, each time a negative review was posted. The couple could, however, recoup their money if they deleted the negative review, or were able to convince their guests or wedding attendants to remove their

[1223] Mathilde Ceilles, *Condamnée en justice pour avoir critiqué un restaurant sur Internet*, LE FIGARO, (July 11, 2014, 19:01, updated July 15, 2014, 14:19), https://www.lefigaro.fr/actualite-france/2014/07/11/01016-20140711ARTFIG00327-condamnee-en-justice-pour-avoir-critique-un-restaurant-sur-internet.php.
[1224] Mara Siegel, *Hotel Fines $500 For Every Bad Review Posted Online*, PAGE SIX NEWYORKPOST.COM (Aug. 4, 2014, 1:03 am), http://pagesix.com/2014/08/04/hotel-charges-500-for-every-bad-review-posted-online/.

comments. Professor Eric Goldman, reporting about it on his blog,[1225] quoted the 2003 New York Supreme Court[1226] *People v. Network Assoc., Inc*., case,[1227] where a software company's terms and conditions, printed on diskettes and online on its "download" page, stated that installing the software constituted acceptance of the terms and conditions of its license agreement. Users were advised to "*read the license agreement before installation*" and added that:

> *"[o]ther rules and regulations of installing the software are: … b. The customer shall not disclose the result of any benchmark test to any third party without Network Associates' prior written approval [and] c. The customer will not publish reviews of this product without prior consent from [the software company.*"

In that case, the New York Attorney General (NYAG) filed suit against the company, deceptive acts and practices in violation of New York General Business Law § 349 and sought a permanent injunction based on fraud and illegality. The NYAG argued that including a reference to "rules and regulations" in the restrictive clause was deceptive. The NYAG also argued that it was:

> "*designed to mislead consumers by leading them to believe that some rules and regulations outside exist under state or federal law prohibiting consumers from publishing reviews and the results of benchmark tests"* and that "*the language [was] deceptive because it may mislead consumers to believe that such clause is enforceable under the lease agreement, when in fact it [was] not enforceable under the terms of*

[1225] Eric Goldman, *Fining Customers For Negative Online Reviews Isn't New…Or Smart*, TECHNOLOGY & MARKETING LAW BLOG (Aug. 12,2014), http://blog.ericgoldman.org/archives/2014/08/fining-customers-for-negative-online-reviews-isnt-new-or-smart-forbes-cross-post.htm

[1226] The New York Supreme Court is a court of first instance.

[1227] *People v. Network Assoc., Inc*., 195 Misc.2d 384 (2003).

the lease" and that "*as a result consumers may be deceived into abandoning their right to publish reviews and results of benchmark tests*."

The court held that the language of the agreement implied that it was not the software company's policies which placed a limitation on of reviews, but rather the law or other rules and regulations. Therefore, the court found that, as the language of the agreement may have been deceptive, it was unenforceable and warranted an injunction and the civil sanctions.[1228] The software company was also enjoined "*from including any language restricting the right to publish the results of testing and review without notifying the [NYAG] at least 30 days prior to such inclusion*."[1229] It can be argued that the case could apply if a private party regulates social media speech the same way the New York State hotel had chosen to do.

In another case, when the merchandise they had ordered online, for less than $20, shipping included, failed to be delivered, a Utah couple published on RipoffReport.com negative comments about the retailer, after having tried, in vain, to contact him. They were contacted by the merchant, more than three years later, not in an attempt to finally solve the delivery issue, but to inform them that their online review had violated a "non-disparagement clause" of the site's Terms of Sale and Use, which forbade clients to "*tak[e] any action that negatively impacts [the site], its reputation, products, services, management or employees*."[1230] The couple was also told that failure to remove their comment within seventy-two hours would result in a $3,500 fee. This clause, however, was not part of the

[1228] At 390.
[1229] At 391.
[1230] Complaint at 3, Palmer v. Kleargear.com, D. Utah 2013 (No. 1:13-cv-00175). The complaint is available at http://www.citizen.org/documents/Palmer-v-Kleargear-Complaint.pdf.

Terms of Sales and Use at the time the couple had posted their comments. As the couple did not pay the fee, the retailer reported this alleged debt to credit reporting agencies, which damaged the couple's credit.[1231] They filed suit against the retailer,[1232] claiming, *inter alia*, damages under the Fair Credit Reporting Act and state tort law and sought a declaratory judgment that their "debt" was null and void. The U.S. District Court for the District of Utah granted Plaintiffs a default judgment in May 2014 and found the retailer liable to the couple for violating the federal Fair Credit Reporting Act, defamation, intentional interference with prospective contractual relations, and intentional infliction of emotional distress.[1233] The couple was awarded $102,250 in compensatory damages and $204,500 in punitive damages.[1234]

It should be noted that online reviews are an effective way to inform fellow consumers, but that it can also be used by companies to disparage competitor's products, wither by hiring low-pay freelancers to write review or even by using bots.[1235] The Northern District of California held in 2011 that small business owners who alleged that the *Yelp* platform had created negative reviews of their businesses and manipulated review and ratings content to induce them to purchase advertising through Yelp, had failed to state

---

[1231] See Written Testimony of Jennifer Kulas Palmer before the U.S. Senate Committee on Commerce, Science & Transportation at the hearing "Zero Stars: How Gagging Honest Reviews Harms Consumers and the Economy" (Nov. 4, 2015), available at https://www.commerce.senate.gov/services/files/2966d0eb-8812-4035-ae59-4b75979864e4. It took 18 months for the couple to succeed having the bad report deleted from their credit report.

[1232] The non-profit organization Public Citizen represented them pro-bono.

[1233] Palmer v. Kleargear.com, D. Utah 2015 (No. 1:13-cv-00175).

[1234] *Judge Awards Utah Couple $306,750 in Case Against Retailer That Tried to Impose Fine for Critical Online Review*, PUBLIC CITIZEN, June 26, 2014, http://www.citizen.org/pressroom/pressroomredirect.cfm?ID=4234.

[1235] See Madeline Lamo & Ryan Calo, Regulating Bot Speech, 66 UCLA L. REV. 988,997 (2019), explaining that "[*b*]*ots can… skew the marketplace… by creating confusion in product review*" and that "*bots are an effective way to create large number of fake review in a short amount of time*."

a claim and dismissed the case. The Ninth Circuit affirmed in 2014.[1236] The business owners had filed a class-action lawsuit against Yelp, claiming civil extortion, and attempted civil extortion under California law. One busines owner claimed that several five-star reviews of his business had disappeared from the platform after he had refused to purchase advertising. The Ninth Circuit reasoned that, while the plaintiff claimed she had been deprived of the benefit of positive Yelp reviews, she:

> *"had no pre-existing right to have positive reviews appear on Yelp's website. She alleges no contractual right pursuant to which Yelp must publish positive reviews, nor does any law require Yelp to publish them. By withholding the benefit of these positive reviews, Yelp is withholding a benefit that Yelp makes possible and maintains. It has no obligation to do so, however. [Plaintiff] does not, and could not successfully, maintain that removal of positive user-generated reviews, by itself, violates anything other than Yelp's own purported practice."*[1237]

Another plaintiff claimed that Yelp had contacted the business offering to manipulate the business' listing page in exchange for purchasing advertising, and yet another claimed that the platform had published fake reviews about his business, "*as a threat to induce* [*him*] *to advertise*." The Ninth Circuit found that:

[1236] Levitt v. Yelp! Inc., 765 F. 3d 1123 (9th Circuit 2014).
[1237] Levitt v. Yelp! Inc., at 1133.

"Yelp's manipulation of user reviews, assuming it occurred, was not wrongful use of economic fear, and, second, that the business owners pled insufficient facts to make out a plausible claim that Yelp authored negative reviews of their businesses."[1238]

b. Controlling Speech Through Contract or Copyright

Other companies have tried to protect themselves against negative online reviews by using copyright laws.[1239] For instance, a Florida apartment complex in Florida asked prospective tenants to sign a "Social Media Addendum" assigning their copyrights in written or photographic comments about the property to the owner. By signing this Addendum, tenants also agreed, in consideration for the lease of their unit, not to publish negative commentaries about the apartments or the property.[1240]

A rather creative way for business or professionals to control critics made by their clients or patients is to have them sign an agreement assigning the copyright to any comments they may publish about the business or the services rendered. This allows them, as copyright holder to ask Online Service Providers (OSPs) to take down the comments using section 512 of the Digital Millennium Copyright Act of 1998 (DMCA) notice-and-takedown procedures. Copyright is thus used to chill legitimate speech, a far cry from its original purpose, under Article I, Section 8, Clause 8, of the U.S. Constitution, that is,

---

[1238] Levitt v. Yelp! Inc., at 1130.
[1239] Professor Eric Goldman called these clause "anti-review clauses," See Eric Goldman, *Understanding the Consumer Review Fairness act of 2016*, 24 MICH.TELECOM. & TECH. L. REV. 1 (2017). Professor Goldman noted that they are also called "gag clauses" or "non-disparagement clauses" and that Congress used both terms, citing H.R. Rep. No. 114-731, at 5 (2016).
[1240] Lisa Vaas, *Apartment complex threatens residents with $10k fines for negative online reviews*, NAKED SECURITY, March 11, 2015, https://nakedsecurity.sophos.com/2015/03/11/apartment-complex-threatens-residents-with-10k-fines-for-negative-online-reviews.

promote the progress of science and useful arts. Professor Eric Goldman stated that these clauses "*distort the marketplace benefits society gets from consumer reviews.*"[1241]

This route was however taken by a Manhattan dentist in 2010, using forms provided by a North Carolina company which provided such forms to health care professionals so that they can suppress negative comments about their practices.[1242] A patient visited her office in 2010 as he suffered to a violent toothache. But before curing the teeth, the dentist asked the patient to sign an agreement assigning her "*all Intellectual Property rights, including copyrights, … for any written, pictorial, and/or electronic commentary*." The patient later complained on *Yelp* and some similar sites about the high dental fee and the alleged lack of cooperation of the dentist's office in providing the right paperwork to the patient's insurance company. The dentist unsuccessfully asked the sites to take down the negative comments and then threatened the patient with a copyright infringement suit. Instead, he filed a class action suit, claiming that the agreement was void and, in the alternative, that his use of the copyrighted comments was fair use.[1243] Judge Paul Crotty rendered a default judgment in favor of Plaintiff on February 27, 2015.[1244] He found the use of the comments to be non-infringing fair use, that they "*constitute[d] breaches of fiduciary duty and violations of dental ethics and are subject to the equitable defenses of unclean hands, and, as*

[1241] Eric Goldman, *Understanding the Consumer Review Fairness act of 2016*, 24 MICH.TELECOM. & TECH. L. REV. 1 (2017).
[1242] This company is no longer providing these forms. The Center for Democracy and Technology (CDT) had filed a complaint with the FTC about these forms in November 2011, and the company ceased to provide them in December 2011., see https://cdt.org/press/medical-justice-terminates-illegal-doctor-review-contracts.
[1243] Complaint at 7, Lee v. Makhnevich, 1:11-cv-08665 (S.D.N.Y. 2011).
[1244] Lee v. Makhnevich, 1:11-cv-08665 (S.D.N.Y. 2015).

to such assignment and assertion, constitute copyright misuse," and ruled further that the agreement was void because it was unconscionable and lacked consideration.[1245]

Using such ill-advised methods to control speech about oneself is no longer legal, as the Consumer Review Fairness Act of 2016 (CRFA)[1246] was signed into law on December 14, 2016.[1247] The CRFA prohibits asking a party to a contact to fill a form contract containing a provision prohibiting or restricting the party:

> "*to engage in a covered communication*,"[1248] defined as "*a written, oral, or pictorial review, performance assessment of, or other similar analysis of, including by electronic means, the goods, services, or conduct of a person by an individual who is party to a form contract with respect to which such person is also a party.*"[1249]

It also prohibits imposing a penalty or fee to a party for having engaged in a "covered communication," or transferring or requiring the transfer of intellectual property rights to the content of the communication. Violation of the CRFA is an unfair or deceptive act violating the FTC Act.[1250] The CRFA does not, however, preempt state laws.[1251]

[1245] The agreement stated that, in consideration for the copyright assignment, the dentist agreed not to provide patient's personal information to any marketing companies, which the agreement claimed the dentist could do was because of because of a Health Insurance Portability and Accountability Act of 1996 (HIPAA) "loophole." However, HIPAA indeed forbids dentists, from disseminating patient information for marketing purpose, so thus there was no need for the patient to enter into an agreement to protect his privacy.
[1246] An Act to prohibit the use of certain clauses in form contracts that restrict the ability of a consumer to communicate regarding the goods or services offered in interstate commerce that were the subject of the contract, and for other purposes, 15 U.S.C. 45(b).
[1247] 15 U.S.C. § 45b.
[1248] 15 U.S.C. § 45c.
[1249] 15 U.S.C. § 45 a (2).
[1250] 15 U.S.C. § 45b (d)(1).
[1251] 15 U.S.C. § 45b (g).

The FTC announced in May 2019 that it had issued three separate proposed administrative complaints and orders enforcing the CRFA. The three separate companies had each used non-disparagement clauses in form contracts for goods and services.[1252] One of such clauses read that the company:

> "*takes customer service very seriously. We want all of our customers to be 100% satisfied. We also take our reputation very seriously. By signing this purchase order you are agreeing, under penalty of civil suit, for an amount not to exceed three times the monetary value of this order, plus attorney's fees for [company], not to publicly disparage or defame [company] in any way or through any medium*."

This clause is so broad as it could have for effect for a customer having posted a negative comment on s social media platform to receive a cease-and-desist letter from the company, claiming disparagement…

Publishing negative online reviews may still have negative consequences in other countries, but few have laws allowing the reviewer to be sent to jail. Such was, however, the case in Thailand, where an American was found guilty in October 2020 of defamation, a crime in Thailand, for having posted a derogatory review of a resort on the platform

[1252] *FTC Announces First Actions Exclusively Enforcing the Consumer Review Fairness Act*, FEDERAL TRADE COMMISSION, (May 8, 2019), https://www.ftc.gov/news-events/press-releases/2019/05/ftc-announces-first-actions-exclusively-enforcing-consumer-review. For instance, one company used this clause in their consumers contract: "*CUSTOMER and COMPANY agree that the within contract is a private and confidential matter and that the terms and conditions of the contract, including the estimates and all pricing shall remain private and confidential and shall not be made public, or given to anyone to make public, INCLUDING THE BETTER BUSINESS BUREAU. Customer also agrees not to file any complaints with the Better Business Bureau and agrees to attempt to resolve their complaints by contacting COMPANY in writing directly. Should the CUSTOMER breach this confidentiality clause, the CUSTOMER agrees to pay COMPANY liquidated damages equal to the actual amount of damages suffered or two times the contract price, whichever shall be higher. THE COMPANY MAY ALSO BE AWARDED COUNCIL [sic] FEES AND COSTS AS REQUESTED BY COMPANY*."

Tripadvisor.[1253] The parties settled as the American tourist accepted to issue an apology but had nevertheless to go to jail. The resort may not have many bookings from the U.S. in the next years or so, as *Tripadvisor* posted this message, once the American was out of jail and out of the country, on the resort page on its site:

> "*This hotel or individuals associated with this hotel filed criminal charges against a Tripadvisor user in relation to the traveler writing and posting online reviews. The reviewer spent time in jail as a result. Tripadvisor serves its users best when travelers are free to share their opinions and experiences on our platform – both positive and negative. The hotel may have been exercising its legal rights under local law, however, it is our role to inform you so you may take this into consideration when researching your travel plans*."[1254]

Interestingly, it seems that consumers consider that "*negative reviews establish credibility*"[1255] as they are viewed as ensuring the authenticity of the reviews, whereas all positive reviews are viewed as suspicious.

[1253] Richard C. Paddock, *He's Sorry for His Bad Reviews. He May Now Avoid Prison*, THE NEW YORK TIMES, (Oct. 9, 2020, Updated Nov. 11, 2020), https://www.nytimes.com/2020/10/09/world/asia/thailand-review-american-apology.html.

[1254] Richard C. Paddock, *Thai Hotel That Put American in Jail Gets New Label on Tripadvisor*, THE NEW YORK TIMES, (Nov. 11, 2020), https://www.nytimes.com/2020/11/11/world/asia/thailand-hotel-tripadvisor-jail.html?searchResultPosition=1.

[1255] See How Online Reviews Influence Sales. Companies cannot ignore the power of online consumer reviews, SPIEGEL RESEARCH CENTER, at 10, https://spiegel.medill.northwestern.edu/_pdf/Spiegel_Online%20Review_eBook_Jun2017_FINAL.pdf.

B. Public Shaming on Social Media

Publishing on social media innocuous and private instance which were meant to be private can lead to great embarrassment, even depression.[1256] Being shamed online may have dire real-life consequences. An American social worker was fired from her job after an online campaign asked for her being fired: her colleague had posted on *Facebook* a picture of her, with her consent, mock-yelling and really 'flipping the bird' at Arlington Cemetery, in front of a sign urging visitors "Silence and Respect."[1257] As the privacy settings of her colleague allowed the picture to be seen outside her networks of 'friends,' the image eventually became viral, and the threatening comments starting rolling. When Minnesota dentist Walter Palmer shot beloved Cecil the lion, while on a hunting trip in Zimbabwe, online shaming eventually led to people gathering in from of his dental office shouting "Murderer! Terrorist!" through a megaphone, forcing him to close it for a while.[1258] Online comments have wished for Dr. Palmer to "rot in hell." Similar comments about a hunter who regularly posted images on social media images of animals she had killed, such as foxes, fish, and deer, led to several criminal charges in Germany in 2020.[1259] Some of the comments called her a "*slut*" *while another read "ugly woman we'll find you, watch out for*

[1256] The plight of the "Star Wars Kid" , a Canadian teenager who videotaped himself plying with a golf ball retriever as it was a light saber, to have it later maliciously uploaded by third parties to mock him, would have been even worse if it had not occurred in 2002, but today. The teenager would have become an instant meme, a hashtag, and maybe even an emoji. Even so, he suffered great embarrassment for a few moments of playful abandon. He spoke ten years later about the issue and the dangers of cyberbullying. See *10 years later, 'Star Wars Kid' speaks out*, MCLEANS, (May 9, 2013), https://www.macleans.ca/news/canada/10-years-later-the-star-wars-kid-speaks-out.

[1257] Jon Ronson, *'Overnight, everything I loved was gone': the internet shaming of Lindsey Stone*, THE GUARDIAN (Feb. 21, 2015, 02:00 EST), http://www.theguardian.com/technology/2015/feb/21/internet-shaming-lindsey-stone-jon-ronson

[1258] Christina Capecchi and Katie Rogers, *Killer of Cecil the Lion Finds Out That He Is a Target Now, of Internet Vigilantism*, THE NEW YORK TIMES (July 29, 2015), http://www.nytimes.com/2015/07/30/us/cecil-the-lion-walter-palmer.html.

[1259] Ben Knight, *German online commenters ordered to pay hate speech fines*, DEUTSCHE WELLE, (Sept. 26, 2020), https://p.dw.com/p/3j3Gm, (last visited Dec. 30, 2020).

your health." Justine Sacco, a New York City public relations executive, was fired in December 2013 by her employer IAC after a tweet she sent just before boarding a plane to South Africa made her a number one trend on Twitter. The tweet read: "*Going to Africa. Hope I don't get AIDS. Just kidding. I'm white!*" Writer Jon Ronson compared the Sacco's fate to public shaming sentences which were carried on in 17th century New England.[1260] Indeed, "shaming punishments," injuring the dignity or the reputation of the convict were used in the U.S. from the colonial period to the 19th century.[1261]

Dan Markel wrote in 2001 that, in Colonial New England, "[*b*]*ecause Americans lived in smaller communities and were less likely than today to move between communities, they were more sensitive to threats to their reputations and good name.*"[1262] France used the pillory too in the Middle Age through the eighteenth century, either as a preliminary punishment before the main sentence is carried out, or, for lesser crimes, as main punishment.[1263] Another shaming punishment carried out in France was the run, often used to punish adulterers or thieves. The culprits had to run, sometimes naked in the case of adulterers, or carrying out the object they had stolen, in the case of thieves, while the town crier loudly informed the public about the passage of the shamed convict. It can be argued that members of social media networks are also living in small communities, not in terms of size, as social media sites have millions or even billions of users, but because the

---

[1260] Jon Ronson, *How One Stupid Tweet Blew Up Justine's Sacco's Life,* THE NEW YORK TIMES, Feb. 12, 2015, http://www.nytimes.com/2015/02/15/magazine/how-one-stupid-tweet-ruined-justine-saccos-life.html?ref=magazine

[1261] Dan Markel, *Are Shaming Punishments Beautifully Retributive: Retributivism and the Implications for the Alternative Sanctions Debate*, 54 VAND. L. REV. 2157, 2167, (2001)

[1262] Markel, at 2167-2168.

[1263] JEAN-MARIE CARBASSE, HISTOIRE DU DROIT PÉNAL ET DE LA JUSTICE CRIMINELLE, 296 (PUF), 2nd ed. 2009.

speech and the ease of sharing news may make all users aware of one single 'town crier.' This is particularly true in the case of Twitter, where most users provide unlimited access to their tweets and whose online conversations are carried on in the open.

The tweet which got Sacco into trouble raised furor because it was interpreted by most, including her employer, as racist. Ronson wrote that he believed, "*after thinking about her tweet for a few seconds more… that it wasn't racist, but a reflexive critique of white privilege,"* an interpretation confirmed by Sacco during an interview with Ronson. Other journalists, albeit a minority, expressed the same opinion.[1264] In 2020, it was Amy Cooper, a New York City executive who was fired by her employer after a video of her making a 911 call to report a bird watcher who had asked her to leash her dog in the Ramble, an area of Central Park home of more than two hundred and thirty bird species,[1265] and where dogs must be kept on leash.[1266] The man, Christian Cooper, filmed the Memorial Day encounter on his phone and posted it on his Facebook page, where it became viral.[1267] Amy Cooper is seen telling Mr. Cooper "*I'm going to tell them there's an African American man threatening my life*," then telling the police on the phone "*There is an African American man, he is recording me and threatening myself and my dog*," while holding her dog by the collar, as it whimpers. By the end of the day, the woman had been identified, and was placed on

---

[1264] Jeff Bercovici, *Justine Sacco And The Self-Inflicted Perils Of Twitter*, FORBES, Dec. 23, 2013, 8:43AM, http://www.forbes.com/sites/jeffbercovici/2013/12/23/justine-sacco-and-the-self-inflicted-perils-of-twitter

[1265] *The Ramble*, CENTRAL PARK NYC, https://www.centralparknyc.org/attractions/the-ramble, (last visited Dec. 30, 2020)

[1266] Sarah Maslin Nir, *White Woman Is Fired After Calling Police on Black Man in Central Park*, THE NEW YORK TIMES, (May 26, 2020, Updated May 27, 2020, 7:31 a.m. ET), https://www.nytimes.com/2020/05/26/nyregion/amy-cooper-dog-central-park.html?searchResultPosition=3.

[1267] Christian Cooper (May 25,1:34 PM), https://www.facebook.com/671885228/posts/10158742137255229/?d=n.

administrative leave by her employer, who announced it on Twitter.[1268] The dog had been identified as well (Henry), and its owner had "*voluntarily surrendered*" it to the rescue where it had been adopted "*while this matter [was] being addressed*," as announced by the rescue on its Facebook page.[1269] Amy Cooper was fired the next day, and the announcement was made by her employer on Twitter, writing that the company "*do*[*es*] *not tolerate racism of any kind*."[1270] The hashtag #AmyCooper was trending in first position on Twitter on both days.[1271] Two journalists on both side of the political spectrum, writing on the issue, both wrote that firing may have not been warranted.[1272]

Social media may be used denounce crimes, whether such allegations are grounded or not. An Australian woman saw a man taking a photograph at a Melbourne shopping mall and believed that he was a pedophile taking pictures of children nearby.[1273] She took a

---

[1268] Franklin Templeton, @FTI_US, Twitter, (May 25,2020 10 :43 PM), https://twitter.com/FTI_US/status/1265111264335986689.
[1269] Abandoned Angels Cocker Spaniel Rescue, Inc., Facebook (May 25, 6:28 PM), https://www.facebook.com/AbandonedAngels/posts/10157503306378723?__tn__=-R. The dog was later returned to its owner.
[1270] Franklin Templeton, @FTI_US, Twitter, (MAY 26,2020 2 :24 PM), https://twitter.com/FTI_US/status/1265348185201008641.
[1271] Robin Abcarian, *We've seen a lot of toxic white privilege lately, but Amy Cooper's tops it all*, THE LOS ANGELES TIMES, (May 26, 2020, 1:55 PM), https://www.latimes.com/opinion/story/2020-05-26/column-race-central-park-birder-dog-walker.
[1272] Emily Jashinsky, *Amy Cooper Doesn't Deserve Sympathy, But Social Media Shaming Is An Unhealthy Norm*, THE FEDERALIST, (May 27, 2020), https://thefederalist.com/2020/05/27/amy-cooper-doesnt-deserve-sympathy-but-social-media-shaming-is-an-unhealthy-norm/, calling Amy Cooper "*a thoroughly unsympathetic character* ", but expressing unease at how easy it has become to shame someone on social media. *See* also Joan Walsh, *Birding While Black: Just the Latest Bad Reason for White People to Call Police*, THE NATION, (May 27, 2020, 3:55 PM), https://www.thenation.com/article/society/amy-cooper-birding-police/. writing "*I don't know if [firing her] was the right outcome; I do know Amy Cooper should be charged with making a false police report*", adding that "[*a*]*ny potential empathy for Amy Cooper has to end with her calling the police on a black man who asked her to obey the law, rather than simply leashing her dog.*
[1273] Sarah Michael, *'OK people, take a look at this creep!': Man who mum shamed on Facebook because she thought he was taking photos of her kids... was just taking a selfie in front of a Darth Vader display to show HIS children*, DAILY MAIL (May 7, 2015, 22:07 EST), http://www.dailymail.co.uk/news/article-3073095/Mother-mistakenly-shames-dad-thought-taking-photos-kids-Facebook-post-shared-hundreds-actually-taking-selfie-Star-Wars-display-children.html#ixzz3bdFuWVRv

picture of the 'offender' and posted it on her Facebook page, relating the incident.[1274] The man was actually taking a 'selfie' in front of a Dark Vader poster to email it to his own children, as a joke. The post was shared more than 20 000 times, and a friend of the man alerted him that he had been 'exposed' as a sexual pervert on Facebook.[1275] He went to the police station, the police searched his phone, and he was cleared. That is the power of denouncing a crime: the police will investigate the issue, and one may become a suspect, if only for a fleeting moment. In this case, the man chose not to file a defamation suit, even though the Facebook post was defamatory ("Take a look at this creep") and allowed the man to be identified by his picture.

If an individual is arrested, another legal issue may rise, at least in the U.S. Booking photographs, colloquially known as "mugshots" public records, and as, such, have been seen by some commercial websites as a business opportunity. The "mugshots" are published online for the public to see who has been recently booked by the police. These practices, while provide the public information about police activity in their areas, comes also at a cost for the privacy of the individuals arrested.[1276] In some instances, these websites charged a fee for removing the picture.[1277] An Ohio attorney filed a suit against one of these sites in 2015, arguing that such practices may be a violation of the individual's right of publicity rights, distinguishing booking photographs, publicly available document,

---

[1274] She wrote: "*Ok people, take a look at this creep. Today at Knox, he approached my children when they were sitting at the frozen movie in the children's clothing section, he said "hey kids" they looked up and he took a photo, then he said I'm sending this to a 16 yr old."*

[1275] Lisa Vaas, *"Creep" shamed on Facebook was actually man taking selfie with Darth Vader*, NAKED SECURITY (May 11, 2015), https://nakedsecurity.sophos.com/2015/05/11/creep-shamed-on-facebook-was-actually-man-taking-selfie-with-darth-vader/.

[1276] Stephanie Francis Ward, *Hoist Your Mug*, ABA JOURNAL (Aug. 2012), 17-18. The article quotes Professor Annemarie Bridy: "*They're taking advantage of the public information and using it as a kind of shaming*."

[1277] Ibid. One of these sites had a link directly above the booking photograph, which read "*Click here for instant mug short removal"* and led to another site claiming it would delete the photograph for $99.

from the practice of selling removal services.[1278] In other instances, the police departments themselves published them on their social media accounts. The New York Times reported in June 2015 on the plight of a woman who had been arrested and booked by the South Burlington, Vermont, police department for rolling a stop while driving with an expired license.[1279] Her booking picture had been published on the police department Facebook page, to her great embarrassment. She was quoted saying: "*I actually felt like I was a murderer or selling drugs or something. Are they doing this because they think if they put your picture up you won't do it again?*" The South Burlington police department announced on July 4, 2015, that it had decided no longer to post mugshots on its Facebook page and deleted the pictures previously posted. Chief of Police Trevor Whipple wrote:

> "*After weighing the public transparency versus the posting of pre-adjudication images of those arrested it was felt that not posting is the best course of action…. Many in our community have seen value in the posting of mugshots and appreciate seeing who is being arrested. Others have presented arguments and positions as to why posting mugshots can be particularly harmful to someone who may be successful in a restorative justice process, has charges not filed, has charged dropped or is found to be not guilty of the accused*

[1278] David Kravets, *Shamed by Mugshot Sites, Arrestees Try Novel Lawsuit*, WIRED (Dec. 12.2012 06:30 AM), https://www.wired.com/2012/12/mugshot-industry-legal-attack. See also Debra Cassens Weiss, *Ohio Lawyer Sues Mugshot Websites, Claims Right-of-Publicity Violation*, THE ABA JOURNAL, (December 7, 2012, 12:00 pm CST), https://www.abajournal.com/news/article/ohio_lawyer_sues_mugshot_websites_claims_right-of-publicity_violation.

[1279] Jess Bidgood, *After Arrests, Quandary for Police on Posting Booking Photos*, THE NEW YORK TIMES (June 26, 2015), http://www.nytimes.com/2015/06/27/us/after-arrests-quandary-for-police-on-posting-booking-photos.html?_r=0.

crime. The posting of mugshots also brought about a flurry of inappropriate comments."[1280]

The New York Legislature passed a law, as part of the 2020 executive budget, which prohibits "*disclosure of law enforcement arrest or booking photographs of an individual, unless public release of such photographs will serve a specific law enforcement purpose and disclosure is not precluded by any state or federal laws*."[1281]

C. Choosing Not to Post and Choosing to Delete

Snapchat, now *Snap,* which original name was *Picaboo,* started as an application to be used for 'sexting.'[1282] As *Snap* messages disappear after ten seconds or less, depending on the time frame set by the sender, the sender of a racy message or picture may feel safe into believing that the message will not be forwarder or shared without his permission or knowledge. Indeed, Snap has a feature informing the sender the recipient is taking a screenshot of the picture. But, as noted by journalist Graham Cluley, "*what action* [*is the sender*] going to take if [*s/he*] *share*[*s*] *a photo in confidence, only to discover that someone has chosen to keep a permanent record?*"[1283] Also, a person other than the sender can take a picture of the telephone screen, unbeknownst to Snap and thus unbeknownst to the sender.

[1280] South Burlington Police Department, Facebook, (July 4, 2015), https://www.facebook.com/SouthBurlingtonPolice/posts/834326616651641.

[1281] N.Y. Pub. Off. Law § 89(2)(b)(viii) General provisions relating to access to records; certain cases.

[1282] While Snap did not advertise the new app as such, when it launched in 2012, its choice of the pictures used to illustrate the app were telling: two young women appearing to be naked in a pool as their body was hidden from the shoulders down by the screenshot of the app and the timer, setting how long the picture could be seen. The two models, two sisters, later sued Snapchat of their right of publicity. They also claimed these Snapchat images come first on Google Image search results when typing "Snapchat Sluts" or "Snapchat Whores." Alyson Shontell, *These 2 Sisters Became 'The Faces Of Snapchat' — Now They're Suing Because They Didn't Get Paid*, BUSINESS INSIDER, (Sept. 24, 2014, 6:05 PM), http://www.businessinsider.com/sarah-and-elizabeth-turner-sisters-sue-snapchat-for-slut-modeling-photos-2014-9.

[1283] Graham Cluley, *Does Snapchat offer safe sexting from smartphones, or a false sense of security?*, NAKED SECURITY, (Nov. 6, 2012), https://nakedsecurity.sophos.com/2012/11/06/snapchat-sexting-app-security/.

Instagram's Stories disappear twenty-four hours after having been posted, unless the user added them to her or his "Story Hightlights."[1284] Twitter announced in November 2020 that it would offer "Fleets"[1285] to allow users to share their "fleeting thoughts." The posts disappear 24 hours after being posted. WhatsApp introduced disappearing messages in November 2020. When the feature is turned on by the user, chart messages disappear withing seven days.[1286]

a. Thinking Before Posting

As any regular user of a social platform can attest, nice is the exception, not the rules. However, a Pew Research Center analysis found that lawmakers received more 'love' than 'anger' reactions to their Facebook posts in 2019 and 2020, while they had received more negative reactions than positive ones during the 2016 Presidential campaign.[1287] Evon Musk vowed to be nicer on Twitter in a 2018 interview.[1288] Scholar Timothy Garton Ash proposed in 2016 ten principles of free speech, among them not making threats of

---

[1284] *How do I add a story to my Story Highlights?*, INSTAGRAM HELP CENTER, https://help.instagram.com/813938898787367 (last visited Dec. 30, 2020).
[1285] Joshua Harris and Sam Haveson, *Fleets: a new way to join the conversation*, TWITTER BLOG, (Nov. 17, 2020), https://blog.twitter.com/en_us/topics/product/2020/introducing-fleets-new-way-to-join-the-conversation.html.
[1286] *Introducing disappearing messages on WhatsApp*, WHATSAPP BLOG, (Nov. 5, 2020), https://blog.whatsapp.com/introducing-disappearing-messages-on-whatsapp. In another post, the company advised its users to "[o]*nly use disappearing messages with trusted individuals. For example, it's possible for someone to: Forward or take a screenshot of a disappearing message and save it before it disappears. Copy and save content from the disappearing message before it disappears. Take a photo of a disappearing message with a camera or other device before it disappears*." *About disappearing messages*, WHATSAPP, https://faq.whatsapp.com/general/chats/about-disappearing-messages.
[1287] Regina Widjaya, *'Love' reaction steadily overcomes 'anger' as response to lawmakers' posts on Facebook*, (Sep. 11, 2020), THE PEW RESEARCH CENTER, https://www.pewresearch.org/fact-tank/2020/09/11/love-reaction-steadily-overcomes-anger-as-response-to-lawmakers-posts-on-facebook/.
[1288] "*I have made the mistaken assumption—and I will attempt to be better at this—of thinking that because somebody is on Twitter and is attacking me that it is open season."* See Tom Randall, '*The Last Bet-the-Company Situation': Q&A With Elon Musk*, BLOOMBERG, (July 13, 2018, 6:00 AM EDT), https://www.bloomberg.com/news/features/2018-07-13/-the-last-bet-the-company-situation-q-amp-a-with-elon-musk.

violence (number 2), expressing oneself "*with robust civility about all kind of human difference*" (number 5) and "*respect*[*ing*] *the believer but not necessarily the content of the belief* (number 6).[1289]

Generally, "thinking before posting" is a commonsense advice still relevant in our times. Technology may help us doing so: a thirteen-year-old teenager, Trisha Prabhu, presented in 2014 her "Re-think" project at the Google Science Fair.[1290] The system is designed to make adolescents "re-think" before posting a mean or hurtful message online. Ms. Prabhu's theory was that the prefrontal cortex, the frontal part of the brain which controls decision-making and impulse skills, does not develop fully until one reaches twenty-five years of age. Therefore, people below that age may have difficulty to control their emotions and also may be more likely to act on impulse, including posting messages which may be hurtful. Ms. Prabhu wanted to design a system which would prevent cyber-bullying, by giving the internet user about to post a bullying message the opportunity to think about it before hitting the "send" button. Therefore, the message is never posted, instead of being posted, hurting someone. Even if the bully may be blocked or banished from the site where the message was posted, such solutions are only temporary, notes Ms. Prabhu, and thus it is best to completely prevent cyber-bullying from occurring. Ms. Prabhu created a software program which would analyze messages about to be posted and ask:

[1289] Tom Rachman, *Timothy Garton Ash Puts Forth a Free-Speech Manifesto*, THE NEW YORK TIMES, (May 22, 2016), https://www.nytimes.com/2016/05/23/books/timothy-garton-ash-puts-forth-a-free-speech-manifesto.html, reporting on the publication of Tom Rachman's book, *Free Speech: Ten Principles for a Connected World*. The book is more than an exhortation to be nice online; the author argues that speech should not be censored.

[1290] Micah Singleton, *This 14-year-old wants cyberbullies to think before they post*, DAILY DOT, (Aug. 9, 2014, 2:54 pm), https://www.dailydot.com/debug/science-project-to-end-cyberbullying.

"*Alert message "This message may be hurtful to others. Would you like to pause, review and rethink before posting?*

The Federal Bureau of Investigation of Chicago produced a public service announcement video urging social media users to think before they post and to warn about the danger of posting on social media without considering the possible effects and consequences. The video features a young man explaining that he "*used social media to vent*," and that one of his posts was taken as a terrorist threat, "*the university got shut down; I got arrested by the FBI; and now, I don't know what my future looks like. I search my name on the web almost every day and look at this stuff. It's not going away*."[1291] He urged viewers to think before they post.[1292] A 2020 study by researchers from the University of Regina and the Massachusetts Institute of Technology found that social media users are more likely to consider whether a particular information about COVID-19, such as a headline. The study suggests that "*accuracy nudges are straightforward for social media platforms to implement on top of the other approaches they are currently employing*," and argued that *"with further optimization, interventions focused on increasing the salience of accuracy on social media could have a positive impact on countering the tide of misinformation*."[1293] During the COVID-19 pandemic, the United Nations launched the "Pause" initiative "*to foster behavior change and counter the spread of misinformation regarding COVID-19*

[1291] There is no right to be forgotten in the U.S.
[1292] *Think Before You Post PSA*, FEDERAL BUREAU OF INVESTIGATION, https://www.fbi.gov/video-repository/think-before-you-post-psa.mp4/view (last visited Dec. 30, 2020).
[1293] Jonathon McPhetres, Yunhao Zhang, Jackson G. Lu, David G. Rand, *Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention*, PSYCHOLOGICAL SCIENCE, (June 30, 2020), available at https://journals.sagepub.com/doi/10.1177/0956797620939054.

pandemic,[1294] asking social media users to take a #PledgetoPause but either making a short selfie video to demonstrate a 'pause, ' pausing quietly for a few , then sharing the video and pledging to pause before sharing information online during the Covid-19 pandemic.[1295]

New parents often publish photographs of their children on their social media accounts. Some parents even create social media accounts for their child. For instance, fashion model Coco Rocha created an *Instagram* account for her daughter, immediately after her birth in March 2015, which attracted 17,000 followers within two days.[1296] The French military police, the *gendarmerie nationale*, posted a warning on its Facebook page in February 2016, urging French parents to think twice before posting pictures of their children on social media. This post was triggered by the *Motherhood Challenge* social media campaign, which urged parents to post three pictures of their children on social media and nominate ten other parents to do the same.[1297] The *gendarmerie* reminded proud parents « *that it is important to protect the private life and the image of minor children on social media sites*." On the other side of the Rhine, the North Rhine-Westphalia (NRW) police, in Germany, urged parents in October 2015 to think twice before posting images of their children on social media sites. The NRW police explained, via its own Facebook page, that, while these pictures may now look "cute," they may later embarrass their children, who have a right to privacy, and that these images, especially if children are photographed naked, may even lead to the child being bullied, or even worse, for the photograph to be

---

[1294] *Pause. Take care before you share*, UNESCO (July 7, 2020), https://en.unesco.org/news/pause-take-care-you-share.
[1295] TAKE CARE BEFORE YOU SHARE, https://www.takecarebeforeyoushare.org/ (last visited Dec. 30, 2020).
[1296] Sam Reed, *Coco Rocha Gives Birth, Creates Instagram Account for Baby*, PRET-A-REPORTER (March 30, 2015), http://www.hollywoodreporter.com/news/coco-rocha-gives-birth-creates-785301.
[1297] Gendarmerie nationale, *[PRÉVENTION] Préservez vos enfants !*, Facebook (Feb. 23, 2016), https://www.facebook.com/gendarmerienationale/posts/1046288785435316 (last visited Dec. 30, 2020).

used by pedophiles.[1298] Considering that many pictures contain locational metadata which allow any interested third party to geo-locate where the picture was take, the threat for the security of children is real.[1299] Images of babies and young children have also been used to role-play on Instagram, using the #BABYRP hashtag. Some of the players are laying the role of babies, others the role of parents, others the role of a virtual adopting agency and they interact on social media. 'Parents' may contact the 'adopting agency' as they want to 'adopt' a child, say, blonde with blue eyes, and the 'agency' produces a picture found on *Instagram*.[1300] The exchange can be merely playful, but some are more sexual.[1301] The real parents of the children whose pictures are thus being used are understandably being shaken. Some started a social media campaign to inform the public, using the hashtag #downwithbabyrp, others started an online petition to ask *Instagram* to "[*p*]*ut an end to the baby and child role play accounts* which did not gather enough votes.[1302]

[1298] Polizei NRW Hagen, *Hören Sie bitte auf, Fotos Ihrer Kinder für jedermann sichtbar bei Facebook und Co zu posten!*, Facebook, (Oct. 13, 2015), https://www.facebook.com/Polizei.NRW.HA/photos/a.215738981931747.1073741830.208563659315946/474114729427503/?type=3&theater

[1299] Professor Owen Mundy from Florida State University built a site, I Know Where You Cat Lives, to show how the locational metadata of pictures of cates uploaded on social media allow any interested third party to precisely locate geographically where the picture was taken. Professor Mundy used pictures tagged #cats to make the point that pictures of human beings, including children, may also be geo-located using metadata. See I KNOW WHERE YOUR CAT LIVES, http://iknowwhereyourcatlives.com, (last visited Dec. 30, 2020).

[1300] Blake Miller, *The Creepiest New Corner Of Instagram: Role-Playing With Stolen Baby Photos,* FAST COMPANY, (Sept. 23, 2014), http://www.fastcompany.com/3036073/the-creepiest-new-corner-of-instagram-role-playing-with-stolen-baby-photos

[1301] Jeff Skrzypek, *South Florida mom says photos of her infant being used on Instagram in sexually explicit way*, WPTV, (Nov. 13, 2013, 10:24 PM), http://www.wptv.com/news/region-c-palm-beach-county/west-palm-beach/south-florida-mom-says-instagram-photos-of-her-infant-being-used-in-sexually-explicit-way#ixzz2l4KjjP1Qv.

[1302] Put an end to the baby and child role play accounts , CHANGE, ORG, https://www.change.org/p/instagram-put-an-end-to-the-baby-and-child-role-play-accounts?utm_campaign=petition_created&utm_medium=email&utm_source=guides, (last visited Dec. 30, 2020).

The answer to this issue may be technology. Some features, some as privacy settings, are already available to parents wishing to protect their children online privacy. But only a minority of parents seem to use them. A study commissioned in 2015 by a U.K. internet registry company found that, while 53% of parents had uploaded images of their children on Facebook, 14 % of them on Instagram and 12% on Twitter, 17% of the parents had never checked their Facebook privacy settings, and 46% had only checked once or twice.[1303] Jay Parikh, Facebook's Vice president of Engineering, had announced in November 2015 that the site would start to automatically warn parents about to post pictures of their children which could be seen by all. Mr. Parikh cited progress in artificial intelligence and deep learning which will further allow image analysis to warrant the necessity of warning parents about to allow images of their children to be analyzed.[1304]

b. Deleting One's Post

During the controversy over the *Charlie Hebdo* Pen award,[1305] Joyce Carol Oates, who had signed the letter protesting PEN America's freedom of expression award to the French satirical magazine,[1306] posted on Twitter: "*Twitter confers upon provisional thoughts*

[1303] *Today's children will feature in almost 1,000 online photos by the time they reach age five* , KNOWTHENET (May 26, 2015), http://www.knowthenet.org.uk/articles/today%E2%80%99s-children-will-feature-almost-1000-online-photos-time-they-reach-age-five#sthash.bM21eJju.dpuf.

[1304] Mark Blunden, *Facebook 'will automatically warn parents if they share pictures of their children with the public by accident'*, EVENING STANDARD, (Nov. 12, 2015), http://www.standard.co.uk/news/techandgadgets/facebook-will-automatically-warn-parents-if-they-share-pictures-of-their-children-with-the-public-by-a3112681.html

[1305] See Hillel Italy, *After weeklong controversy, Charlie Hebdo receives PEN award at literary gala in NYC*, PEN AMERICA, https://pen.org/press-clip/after-weeklong-controversy-charlie-hebdo-receives-pen-award-at-literary-gala-in-nyc.

[1306] Alan Yuhas, *Two dozen writers join Charlie Hebdo PEN award protest*, THE GUARDIAN (29 Apr 2015 14.46 EDT), https://www.theguardian.com/books/2015/apr/29/writers-join-protest-charlie-hebdo-pen-award.

a spurious quasi-permanence. In life we tend to discuss, revise opinions. But a tweet seems final."[1307]

Deleting one's message or choosing not to write about a subject to avoid possible misunderstanding are both forms of self-censorship. Even if it is legal to post about a particular topic, it may not be advisable. Some social media users may fear that posting about a controversial topic, say, abortion or women's rights, may lead to a cascade of insulting, even threatening messages and thus are choosing not to participate in a social media debate, even if the topic is of interest to them. Therefore, the First Amendment theory of the marketplace of ideas is not well served.

D. Not Everyone Has the Right to Delete One's Tweet: When a Social Media is a Public Forum

Speech is conveyed through channel of communications. The government may regulate the obnoxiousness of the channel, but not the obnoxiousness of the speech. This is why a city ordinance prohibiting use of sound amplification devices, except with the permission of the Chief of Police, was held to be unconstitutional, as its large scope had prevented a religious proselyte to amplify his message by using loudspeakers in a public park. Speech is also noise, but noise "*can be regulated by regulating decibels*, wrote Justice Douglas in *Saia v. New York*, adding that "[*a*]*nnoyance at ideas can be cloaked in annoyance at sound. The power of censorship inherent in this type of ordinance reveals its vice*."[1308] Speech can also be printed on pamphlets which, when discarded, become litter, which may

[1307] @JoyceCarolOates, Twitter, (May 4, 2015, 1:35 PM), https://twitter.com/JoyceCarolOates/status/595280727228354560.
[1308] Saia v. New York, 334 US 558, 562 (1948).

be regulated as litter, not as speech. As explained by Justice Roberts in *Schneider v. State*, the State may well wish to keep streets clean, but an ordinance prohibiting a person from "*handing literature to one willing to receive it*" is not the way to do so, as:

> "[*a*]*ny burden imposed upon the city authorities in cleaning and caring for the streets as an indirect consequence of such distribution results from the constitutional protection of the freedom of speech and press*."[1309] The State may still act to prevent street littering, such as laws making throwing papers on the streets illegal.

Justice Douglas wrote in 1948, in *Saia v. New York*:

> "*Loud-speakers are today indispensable instruments of effective public speech. The sound truck has become an accepted method of political campaigning. It is the way people are reached. Must a candidate for governor or the Congress depend on the whim or caprice of the Chief of Police in order to use his sound truck for campaigning? Must he prove to the satisfaction of that official that his noise will not be annoying to people?"*

Social media accounts have now replaced loud-speakers as *"indispensable instruments of effective public speech.*" Could social media platforms or accounts be one day considered public forums, if other government follow more and more the social media playbook set by the Forty-fifth President? Donald Trump appeared to argue that social media platforms are public forum in his Preventive Online Censorship Executive Order,[1310]

[1309] Schneider v. State (Town of Irvington), 308 US 147, 162 (1939).
[1310] 85 FR 34079, available at https://www.federalregister.gov/documents/2020/06/02/2020-12030/preventing-online-censorship.

citing *PruneYard Shopping Center v. Robins.*[1311] In that case, the Supreme Court found that a California shopping center, a private property, could not bar people from exercising their freedom of speech inside the shopping center, as this right was protected by both the California Constitution and the First Amendment. The shopping center had as policy not to allow visitors or tenants to engage in expressive activity not directly related to the mall commercial purposes, a policy which had been enforced in a nondiscriminatory way. Appellees were high school students who distributed pamphlets and asked people to sign their petition in support of a United Nations resolution against Zionism. The Supreme Court reasoned that the shopping center is a business establishment open to everyone, and that, therefore, the views expressed by members of the public distributing pamphlets or soliciting signatures on a petition "*will not likely be identified with those of the owner* [*of the shopping center*]." As the State did not dictate a specific message to be displayed on the private property, there was "*no danger of governmental discrimination for or against a particular message,"* and the owners could have "*expressly disavow any connection with the message by simply posting signs in the area where speakers or handbillers stand*." *PruneYard* may not have been the best case to be cited in the Executive Order, as Twitter, a private party owning a private, albeit digital, space, had not deleted the President's tweets, but had indeed "*expressly disavowed any connection with the message by posting* [*a sign*]." However, the right of Donald Trump to access social media sites had not been infringed by Twitter. As we saw earlier, Twitter had only added a link to the President's tweets about the mail-in ballots beings sent to every registered voter in California, warning users to *"! Get the facts*

[1311] PruneYard Shopping Center v. Robins, 447 U.S. 74, 85-89 (1980).

about mail-in ballots" and leading to a page offering counterviews and facts about the issue. Twitter explained in a tweet that it had done so:

> "*as part of our efforts to enforce our civic integrity policy. We believe those Tweets could confuse voters about what they need to do to receive a ballot and participate in the election process.*"[1312]

The Executive Order on Preventing Online Censorship also cited the 2017 Supreme Court *Packingham v. North Carolina* case,[1313] where the Court had found that a North Carolina law making a felony for registered sex offenders to access a social media sites, knowing that the site allows minors to become members, or to create or maintain personal Web pages, violated the First Amendment, as such sites *"can provide perhaps the most powerful mechanisms available to a private citizen to make his or her voice heard*." The Supreme Court had indeed noted that social media "*allows users to gain access to information and communicate with one another on any subject that might come to mind*," and that barring access to sex offenders would prevent them from "*speaking and listening in the modern public square.*"

The Supreme Court identified three types of public forums in *Perry Educ. Ass'n v. Perry Local Educators' Ass'n*:[1314] the traditional public forum, the public forum created by government designation, and the nonpublic forum. The Court added a fourth type of public forum to the list in 2010, the "*limited public forum*," which may be "*limited to use by certain*

[1312] @TwitterSafety, Twitter (May 27, 2020, 10:54 PM), https://twitter.com/TwitterSafety/status/1265838823663075341
[1313] Packingham v. North Carolina, 137 S. Ct. 1730, 1737 (2017).
[1314] Perry Ed. Assn. v. Perry Local Educators' Assn., 460 US 37, 45-47 (1983).

groups or dedicated solely to the discussion of certain subjects." For this type of forum "*access barrier must be reasonable and viewpoint neutral*." The government cannot prohibit communicative activities in traditional public forums, such as public streets and parks. Time, place, and manner of expression for this forum can be regulated but content-based exclusion from this type of forum to be legal, the State must show that the regulation is necessary to serve a compelling state interest and that it is narrowly drawn to achieve that end. The regulation must also leave open "*adequate alternative channels for communication."*[1315] The Supreme Court explained in *Perry Educ. Ass'n v. Perry Local Educators' Ass'n* that "[*a*] *public forum may be created for a limited purpose such as use by certain groups… (student groups), or for the discussion of certain subjects, … (school board business)*."[1316] It is not required however, to keep this forum indefinitely open, and it may be regulated the same way as a traditional public forum and time, place, and manner of expression there can be regulated, as long it such regulations are content-neutral, narrowly tailored, serve a significant government interest, and leave open ample alternative channels of communication. The government can also designate a place or a channel of communication to be a public forum, and the public at large can then use it for assembly and speech. However, carefully defining the limits of a particular forum may not be enough to create a legitimate designated public forum. For instance, a Memorandum from the Fairfax County in Virginia limited use of the large grassy mall in front of the Fairfax County Government Center Complex mall to "[*a*]*ny nonprofit organization which has an office in Fairfax County and/or serves the citizens of Fairfax County*...." The purpose of using this

[1315] Perry Ed. Assn, at 45.
[1316] Perry Ed. Assn, note 7.

grassy area was "*to encourage use of the common areas of the Government Center Complex by [qualified persons] for civic, cultural, educational, religious, recreational and similar activities...*" A woman who had been denied the right to put a religious display in the mall because she was not part of the class of people allowed to speak in this particular forum sued Fairfax County, claiming violation of her freedom to speak. While the District Court held the mall was a designated limited public forum, and that therefore the First Amendment had not been violated, the Fourth Circuit requalified it as a traditional public forum, as the mall had "*the objective use and purpose of a traditional public forum*."[1317]

Are social media platforms public forums? Justice Kennedy compared, in dicta, social media sites to the "*modern public square*" in *Packingham v. North Carolina.*[1318] The Supreme Court held in 2019 in *Manhattan Community Access Corp. v. Halleck* that "*a private entity who provides a forum for speech is not transformed by that fact alone into a state actor*."[1319] This case was cited in 2020 by the Ninth Circuit Court of Appeals in *Prager university v. Google , LLC*, holding that "[*d*]*espite YouTube's ubiquity and its role as a public-facing platform, it remains a private forum, not a public forum subject to judicial scrutiny under the First Amendment*."[1320]

Are the social media accounts of elected officials public forums? The First Amendment Coalition released documents in September 2017 showing that then-California

[1317] Warren v. Fairfax County, 196 F. 3d 186, 189-190 (4th Cir. 1999)
[1318] Packingham v. North Carolina, 137 S. Ct. 1730, 1737 (2017): "*By prohibiting sex offenders from using those websites, North Carolina with one broad stroke bars access to what for many are the principal sources for knowing current events, checking ads for employment, speaking and listening in the modern public square*."
[1319] Manhattan Community Access Corp. v. Halleck, 139 S. Ct. 1921, 1931 (2019).
[1320] Prager University v. Google, LLC , No.18-15712, (Feb. 26, 2020), https://cdn.ca9.uscourts.gov/datastore/opinions/2020/02/26/18-15712.pdf.

Governor Jerry Brown had blocked some 1,500 individual accounts on Facebook and several hundred on Twitter. Governor Brown no longer blocked these accounts when these documents were released but could not specify when he had blocked them. According to the First Amendment Coalition, Governor Brown had first refused to make public the blocked accounts list, as they had been blocked from his personal social media accounts and thus their disclosure was not required by the California Public Records Act. Can politicians block social media users?

President Trump is an active user of Twitter, to the point he has been called "Commander-In-Tweet."[1321] Even after being elected President of the United States in 2016, he continued to 'tweet' opinions, even policies, from his own Twitter account, @RealDonaldTrump, which he created in May 2009, not from the official Twitter account of any U.S. President, @POTUS. President Trump blocked several Twitter users from seeing his account, apparently after they had directed messages at @RealDonaldTrump, criticizing the President. The Knight First Amendment Institute at Columbia University and seven Twitter users who have been thus blocked filed a suit in July 2017 against Donald Trump, in his official capacity, Sean Spicer, at the time White House Press Secretary, and Daniel Scavino, then White House Director of Social Media and Assistant to the President, who sometimes operated the President's social media accounts. Plaintiffs claimed that, by blocking them, Defendants have violated the First Amendment rights. They took the position that the @RealDonaldTrump Twitter account is a public forum:

[1321] Tamara Keith, *Commander-In-Tweet: Trump's Social Media Use And Presidential Media Avoidance*, NPR, (Nov. 18, 2016 3:46 PM ET), https://www.npr.org/2016/11/18/502306687/commander-in-tweet-trumps-social-media-use-and-presidential-media-avoidance.

"[*b*]*ecause of the way the President and his aides use the @realDonaldTrump Twitter account*," explaining further that Defendants "*have promoted the President's Twitter account as a key channel for official communication. Defendants use the account to make formal announcements, defend the President's official actions, report on meetings with foreign leaders, and promote the administration's positions on health care, immigration, foreign affairs, and other matters*."[1322]

It should also be noted that the National Archives and Records Administration had informed the White House that the tweets published by @realDonaldTrump are official records and had thus to be preserved under the Presidential Records Act (PRA) of 1978,[1323] which established that the Presidential Records,[1324] as they are publicly owned and must thus be archived by the National Archives and Records Administration (NARA), which has been established by Congress to preserve and care for the records of the United States government.[1325] Presidential records include records in electronic forms and thus include social media postings. Personal records,[1326] however, are not publicly owned. The President of the United States has his own Twitter account since 2012, @POTUS. President

[1322] Plaintiff gave as one example the use of the @realDonaldTrump account to announce on June 7, 2017 that the President intended to nominate Christopher Wray as FBI director.
[1323] 44 U.S.C. §2201-2209.
[1324] They are defined by the PRA, 44 U.S.C. § 2201(2), as "*documentary materials, or any reasonably segregable portion thereof, created or received by the President, the President's immediate staff, or a unit or individual of the Executive Office of the President whose function is to advise and assist the President, in the course of conducting activities which relate to or have* an *effect upon the carrying out of the constitutional, statutory, or other official or ceremonial duties of the President.*"
[1325] The Vice-Presidential Records, which were created on or received after January 20, 1981 must also be preserved under the Act. Vice President Mike Pence has a personal Twitter account, @Mike_Pence.
[1326] They are defined by the PRA, as ""*documentary materials or any reasonably segregable portion thereof, of a purely private or nonpublic character, which do not relate to or have an effect upon the carrying out of the constitutional, statutory, or other official or ceremonial duties of the President"* which include *"diaries, journals, or other personal notes serving as the functional equivalent of a diary or journal which are not prepared or utilized for, or circulated or communicated in the course of, transacting Government business," "private political associations"* and *"materials relating exclusively to the President's own election to the office of the Presidency.*"

Obama was the first to be able to use @POTUS.[1327] At the end of each President's term,[1328] NARA takes legal custody of the Presidential records of the Administration, which are "*automatically transferred to the legal custody of the Archivist of the United States and the National Archives and Records Administration. The records are eventually housed in a Presidential Library maintained by NARA*."[1329] President Obama had kept his personal Twitter account, @BarackObama,[1330] while President, but did not use it to communicate about Presidential affairs. Donald Trump used his personal Twitter account, @realDonaldTrump, to communicate about all sort of topics, including official ones. This led to the question: were these tweets Presidential or personal records? Senator Claire McCaskill, at the time Ranking Member of the Homeland Security and Governmental Affairs Committee, and Senator Tom Carper, a member of the Committee, sent a letter on March 7, 2017 to David S. Ferriero, Archivist of the United States, asking him whether NARA considered "*President Trump's tweets as presidential records that need to be preserved for historic purposes*" and if NARA had made "*a determination of whether the Trump Administration must also preserve altered or deleted tweets*."[1331] Mr. Ferriero answered on March 30, 2017, that:

[1327] The account is always used by the current President. The tweets of a former President are archived as well.
[1328] Twitter announced in December 2020 that the @POTUS account would lose of its followers the day of Joe Biden's inauguration as President of the United States. The move was criticized by President Biden's team, as the follower's counts had not been erased when Donald Trump had become President, *see* Reese Oxner, *Twitter Will Reset @POTUS Account To Zero Followers After Biden Transition*, NPR, (Dec. 23, 2020, 2:38 PM ET), https://www.npr.org/sections/biden-transition-updates/2020/12/23/949632440/twitter-will-reset-potus-account-to-0-followers-after-biden-transition.
[1329] NARA, *Guidance on Presidential Records, from the National Archives and Records Administration*, https://www.archives.gov/files/presidential-records-guidance.pdf.
[1330] The account was created in March 2007.
[1331] The letter is available at https://www.archives.gov/files/press/press-releases/2017-03-07-mccaskill-carper-letter%20to-aotus%283%29.pdf.

"*NARA has advised the White House that it should capture and preserve all tweets that the President posts in the course of his official duties, including those that are subsequently deleted, as Presidential records, and NARA has been informed by White House officials that they are, in fact, doing so*" but that NARA, however, "*does not make "determinations" with respect to whether something is or is not a Presidential record*."[1332]

In *Knight First Amendment Inst. at Columbia Univ. v. Trump*, Defendants moved for summary judgment, arguing that "[*a*] *First Amendment claim may be directed only at state action, not the President's personal use of social media*." They also argued that Plaintiffs lacked standing, because the Knight Institute had not been blocked by Trump, and the users who had been blocked could still view Donald Trump's tweets, interact with other Twitter users about the President, and that thus their freedom of expression had not been abridged. They also argued that @realDonaldTrump is a personal account, that the choice of blocking some accounts is made by Trump alone, and "*courts are prohibited from enjoining the discretionary conduct of the President.*" One of Defendant's arguments was that blocking users was not state action, quoting *Carlos v. Santos*, where the Second Circuit found that there is no state action if any citizen can perform the act being challenged in court.[1333] Defendants had to reluctantly admit however that "*the President's account identifies his office, and his tweets make official statements about the policies of his administration*," but then argued that:

[1332] The answer is available at https://www.archives.gov/files/press/press-releases/aotus-to-sens-mccaskill-carper.pdf.
[1333] Carlos v. Santos, 123 F. 3d 61 (2nd Cir.1997).

"*the fact that the President may announce [actions of the State] through his Twitter account does not mean that all actions related to that account are attributable to the state. Public officials may make statements about public policy and even announce a new policy initiative in a variety of settings, such as on the campaign trail or in a meeting with leaders of a political party. The fact that an official chooses to make such an announcement in an unofficial setting does not retroactively convert into state action the decision about which members of the public to allow into the event.*"[1334]

Judge Naomi Buchwald from the Southern District of New York held on May 23, 2018 that a public official cannot block a person from his Twitter account "*in response to the political views that person has expressed*," even if the public official is the President of the United States, without violating the First Amendment.[1335] Judge Buchwald reasoned that "*portions of the @realDonaldTrump account -- the "interactive space" where Twitter users may directly engage with the content of the President's tweets -- are properly analyzed under the "public forum" doctrines set forth by the Supreme Court, that such space is a designated* public forum, and that the blocking of the plaintiffs based on their political speech constitutes viewpoint discrimination that violates the First Amendment.

Judge Buchwald first considered whether Plaintiffs had Article III standing to sue Defendants and considered whether plaintiffs (1) had suffered an injury in fact, (2) that is fairly traceable to the challenged conduct of the defendants, and (3) which is likely to be

[1334] Memorandum of Law in support of motion for summary judgment. A copy of this motion is available at https://www.scribd.com/document/361606197/Knight-v-Trump-WH-Opening-Brief#from_embed.
[1335] Knight First Amendment Inst. at Columbia Univ. v. Trump, 302 F. Supp. 3d 541 (S.D.N.Y. 2018).

redressed by a favorable judicial decision.[1336] Blocking users and thus limiting their access to the President's tweets were cognizable injuries-in-fact. Defendant Scavino had access to the @realDonaldTrump and the power to unblock the plaintiffs and thus any future injuries could be traceable to him, even though the record could not establish whether he had indeed blocked Plaintiffs.[1337] The record, however, "*definitively establishe*[*d*] *that the plaintiffs' injuries-in-fact [were] directly traceable to the President's*." As for the third part of the Article III standing test, redressability, Judge Buchwald considered that a declaratory or injunctive relief resulting in the unblocking of plaintiff's Twitter accounts would redress at least some of their future injury, even if the President would later block their account again. The Knight Institute had standing, as it had suffered an injury-in-fact for not being able to read comments that would have been posted by plaintiffs in reply of tweets posted by @realDonaldTrump. Such injury is a direct consequence of plaintiffs not being able to reply to the @realDonaldTrump tweets because they have been blocked, and the injury is redressable if the plaintiffs are unblocked and will then be able to reply to the President's tweets again.

Judge Buchwald then considered whether a public official's blocking individuals on Twitter implicates a forum for First Amendment purposes.[1338] Plaintiffs sought to engage in

---

[1336] Quoting Lujan v. Defenders of Wildlife, 504 US 555, 560 (1992). The Supreme Court explained that "…the irreducible constitutional minimum of standing contains three elements. First, the plaintiff must have *suffered an "injury in fact"—an invasion of a legally protected interest which is (a) concrete and particularized, … and (b) "actual or imminent, not `conjectural' or `hypothetical,' … Second, there must be a causal connection between the injury and the conduct complained of—the injury has to be "fairly. . . trace[able] to the challenged action of the defendant, and not . . . th[e] result [of] the independent action of some third party not before the court…. Third, it must be "likely," as opposed to merely "speculative," that the injury will be "redressed by a favorable decision*.".

[1337] Judge Buchwald did not find, however, that Sarah Huckabee Sanders, then White House Press Secretary, had access to the @realDonaldTrump account and thus granted summary judgment in her favor.

[1338] Citing Cornelius v. NAACP Legal Defense & Ed. Fund, Inc., 473 US 788, 797 (1985). The Supreme Court explained that to resolve the issue of whether a party's First Amendment right was violated by being excluded

political speech, and nothing in the record suggests that they sought to engage in unprotected speech, such as obscenity or fraud. Plaintiffs' speech was thus protected. Judge Buchwald then identified the nature of the forum, using the test set out by the Supreme Court in *Cornelius v. NAACP Legal Defense & Ed. Fund, Inc*. for defining the forum, which is to focus on the access sought by the speaker.[1339] Judge Buchwald noted that Plaintiffs did not seek to access the @realDonaldTrump account as a whole, and thus the whole account was not the forum to be analyzed. Instead, the forum doctrine had to be applied to different aspects of the @realDonaldTrump account, such as:

> "*the content of the tweets sent, the timeline comprised of those tweets, the comment threads initiated by each of those tweets, and the "interactive space" associated with each tweet in which other users may directly interact with the content of the tweets by, for example, replying to, retweeting, or liking the tweet*."

A public forum must be owned or controlled by the government. In this case, the forum was controlled by the government as even if Twitter is a private company because Defendants "*exercise*[*d*] *control over various aspects of the @realDonaldTrump account*," such as controlling the content of the tweets, being able to block other Twitter users, thus preventing them to access the @realDonaldTrump timeline and to participate "*in the interactive space associated with the tweets sent by the @realDonaldTrump account.*" Judge

---

from a particular forum, courts must first decide whether the speech is protected by the First Amendment. If it is protected, the court must then identify the nature of the forum, "*because the extent to which the Government may limit access depends on whether the forum is public or nonpubli*c." The courts must finally assess if the justification for exclusion from the forum satisfies the requisite standard.

[1339] Cornelius v. NAACP Legal Defense & Ed. Fund, Inc., at 801: "When speakers seek general access to public property, the forum encompasses that property. … In cases in *which limited access is sought, our cases have taken a more tailored approach to ascertaining the perimeters of a forum within the confines of the government property*."

Buchwald noted that Twitter maintained control over the @realDonaldTrump account, but that defendants had sufficient control over account to:

> "*establish the government-control element as to the content of the tweets sent by the @realDonaldTrump account, the timeline compiling those tweets, and the interactive space associated with each of those tweets*."

Also, "[*defendants*]*'control over the @realDonaldTrump account is … governmental,"* as it is presented as being the account of the Forty-fifth President of the United States, the tweets are preserved under the Presidential Record Act, and the account "*has been used in the course of the appointment of officers (including cabinet secretaries), the removal of officers, and the conduct of foreign policy*." As such, defendants exercised government control over the account, in their official capacities. Judge Buchwald compared the power to exclude to the power to preserve property, both powers reserved to the property owner, reasoning that "[*w*]*hen a government acts to "legally preserve the property under its control for the use to which it is dedicated," it behaves "like the private owner of property*."[1340] The @realDonaldTrump account was not a private account, as defendants argued, even though it had been created in 2009, when Donald Trump was a private citizen, as "*the entire concept of a designated public forum rests on the premise that the nature of a (previously closed) space has been changed*."[1341] However, the government has no control over the thread of comments posted under a tweet and thus are not a forum.

[1340] Quoting Rosenberger v. Rector and Visitors of Univ. of Va., 515 US 819, 829 (1995), quoting Lamb's Chapel v. Center Moriches Union Free School Dist., 508 U. S. 384, 390 (1993).
[1341] Quoting Cornelius, 473 U.S. at 802.

Judge Buchwald then examined whether use of the public forum thus established was consistent with the purpose, structure, and intended use of the three aspects of the @realDonaldTrump account found to be government-controlled, the content of tweets, the timeline comprised of the account's tweets, and the interactive space of each tweet. Are they government speech or private speech? The answer to this question is essential as government speech "*do*[*es*] *not normally trigger the First Amendment rules designed to protect the marketplace of ideas.*"[1342] Judge Buchwald considered the three factors used by the Supreme Court in *Walker v. Sons of Confederate Veterans* to determine is a particular speech is government speech: (1) whether government has historically used the speech to convey state messages, (2) whether the public closely identifies it with the government, and (3) the extent to which government maintains direct control over this speech. The content of the speech is government speech, as the record had established that the President, with the help of Scavino, used the @realDonaldTrump account to announce policies or official decisions, to promote his political agenda, even to engage with foreign leaders. As the timeline merely aggregates all the tweets, it also is government speech. However, the "*interactive space for replies and retweets created by each tweet sent by the @realDonaldTrump account*" is not government speech, as they are "*most directly associated with the replying user rather than the sender of the tweet being replied to*": it is the user replying who is in control, not the government.

Judge Buchwald asked an interesting question in footnote 20, "[*w*]*hether the content of retweets initially sent by other users constitutes government speech presents a somewhat*

[1342] Walker v. Sons of Confederate Veterans, 135 S. Ct. 2239, 2246 (2015).

closer question," noting that "[*t*]*he content* of a retweet of a tweet sent by another "*governmental account… is still squarely government speech. The content of the retweet of a tweet sent by a private non-governmental account… would still likely be government speech… despite the private genesis of the content, the act of retweeting by @realDonaldTrump resembles the government's acceptance of the monuments in Pleasant Grove and the government's approval of the license plate designs in Walker, which were sufficient to render the privately originated speech governmental in nature.*" This is an interesting argument, as if a retweet by @realDonaldTrump is government speech, then it can be argued that the U.S. government condoned the statement "*The only good Democrat is a dead Democrat*," offered by a New Mexico county commissioner in a video retweeted in May 2020 by the President.[1343]

The question which remained to be answered was the type of forum constituted by the interactive space of a tweet sent by @realDonaldTrump. A forum can be a traditional public forum, defined by the Supreme Court as spaces "*hav*[*ing*] *immemorially been held in trust for the use of the public, and, time out of mind, have been used for purposes of assembly, communicating thoughts between citizens, and discussing public questions*."[1344] Judge Buchwald found that the interactive space of a tweet is not a traditional public forum, even though she noted that the Supreme Court had referred in Reno v. ACLU as the "*vast*

[1343] Aron Blake, *'The only good Democrat is a dead Democrat.' 'When the looting starts, the shooting starts.' Twice in 25 hours, Trump tweets conspicuous allusions to violence,* THE WASHINGTON POST, (May 29, 2020 10:10 a.m.), https://www.washingtonpost.com/politics/2020/05/28/trump-retweets-video-saying-only-good-democrat-is-dead-democrat/.
[1344] Perry Educ. Ass'n v. v. Perry Local Educators' Assn, 460 U.S. 37, 45 (1983).

democratic forums of the Internet,"[1345] and, in *Packingham*, to social media platforms as one of "*the most important places (in a spatial sense) for the exchange of views.*"[1346]

Judge Buchwald found that the interactive space of a tweet was a designated public forum, defined by the Supreme Court in *Cornelius* as a "*public forum… created by government designation of a place or channel of communication for use by the public at large for assembly and speech, for use by certain speakers, or for the discussion of certain subject,*"[1347] noting that "*governmental intent*" is "*the touchstone for determining whether a public forum has been created*."[1348] The @realDonaldTrump account could be accessed by any member of the public, regarding their political affiliation, who could reply or retweet the messages. Furthermore, Scavino had described the account as a way Donald Trump "*communicates directly with you, the American people!* "[1349] The government cannot exclude public forum speakers from a designated public forum without a compelling governmental interest,[1350] but viewpoint discrimination "*is presumed impermissible when directed against speech otherwise within the forum's limitations*."[1351] Plaintiffs had "*indisputably* [*being*] *blocked as a result of viewpoint discrimination*," as they had been blocked because they had criticized the President, and thus blocking them was unconstitutional.

On appeal, the government argued that the @realDonaldTrump account is not a public forum, noting that it "*belongs to him personally and will remain his account after he*

[1345] Reno v. ACLU, 521 U.S. 844, 868 (1997).
[1346] Packingham, at 1735.
[1347] Cornelius, 473 U.S. at 802
[1348] Citing Gen. Media Commc'ns, Inc. v. Cohen, 131 F.3d 273, 279 (2d Cir. 1997).
[1349] Citing Scavino's stipulation, Stip. 37.
[1350] Cornelius, 473 U.S. at 800
[1351] Rosenberger v. Rector and Visitors of Univ. of Va., 515 U.S. 819, 830 (1995).

leaves office," unlike the official @POTUS Twitter account, which is used by the sitting President. They further argued that the "*essential nature*" of the account had not changed since Donald Trump became President, and that it remained "*a private mechanism that Donald Trump possesses to communicate statements he wishes to make to his followers on Twitter and to any other person who visits the @realDonaldTrump page*." Appellees argued in response that "[*t]he account is akin to a digital town hall, with the President speaking from the podium at the front of the room and assembled citizens responding to him and engaging with one another about the President's statements*." This is a compelling argument, evoking the "*interactive space of a tweet"* found by Judge Buchwald to be a public forum.

The government compared the @realDonaldTrump account as Mar-a-Lago or Hyde Park, residences privately owned by Presidents, where they sought refuge from their duties. This argument would have been more persuasive if the @realDonaldTrump account had been made private: a Twitter account can be made private, and then only guests have the privilege of being able to read and comment the tweets posted on the private account, just as guests may be admitted, or not, in a private residence, owner by a President or by a private citizen. Their arguments that Donald Trump "*tweeted about public affairs even before becoming President, and since assuming the Presidency, he has continued to use @realDonaldTrump to discuss matters unrelated to government business, including purely personal topics*" are not convincing either, as discussing public affairs as a private citizen, as did Donald Trump when he created the account, up until his election, and discussing public affairs as President of the United States do not carry the same weight, nor seek to fulfill the same mission.

The government also argued that "*the public-forum doctrine does not come into play unless a plaintiff has been excluded from a space that is owned or controlled by the government*,"[1352] and that "*because the Constitution protects only against government abridgement of speech, exclusion from such space must be attributable to the use of governmental, rather than private, authority*."[1353] For the government, blocking plaintiffs was a "*private action,*" not a state action, noting further that the account is "*governed exclusively by rules established by Twitter* [*which] controls every aspect of* [*the*] *platform*." The government also argued that blocking the plaintiffs had merely limited their ability to reply directly to the @realDonaldTrump's tweets, but that "*that limitation does not implicate the First Amendment,*"[1354] as *"@realDonaldTrump is used to disseminate Donald Trump's speech, but "is not a forum designated to facilitate the speech of others*." The government described the decision to block as Donald Trump's choice "*to pass over one member of the audience who has previously made hostile remarks*." The government also argued that "[*b*]*locking the plaintiffs does not prevent them from interacting with others on Twitter or from continuing to criticize Donald Trump or his administration on* [*Twitter*]."

The United States Court of appeals for the Second Circuit affirmed the Southern District of New York's judgment in July 2019, holding that President Trump had indeed

---

[1352] Citing West Farms Assocs. v. State Traffic Comm'n, 951 F.2d 469, 473 (2d Cir. 1991). The Second Circuit Court of appeals had held that *"public forum analysis applies only where a private party seeks access to public property, such as a park, a street corner, or school auditorium, in order to communicate ideas to others*."

[1353] Citing Flagg v. Yonkers Sav. & Loan Ass'n, 396 F.3d 178, 186 (2d Cir. 2005), which is not a First Amendment case, but a Fifth Amendment case. The Second Circuit Court of appeals defined there state action as requiring " *both an alleged constitutional deprivation caused by the exercise of some right or privilege created by the State or by a rule of conduct imposed by the State or by a person for whom the State is responsible*", adding that "*the party charged with the deprivation must be a person who may fairly be said to be a state actor*."

[1354] Citing Minnesota State Bd. for Community Colleges. v. Knight, 465 U.S. 271, 288 (1984) (*"A person's right to speak is not infringed when the government simply ignores that person while listening to others.")*

engaged in unconstitutional viewpoint discrimination when he had blocked plaintiffs.[1355] The Court took care to note that it "[*did] not consider or decide whether an elected official violates the Constitution by excluding persons from a wholly private social media account."* What is a "*wholly private social media account*"? If Donald Trump would make his account private, describing by Twitter as "*protecting* [*one's*] *tweets*"?[1356] If a Twitter user chooses this option, he or she receives a request each time another user wishes to follow the account: this request can be granted or denied. The tweets posted from such a protected account are only visible to the followers, which are not, however, be able to retweet them, either with or without comments. As such, a protected Twitter account cannot create an "interactive space of a tweet." Would @realDonaldTrump be a "*wholly private social media account*" if the President would only use it to publish personal messages? As his wife, as First Lady, but also daughter Ivanka Trump and son-in-law Jared Kushner,[1357] have all official roles in his administration, the line between public and private life would be even more difficult to define for Donald Trump than for other U.S. Presidents, less inclined to provide their relatives a place in the White House. The Second Circuit noted that "[*n*]*o one disputes that before he became President the Account was a purely private one or that once he leaves office the Account will presumably revert to its private status."*[1358] The word "*presumably*" reminds us that a President leaving office may or may not "tweet" thereafter in a private capacity. As for the @realDonaldTrump account, at the time, as Donald Trump

[1355] Knight First Amendment Inst. Columbia v. Trump, 928 F. 3d 226 (2019)
[1356] About public and protected Tweets, TWITTER, https://help.twitter.com/en/safety-and-security/public-and-protected-tweets (last visited Dec. 30, 2020).
[1357] Both were senior advisers to President Trump.
[1358] Knight First Amendment Inst. 928 F. 3d at 231.

was President, its "*public presentation… and the webpage associated with it bear all the trappings of an official, state-run account*."[1359]

The Second Circuit found the @realDonaldTrump account to be a public forum, as it was "*one of the White House's main vehicles for conducting official business*,"[1360] noting, for instance, that in June 2017 the White House responded to a request for official White House records from the House Permanent Select Committee on Intelligence by referring the Committee to a statement made by the President on Twitter, or that the President first announced on @realDonaldTrump the firing of Chief of Staff Reince Priebus and his replacement with General John Kelly. The Second Circuit reasoned that @realDonaldTrump:

> "*was intentionally opened for public discussion when the President, upon assuming office, repeatedly used the Account as an official vehicle for governance and made its interactive features accessible to the public without limitation*."[1361]

The government had conceded during the oral argument that @realDonaldTrump was not "*independent of* [*Trump's*] *presidency*," but argued that blocking was not state action. Indeed, if Donald Trump "*is a government actor with respect to his use of the Account, viewpoint discrimination violates the First Amendment*."[1362] The Second Circuit found that

---

[1359] Knight First Amendment Inst. 928 F. 3d at 231.
[1360] Knight First Amendment Inst. 928 F. 3d at 232.
[1361] Knight First Amendment Inst. 928 F. 3d at 237.
[1362] Knight First Amendment Inst. 928 F. 3d at 236, quoting Manhattan Community Access Corp. et al. v. Halleck et al., 587 U.S. ___, 139 S.Ct. 1921, 204 L.Ed.2d 405 (2019) : "*When the government provides a forum for speech (known as a public forum), the government may be constrained by the First Amendment, meaning that the government ordinarily may not exclude speech or speakers from the forum on the basis of viewpoint* ...."

replying to a tweet, retweeting it, even liking it,[1363] are expressive conducts and thus speech. The government argued that Plaintiffs had not been preventing from speaking when they had been blocked because it had only prevented from speaking directly to Donald Trump by replying to his tweets. This argument did not convince the Second Circuit as, while "*Plaintiffs have no right to require the President to listen to their speech*,"[1364] blocking them prevents them from interacting with speaking with other twitter users "*who may be speaking to or about the President*."[1365] As a public forum had been created in this interactive space, the President, once having it opened up, cannot "*censor selected users because they express views with which he disagrees*."[1366] Blocking users leads to censorship. The government attempted to convince the Court that Plaintiffs may use alternative methods, designated as "workarounds" to read the President's tweets, including creating new accounts, or logging out from their account to read the President's tweets, which can be read even if one is not logged in on Twitter, or even searching tweets about the President had been posted by other users and engage with them. All these alternative methods are burdensome, and "*burdens to speech as well as outright bans run afoul of the First Amendment*"[1367] Blocking plaintiffs violated the First Amendment.

[1363] For the Second Circuit (at 237), "*Liking a tweet conveys approval or acknowledgment of a tweet and is therefore a symbolic message with expressive content*." Even using the "like" features to bookmark a tweet, while not connoting approval, marks interest about a particular message among the thousands of messages one can see on a timeline every day, even every hour.
[1364] Citing Minnesota State Bd. for Cmty. Colleges v. Knight, 465 U.S. 271, 283, 104 S.Ct. 1058, 79 L.Ed.2d 299 (1984): "*a plaintiff has "no constitutional right to force the government to listen to their views*."
[1365] Knight First Amendment Inst. 928 F. 3d at 238.
[1366] Knight First Amendment Inst. 928 F. 3d at 238.
[1367] Knight First Amendment Inst. 928 F. 3d at 238 Citing *Sorrell v. IMS Health Inc*., 564 U.S. 552, 566, 131 S.Ct. 2653, 180 L.Ed.2d 544 (2011) (stating that government "may no more silence unwanted speech by burdening its utterance than by censoring its content.")

The government also argued that the @realDonaldTrump account was government speech, as it is controlled by the government. The First Amendment does not require government speech to respect viewpoint-neutrality, as it "*would be paralyzing*."[1368] However, the Second Circuit reasoned that, while Donald Trump's tweets are government speech, he was not engaged in government speech when he blocked plaintiffs. It is not the tweets which are at stake, but rather, the "*interactive features*" of the @realDonaldTrump account,[1369] which is a speech of multiple individuals, not just Donald Trump, who did not exercise any control on these messages.[1370] The second Circuit quoted the Supreme Court *Matal* case, where the Supreme Court warned that the government speech doctrine is "*susceptible to dangerous misuse*," and that passing private speech as government speech "*by simply affixing a government seal of approval, government could silence or muffle the expression of disfavored viewpoints.*"[1371] For the Second Circuit, finding that the interactive space of the @realDonaldTrump tweets to be government speech "*would produce precisely this result*."[1372]

Donald Trump is not the only politicians to have blocked users on social media. Representative Alexandria Ocasio-Cortez was sued on July 9, 2019, the very day of the Second Circuit decision, by former New York assemblyman Dov Hikind, who described himself in the complaint as a "*staunch advocate for Jewish causes and the State of Israel* [*and*] *the founder of Americans Against Anti-Semitism*."[1373] Mr. Hikind has been blocked by

[1368] Matal v. Tam, 137 S. Ct. 1744, 1757, (2017).
[1369] Knight First Amendment Inst. 928 F. 3d at 239
[1370] Knight First Amendment Inst. 928 F. 3d at 238
[1371] Matal, 137 S. Ct. at 1758.
[1372] Knight First Amendment Inst. 928 F. 3d at 240
[1373] Dov Hikind v. Alexandria Ocano-Cortez, No. 1:19-cv-03956 (S.D.N/Y. July 9, 2019).

Representative Ocasio-Cortez. He claimed that he often criticized Rep. Ocasio-Cortez on Twitter, most recently before filing the complaint "*in response to AOC's claims that the United States Government is running "concentration camps" on the boarder, similar to those in the Holocaust*." Mr. Hikind argued that Rep. Ocasio-Cortez "*used Twitter as an important public forum for speech*" and that "[*i*]*n an effort to suppress contrary views, Defendant has excluded Twitter users who have criticized AOC and her positions as a Congress woman via "blocking". This practice is unconstitutional and must end*." Mr. Hiking claimed that Rep. Ocasio-Cortez had violated the First Amendment by blocking him "*because it imposes a viewpoint-based restriction on the Mr. Hikind's access to official statements AOC otherwise makes available to the general public."* Just like Mr. Trump, Alexandria Ocasio-Cortez uses a personal Twitter account, @AOC, to communicate with the public. As noted by Mr. Hikind in his complaint, the U.S. Congresswoman has also an official Twitter account, @RepAOC, which has many less followers than the @AOC account. The Knight First Amendment Institute sent a letter to Rep. Ocasio-Cortez urging her not to block Twitter users based on viewpoint.[1374] The letter, written by Jameel Jaffer, the Institute's Executive Director, argued that "*the @AOC account is a "public forum*" within the meaning of the First Amendment" that is used:

> "*as an extension of* [*Rep. Ocasio-Cortez*]*' office to share information about congressional hearings, to explain policy proposals, to advocate legislation, and to solicit public comment about issues relating to government*" and that "[*t*]*he account is*

[1374] *Knight Institute Sends Letter to Rep. Ocasio-Cortez in Response to Alleged Blocking of Critics on Twitter*, THE KNIGHT INSTITUTE, (AUG. 29, 2019), https://knightcolumbia.org/content/knight-institute-sends-letter-to-rep-ocasio-cortez-in-response-to-alleged-blocking-of-critics-on-twitter.

a digital forum in which you share your thoughts and decisions as a member of Congress, and in which members of the public directly engage with you and with one another about matters of public policy."

This referred to the "interactive space of a tweet" found to be a designated public forum by the New York courts.

U.S. District Senior Judge Frederic Block of the Eastern District of New York ordered Representative Ocasio-Cortez to testify in the Hikind case on November 5, 2019, but the suit was settled after Rep. Ocasio-Cortez issued an apology to Mr. Hikind on November 4, 2019, writing that blocking Mr. Hikind "*was wrong and improper and does not reflect the values I cherish*."[1375]

In December 2020, the Knight Institute sent a letter to Chicago Mayor Lori Lightfoot, after she had allegedly blocked from the mayor's official Facebook page an independent journalist covering Chicago who had posted in the comment's section a link to a video about rising crime rates in Chicago.[1376] The issue of politicians' rights to block accounts is still evolving.

Conclusion

I wrote this article because I wanted to research whether the First Amendment's almost limitless protection of speech is the best way to foster a vibrant marketplace of

[1375] Michael Gold, *Ocasio-Cortez Apologizes for Blocking Critic on Twitter*, THE NEW YORK TIMES, (Nov. 4, 2019), https://www.nytimes.com/2019/11/04/nyregion/alexandria-ocasio-cortez-twitter-dov-hikind.html.
[1376] *Knight Institute Asks Mayor Lightfoot to Stop Blocking Critics from Her Official Facebook Page*, THE KNIGHT INSTITUTE, (Dec. 4, 2020), https://knightcolumbia.org/content/knight-institute-asks-mayor-lightfoot-to-stop-blocking-critics-from-her-official-facebook-page.

ideas on social media, or if the European Union framework, with its heavier arsenal of barriers, is the best way to achieve this goal. I conclude that the First Amendment is the best way to protect free speech. But this belief does not lead to the conclusion that laws cannot be passed to regulate the marketplace of ideas. Not regulating free speech promotes speech without self-censorship, which allows the public to be made aware of the real opinions of an individual, who is less likely to present a doctored version of his or her more extreme opinions. This, in turn, has negative consequences for the speaker, such as being fired, or having an offer to join a law school rescinded.[1377] But this freedom also produces victims daily, particularly victims of hate speech, and we must strive to provide them protection and aid.

The First Amendment aims at producing a vibrant marketplace of ideas, and a marketplace, even in a free market economy, can be protected and even made stronger by antitrust laws. Which type of laws could be the equivalent of antitrust laws for the marketplace of ideas? I came up with this analogy. If someone eats a piece of candy or chocolate that he or she finds very delicious, despite being bad for health, waistline, or teeth, then throws on the street the shiny and colorful wrapper of the treat, we should find a way to clean the public space without forbidding the individual to eat the candy. To help us to recognize what is candy and what is litter, and to help us make enlightened decisions as to whether we really want to eat candy or not, or approve of someone eating candy in

[1377] Amanda Robert, *Law student's admission is revoked over 'racially offensive behavior' on social media*, THE ABA JOURNAL, (June 12, 2020, 11:32 am CDT), https://www.abajournal.com/news/article/law-students-admission-revoked-over-racially-offensive-behavior-on-social-media. The Southern Methodist University's Dedman School of Law in Dallas revoked the admission of an incoming student, "*based on the student's racially offensive behavior recorded on social media*."

spite of its ill-effect for health, we must be educated outside of social media, and leave the echo chambers which our social media accounts have often become.