www.jusletter-it.eu

Ryan Vannin

Hate Speech and the current role of participative platforms

The Swiss perspective

The language of hatred is a complex phenomenon, which still does not know a universally recognized definition. There are also obstacles in identifying and prosecuting those who publish and disseminate hate content, especially because Internet giants such as Facebook and Twitter benefit from a certain legislative immunity and mostly operate in a self-regulated environment. The Swiss legislator, so far, opted for a collaborative approach, where law enforcement authorities involve the platforms only in case of anonymity of the perpetrator or when the published content is deemed illegal and has to be blocked or removed. The intention is to describe the role played by participative platforms and the current legal perspective in Switzerland in combating Hate Speech.

Category of articles: Internet law
Region: Switzerland
Field of law: Internet law

Citation: Ryan Vannin, Hate Speech and the current role of participative platforms, in: Jusletter IT 30 September 2020

ISSN 1664-848X, jusletter-it.weblaw.ch, Weblaw AG, info@weblaw.ch, T +41 31 380 57 77

Contents

1. Introduction
2. Defining Hate Speech online
   2.1. Freedom of Expression
   2.2. Limits to the Freedom of Expression
   2.3. Applicable provisions against hate speech
   2.4. Relevant punishable acts
      2.4.1. Hate against groups of people
      2.4.2. Hate against individuals
   2.5. How online speech occurs
      2.5.1. The meaning of public reach
      2.5.2. The occurrence of hate speech
      2.5.3. Exposure and sharing of hate content
3. The role of participative platforms
   3.1. Hints on Internet governance
   3.2. Participative platforms as media
      3.2.1. Immunity of the media
      3.2.2. Media as defenders of freedom of expression
      3.2.3. Excursus: The Christchurch massacre
   3.3. Participative platforms as «gatekeepers» and «good samaritans»
      3.3.1. An overview
      3.3.2. Gatekeeping
         3.3.2.1. Content removal requested by Courts or authorities
         3.3.2.2. Pre-emptive measures
      3.3.3. Acting as good Samaritans
         3.3.3.1. The interests at stake
         3.3.3.2. Self-regulation
   3.4. The outlook
4. Conclusions

1. Introduction

[1] The advent of the Internet has given anyone the opportunity to express and spread any kind of message, almost free of charge and with the possibility of reaching a wide and numerous audience. From the first sites born as a meeting point for right-wing extremists to the advent of participative platforms such as Facebook and Twitter, malicious people have always found ways to make the most of the possibilities offered by technology and to pursue their malicious purposes. The language of online hatred has therefore always been a constant since the Internet has existed, and over the years it has reached more and more users with greater speed, greater impact and consequences.[1]

[2] It should therefore come as no surprise that, in the light of developments in case law in Switzerland,[2] and a more stringent judicial and legislative approach that is slowly gaining ground in some other countries, it is again appropriate to focus on the role of participative platforms in curbing the spread of hate speech online.

---

1 Matthew Costello/James Hawdon, in: Thomas J. Holt/Adam M. Bossler (eds.), The Palgrave Handbook of International Cybercrime and Cyberdeviance, Zurich (CH) 2019, p. 2 ff. (hereinafter Author, Handbook, p. ...).

2 Judgement of the Federal Supreme Court (Judgement) 6B_1114/2018 of January 29, 2020, destined for publication; see also Decision of the Federal Supreme Court (BGE) 145 IV 23; 143 IV 380; 133 IV 308.

[3] Participative platforms include Internet publishing and broadcasting platforms that do not themselves create or own the content being published or broadcasted. These are a subcategory of so called Internet intermediaries, which are organizations or businesses that «*bring together or facilitate transactions between third parties on the Internet. They give access to, host, transmit and index content, products and services originated by third parties on the Internet or provide Internet-based services to third parties*».[3]

[4] Participative platforms allow the dissemination of user-generated content, but they can also monitor, filter and influence the flow of what is published.[4] The main focus of our analysis is on the contexts and to what extent, especially from the point of view of Swiss law, these platforms should be considered responsible and involved in countering the drift into hate speech.

2. Defining Hate Speech online

2.1. Freedom of Expression

[5] Since there is no recognized universal definition of hate speech, it is first necessary to mention the concept of freedom of expression in order to understand the implications.[5] Freedom of expression is generally considered the liberty to articulate an opinion in any available form without fear of being censored or punished if it diverges from the opinion of the State and of other fellow citizens.[6]

[6] The texts of international declarations protecting fundamental freedoms, including Art. 19 Universal Declaration of Human Rights (UDHR), Art. 19 International Covenant on Civil and Political Rights (ICCPR), art. 4 International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), Art. 10 European Convention on Human Rights (ECHR), as well as numerous national constitutional acts, are permeated by these principles[7]. Art. 19 UDHR states that «[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*».[8] In Switzerland, freedom of expression and information, together with other liberties, such as freedom of the media or artistic expression, are guaranteed at constitutional level. (Art. 16 ff. Swiss Const.).[9]

---

3 Karine Perset, The Economic and Social Role of Internet intermediaries, in: OECD Digital Economy Papers, vol. 171, April 8, 2010, http://dx.doi.org/10.1787/5kmh79zzs8vb-en (visited on March 28, 2020), p. 9.

4 Costello/Hawdon, Handbook, p. 12. Stéphanie Musy, La répression du discours de haine sur les réseaux sociaux, Semaine Judiciaire (SJ) 2019 II, p. 1.

5 Musy, p. 1; Elena Mihajlova et al., Freedom of expression and hate speech, Skopje (Macedonia) 2012, p. 25 ff.

6 Alexander Brown, Hate Speech Law, New York (USA) 2015, p. 5 f.; James Grimmelmann, Internet Law: Cases & Problems, 9th ed., Oregon City (USA) 2019, p. 517; Bikhu Parekh, in: Michael Herz/Peter Molnar (eds.),The Content and Context of Hate Speech: rethinking regulation and responses, New York (USA) 2012, p. 42 f. (hereinafter Author, Hate Speech, p. . . .).

7 Catherine O'Regan, Hate Speech Online: an (Intractable) Contemporary Challenge?, in: Current Legal Problems vol. 71, London (UK) 2018, p. 406.

8 Art. 18–21 UDHR enshrine the so-called «constitutional» liberties, which – together with the remaining articles – have influenced several national constitutions of the International Community. While the UDHR is in itself neither a treaty nor a legally binding act, it is the fundamental constitutive document of the UN and has served as the foundation of two binding covenants: the ICCPR and the ICESCR.

9 Federal Constitution of the Swiss Confederation of April 18, 1999 (Classified compilation [CC] 101) (hereinafter Const.).

2.2. Limits to the Freedom of Expression

[7] Freedom of expression has also its limits.[10] For example, Art. 19(3) ICCPR states that there may be certain restrictions, if provided by law and necessary to respect the rights and reputation of others or to protect national security, public order or for reasons of public health and morals. At the same time, it is forbidden to support «*any manifestation of national, racial or religious hatred which constitutes incitement to discrimination*»[11], as well as any form of «*propaganda or organisation based on theories of racial superiority and incitement to racial discrimination and acts of violence*»[12].

[8] The restrictions today still reflect the post-war period: in fact, the principles of the UDHR played a role in the formation of modern Western political thought, which was then transposed into the domestic legislation of many countries, including those that, during the two wars, had adopted Nazi-Fascist regimes such as Italy, Germany or Japan.[13] Countries that have implemented Art. 20 ICCPR in their domestic regulations are required by law to protect ethnic and racial groups from threatening, abusive, insulting publications that can stimulate forms of hostility and public contempt.[14] A State is called upon to protect those groups with certain ascriptive characteristics who are subject to discrimination and manifestation of intolerance by individuals or groups who, instead of accepting and tolerating differences, use them to spread hatred and discredit.[15]

[9] Thus also Switzerland, which ratified the ICERD on December 29, 1994[16] and was required to legislate accordingly. Art. 261bis of the Swiss Criminal Code[17] entered into force on January 1, 1995,[18] which, in compliance of Art. 36 Const.[19], was enacted to protect the human dignity and public peace.[20] In this sense, all forms of public expression that offend on grounds of religious affiliation, nationality, ethnic origin or disability (and in Switzerland as of February 9, 2020 also on sexual orientation[21]) are not tolerated and might therefore be punishable.[22]

10 Alexander Brown, Hate Speech Law, New York (USA) 2015, p. 2; O'Regan, p. 407.

11 Art. 20 ICCPR.

12 Art. 4 ICERD; see also Parekh, Hate Speech, p. 37–38.

13 Abrams, Hate Speech, p. 118 ff.; Bosco et al., in: Babak Akhgar/Ben Brewster (eds.), Combatting Cybercrime and Cyberterrorism, Zurich (CH) 2016, p. 101 (hereinafter Author, Combatting Cybercrime: Short Title, p. . . .); see also *UN Historic Archives*, https://legal.un.org/avl/ha/cerd/cerd.html (visited on December 24, 2019).

14 Jeremy Waldron, The Harm of Hate Speech, Cambridge (USA) 2021, p. 13 ff., 29; O'Regan, p. 408.

15 Naganna Chetty/Sreejith Alathur, Hate speech review in the context of online social networks, in: Vincent van Hasselt (ed.), Aggression and Violent Behavior 40, Amsterdam (The Netherlands) 2018, p. 110 ff.; Herz/Molnar, Hate Speech, p. 3; see *Snyder v. Phelps*, 562 U.S. 443 (2011) (for an in depth judicial analysis between freedom of expression and hate speech in the United States).

16 Official compilation of the Federal Legislation (OC) 1994 1164.

17 Swiss Criminal Code of December 21, 1937 (CC 311.0) (hereinafter Crim. Code).

18 OC 1994 2887–2888; Documents of the Federal Gazette (BBl) 1992 III 269.

19 The provision regulates restrictions on fundamental rights justified by the public interest or for the protection of the fundamental rights of others. These restrictions have to be proportionate.

20 BGE 133 IV 308 consid. 8.2.

21 BBI 2018 7861; 63,1% of Swiss voters accepted the amendment, https://www.bk.admin.ch/ch/d/pore/va/20200209/det630.html (visited on March 28, 2020).

22 Musy, p. 2; Chetty/Alathur, p. 110 ff.; Waldron, p. 5.

2.3. Applicable provisions against hate speech

[10] Expressions of hate manifest themselves in various forms, each according to its own characteristic trait.[23] But not only are the forms different, so are the victims to whom they are addressed, the public that is directly or indirectly involved and also the channels through which expressions of hatred can be conveyed.[24] Those are not limited to the discredit of classes or groups of people worthy of protection, but also occur against individuals. According to Titley et al., hate is: «*a deep and emotional dislike, directed against a certain object or class of objects. The objects of such hatred can vary widely, from inanimate objects to animals, oneself or other people, entire groups of people, people in general, existence, or the whole world*».[25]

[11] Every recognized episode of hate speech must therefore be analyzed by resolving the tension created between freedom of expression and the limits imposed on it by the applicable provisions.[26] While Art. 261 and 261bis Crim. Code punish mainly offenses of religious, ethnic, racial, of sexual orientation nature,[27] the gratuitous offense of discrediting another person is defined as defamation, and falls within the scope of crimes against honor and the personal sphere (artt. 173–175 Crim. Code).[28] An attack against personal honor is considered an insult (Art. 177 Crim. Code)[29]. If it also takes a violent and serious form that causes fear or fright, it is considered a threat (Art. 180 Crim. Code),[30] or, when it compels another to carry out an act, to fail to carry out an act or to tolerate an act, coercion (Art. 181 Crim. Code).[31] Punishable as well are those acts that cause fear and alarm among the general public (Art. 258 Crim. Code)[32] and the public incitement to commit a felony or an act of violence (Art. 259 Crim. Code).[33] Also the Swiss Civil Code[34] offers some protection against encroaching messages, for example against a violation of personality (Art. 28 Civ. Code)[35].

[12] The Swiss legislation does not distinguish between off or online offenses.[36] However, it emerges that, with the advent of new technologies, and in particular of participatory platforms,

23 Brown, p. 19; O'Regan, p. 406 f.; Waldron, p. 34 f.

24 BGE 131 IV 23 consid. 3; Musy, p. 2; for more insights at international level, see Mihajlova et al., p. 25 ff.; Waldron, p. 34 f.

25 Titley et al., Starting Points for Combating Hate Speech Online, in: Council of Europe publish., Strasbourg (France) 2014, p. 55.

26 Grimmelmann, p. 24; Musy, p. 2.

27 Brown, p. 31 f.; the Crim. Code has included religiously and racially motivated offenses under the title «Felonies and Misdemeanours against Public Order», considering them a legal asset that concerns a group of people to be protected and which is not limited to the dignity of the individual. For this reason, those who are accused of racial discrimination risk up to three years in prison, while those who are accused of insulting only a fine of up to 90 daily rates.

28 BGE 142 IV 18; 137 IV 313 (on discredit because of personal political sympathies).

29 The subtitle in the Italian version of the code is *injuria*, but it only punishes the offense or attack against another person's dignity or *decorum*, e.g., with a statement such as «you're stupid like a donkey»; Brown, p. 24.

30 Which goes under the title «Felonies and Misdemeanors against Liberty»; see also Sine Selman/Monika Simmler, «Shitstorm» – strafrechtliche Dimensionen eines neuen Phänomens, in: Schweizerische Zeitschrift für Strafrecht (ZStrR) 136, Bern (CH) 2018, p. 268 ff.

31 BGE 141 IV 437 (on coercion by stalking).

32 BGE 141 IV 215 (on the definition of «population»).

33 Also under the title «Felonies and Misdemeanours against Public Order».

34 Swiss Civil Code of December 10, 1907 (CC 210) (hereinafter Civ. Code).

35 BGE 138 III 641 (on the civil violation of personality via Internet).

36 Brown, p. 24; Musy, p. 2.

the diffusion of hate content that can be prosecuted by criminal law or disputed in civil courts has grown exponentially.[37] The perpetrators of hatred use the technologies available on and through the Internet to harm, attack, threaten, denigrate, offend, belittle or express displeasure that is decidedly inappropriate against such objects or class of objects,[38] knowing, or being able to know, that such expressions not only reach the recipient(s), but also propagate and spread, immediately and with a certain persistence, to a more or less numerous third party audience.[39]

2.4. Relevant punishable acts

2.4.1. Hate against groups of people

[13] Art. 261bis Crim. Code punishes expressions addressed against groups of people identified by certain ascriptive characteristics, such as race, ethnicity[40], religion[41], sexual orientation[42] with negative meanings and which could be a source of prejudice and undermine the human dignity (negative stereotyping or stigmatization).[43] The constituent elements of the offense take into account who the victims are, the repressed behavior and the reaching of a public audience.[44] The concept of race refers to groups of people identified by certain ascriptive characteristics, notably hereditary traits such as skin color, physiognomy etc.;[45] while the concept of ethnicity refers to people identified by their history and culture (language, tradition, ways of living),[46] like Kosovar Albanians, Armenians or Romani people.[47]

[14] It also concerns expressions that intentionally (or that suggest an intent) arouse, incite and promote feelings of hatred or hostility towards members of groups of people identified by race, ethnicity, religion and sexual orientation.[48] Incitement to hatred can also be prosecuted without the incitement actually leading to an offense or a criminal act.[49] So a statement published and openly accessible on Facebook such as: «*J'organise une kristallnacht. Qui est partant pour aller bruler du muzz?*»[50] is punishable under the para. 1 of the provision[51] as a form of clear incitement

37 Selman/Simmler, p. 1 ff.

38 Titley et al., p. 55; see also Ben Wagner, Global Free Expression – Governing the Boundaries of Internet Content, Zurich (CH) 2016, p. 147.

39 Judgement 6B_410/2011 of December 5, 2011 consid. 2; BGE 131 IV 160 consid. 3.3.3; 128 IV 53 consid. 1a; see also Chetty/Alathur, p. 108; Graham, Handbook, p. 7; Grimmelmann, p. 129.

40 BGE 124 IV 124 consid. 2b.

41 BGE 143 IV 193 consid. 2.3.

42 Cfr. Judgement 6B_361/2010 of November 1, 2010 consid. 4 ff.

43 Andreas Donatsch/Wolfgang Wohlers, Strafrecht IV, 4th ed., Zurich/Basel/Geneva (CH) 2011, p. 223 ff.; Musy, p. 3.

44 Doritt Schleiminger Mettler, Art. 261bis, in: Basler Kommentar Strafrecht, Basel (CH) 2017, N 2 ff. (hereinafter BSK-StGB-Author, Art. . . . , N . . .).

45 BGE 123 IV 202; Donatsch/Wohlers, p. 227.

46 BSK-StGB-Schleminger Mettler, Art. 261bis, N 15.

47 Musy, p. 4.

48 As for para. 1 of the norm; cfr. BGE 143 IV 193.

49 BGE 123 IV 202, consid. 2a ff.

50 Paraphrasing: «I'm organizing a punitive expedition like the one occurred during the Nazi-regime. Who's joining to go and burn some muslims?».

51 Judgement 6B_267/2018 of May 17, 2018.

to hatred and of promotion of a racist ideology.[52] Instead, an expression of direct racist exclamation by means of a Tweet such as «*Vielleicht brauchen wir wieder eine Kristallnacht... diesmal für Moscheen*»[53] is prosecuted under Art. 261bis para. 4 Crim. Code.[54]

[15] Regardless of the conduct apprehended, the expression must be sufficiently serious for it to be considered an affront to human dignity,[55] so that a xenophobic, tasteless, amoral or morally offensive, or unbecoming or uncivilized statement relating to an ethnic group, race or religion does not immediately constitute racial discrimination.[56] Art. 261bis Crim. Code does not apply to xenophobic expressions that do not refer to one of the protected groups described above: expressions such as «bastard foreigner» or «dirty refugee» do not constitute racial discrimination. According to the Swiss Federal Supreme Court[57] prohibiting those forms would violate the right to freedom of expression.[58] The European Court of Human Rights (ECtHR) found Switzerland in violation of Art. 10 ECHR for applying Article 261bis para. 4 Crim. Code against an individual who denied the Armenian genocide.[59]

[16] In addition, a judge must take into account the meaning conveyed by the expression: it is irrelevant what the author intends or wanted to intend, instead it matters what the average user, without prior knowledge, can understand given the context in which the expression was made manifest.[60] In this sense, hate speech is also given by the publication of an image in the social media portraying a group of individuals performing the «Hitler's greeting» in front of a synagogue, rejecting the argument that it was a tribute to a sketch by the French humorist Dieudonné.[61] The Federal Supreme Court adopted the same criterion regarding the meaning understood by an average user to the case of the post referred to the «muzz» published on Facebook.[62]

[17] Expressions of hatred addressed against groups of people identified by certain ascriptive characteristics under Art. 261bis Crim. Code must be made available to the public.[63] Conduct in a purely private context is not covered by the provision.[64]

---

52 Cfr. BGE 140 IV 102 consid 2.2.2 ff.

53 Paraphrasing: «We need an other night of the crystals, but this time towards Mosques».

54 Judgement 6B_627/2015 of November 4, 2015.

55 BGE 140 IV 67 consid. 2 ff.; Musy, p. 4.

56 Cfr. BGE 143 IV 308 consid. 4.1; Marcel Alexander Niggli, Rassendiskriminierung. Ein Kommentar zu Art. 261bis StGB und Art. 171c MStG, Zurich (CH) 2007, N 945 ff.; Donatsch/Wohlers, p. 235.

57 BGE 140 IV 67 consid. 2 f.; Musy, p. 6.

58 Judgement 6B_168/2011 of July 18, 2011 consid. 3; Donatsch/Wohlers, p. 230; cfr. BGE 131 IV 27.

59 Case *Perinçek vs. Suisse*, Judgement of the ECtHR (n° 27510/08) of October 15, 2015, N 112 ff.; see also BGE 118 IV 153 consid. 4c (on the prevalence of the public interest or freedom of expression).

60 BGE 140 IV 67 consid. 2.1 f.; 133 IV 308 consid. 8.5 f.

61 BGE 143 IV 308 consid. 4 f.; see also Musy, p. 7–8.

62 Judgement 6B_627/2015 of November 4, 2015 consid. 2.3 f.; see also Judgement 6B_805/2017 of December 6, 2018 consid. 2.3 (the author must intentionally express his hatred against one of the protected groups).

63 BGE 123 IV 202 consid. 3; Judgement 6S.148/2003 of September 16, 2003 consid. 2.3; Donatsch/Wohlers, p. 230 f.

64 Cfr. BGE 130 IV 111 consid. 5.2.1; 126 IV 176 consid. 2; 123 IV 202 consid. 3; see also BSK-StGB-Schleminger Mettler, Art. 261bis, N 39.

2.4.2. Hate against individuals

[18] Attacks against honor, which are offences of expressions of thought, can also be taken into account when it comes to repressing hate speech published on participative platforms, in particular when the content of hatred is not addressed against groups of people protected under Art. 261bis Crim. Code.[65] Pursuant to Art. 173 ff. Crim. Code, the victim must be a specific or identifiable person.[66] Unlike expressions of hatred against persons belonging to a protected group, an attack against a group of persons as a whole does not constitute an offense to the honor of each of the individuals belonging to that group, unless a smaller circle can be identified which differs from the group as a whole (or, as mentioned earlier, are members of a protected group).[67]

[19] Offensive expressions of honor, regardless of the means used, which diminish the value or undermine the reputation of the person, dismantle or derogate from his or her status as a human being, are therefore punishable.[68] Thus are offensive epithets with misused references to mental illness, such as psychopath, querulous, horny pervert, mongrel.[69] It is not punishable who instead hurls himself against foreigners as a whole, unless the stigmatized people in that group can be individualized.[70]

[20] The various articles protecting individual honor are applicable according to different criteria.[71] Art. 173 Crim. Code implies a de facto allegation (which is not substantiated, i.e. false), whereas Art. 177 Crim. Code requires a simple value judgement, a bias. A criticism, an appreciation, is a value judgement, which, depending on the circumstances, becomes injurious under Art. 177 Crim. Code. Therefore, expressions such as «crook» or «whore», if they do not state a fact but are simply a manifestation of contempt, are punishable.[72] Not so if the expressions have a connotation to affirm political positions, even if strong (*e.g.*, «lying bastard»).[73] For attacks against honor prosecution is initiated following a complaint by a party.

[21] Finally, attacks on freedom are also punishable, such as threats (Art. 180 Crim. Code) and coercion (Art. 181 Crim. Code) when the content is so intense as to cause fear or frighten the victim, or to hinder the victim's freedom to act, forcing him or her to do, omit or tolerate an act.[74] Typical episodes are cyber stalking or cyber harassment, including trolling, but also the so-called «firestorm effect» against individuals (and also against legal persons), which is triggered perhaps by smaller, but no less serious, episodes of cyberbullying or «revenge porn».[75]

---

65 BGE 105 IV 111 consid. 1 ff.; Musy, p. 10.

66 BGE 100 IV 43 consid. 1 ff.; Andreas Donatsch, Strafrecht III, 10th ed., Zurich/Basel/Geneva (CH) 2013, pp. 372, 374.

67 BGE 105 IV 111 consid. 1 ff.; 124 IV 262 consid. 2 f.; Donatsch, p. 376.

68 Instead of many, Laurent Rieben/Miriam Mazou, Art. 173 CP, in: Commentaire Romand Code pénale II, Basel (CH) 2019; N 1 ff. (hereinafter CR-CP-II-Author, Art. . . . , N . . .).

69 BGE 131 IV 157; 93 IV 21; 96 IV 54; 98 IV 93; cfr. BGE 76 IV 30.

70 BSK-StGB-Riklin, Art. 173, N 55; see also Niggli, N 1792 ff.

71 CR-CP-II-Rieben/Mazou, art. 173, N 7.

72 CR-CP-II-Rieben/Mazou, art. 173, N 8.

73 Cfr. judgement 6B_1270/2017, 6B_1291/2017 of April 24, 2018.

74 BSK-StGB-Delnon/Rüdy, art. 181, N 14 f.; cfr. BGE 129 IV 262.

75 Selman/Simmler, p. 268 f.

2.5. How online speech occurs

2.5.1. The meaning of public reach

[22] Participative platforms allow users to adjust privacy and confidentiality settings of their content. If a user decides to leave all his Facebook posts accessible, then anyone accessing the platform can view his content, including content shared by and with other users.[76] On Twitter in essence every Tweet is publicly accessible and visible, even to those who are not users of the platform.[77] Participative platforms differ in their user experience, but it is typical that a Facebook user limits the sharing of their content to a circle of «friends», or «friends of friends».[78] If the virtual circle of contacts of a user who publishes hate-expressed content is made up of family members or limited to friendships, then the character of openness to the public within the meaning of Art. 261bis Crim. Code is lost.[79]

[23] However, there is no unequivocal interpretation of the concept of availability and accessibility to the public, which requires the judge to assess each individual case in the light of the circumstances.[80] So a grouping of 40–50 skinheads gathered from different places, even if moved by common interests, cannot be considered a restricted circle of acquaintances.[81] The concept of friendship in relation to participatory platforms is not the same as that understood traditionally.[82] Relating on social media, thus establishing a connection with another person only online without ever having interacted live or shared in any way common experiences in proximity, or without having exchanged feelings of appreciation or understanding with each other, cannot be considered a friendship.[83] Having 300 followers on Twitter, according to the Federal Supreme Court, cannot be considered a restricted circle of friends, all the more so if the user does not take any action to limit the possibility of sharing the tweet, which is considered injurious to personality (ex Art. 28 Civ. Code).[84]

[24] Jurisprudence and scholars agree that, rather than the number of friends or followers in the context of the participative platforms, it is rather decisive whether third parties are given the possibility to access the content without particular restrictions or limitations.[85]

---

76 Facebook's Help Center: Your Privacy, https://www.facebook.com/help/238318146535333 (visited April 11, 2020).

77 Twitter's Help Center: About public and protected Tweets, https://help.twitter.com/en/safety-and-security/public-and-protected-tweets (visited April 11, 2020).

78 Musy, p. 9.

79 BGE 130 IV 111 consid. 5.2.1.

80 BGE 141 IV 215 consid. 2.3.4 (definition of population ex Art. 258 Crim. Code: the individuals with whom the author of a statement is connected through friendship or acquaintance in real or virtual life: for example 290 Facebook friends, are not to be regarded as «population»); not so in Judgement 6B_43/2016 of June 23, 2017 consid. 2.4.4; see also BSK-StGB-Schleminger Mettler, Art. 261bis, N 22 ff.; Musy, p. 10.

81 BGE 130 IV 111 consid. 5 f.; BSK-StGB-Schleminger Mettler, Art. 261bis, N 23.

82 BGE 130 IV 111 consid. 5.2 f.

83 BGE 144 I 159 consid. 4 ff. (in connection with a request of recusation of a judge who was among the «friends» of a party on Facebook).

84 Judgement 5A_195/2016 of July 4, 2015 consid. 5; BSK-StGB-Schleminger Mettler, Art. 261bis, N 24.

85 Judgement 6B_1114/2018 of January 29, 2020; BGE 130 IV 111 consid. 3; 111 IV 151 consid. 2; 123 IV 202 consid. 3d; Donatsch, p. 375, 392; Musy, p. 11; BSK-StGB-Riklin, Art. 173, N 4; BSK-StGB-Schleminger Mettler, Art. 261bis, N 22 ff.; Selman/Simmler, S, 261 f.

2.5.2. The occurrence of hate speech

[25] According to Costello/Hawdon, one of the main factors of the occurrence of hate speech on websites and participative platforms recalls a variation of the «lifestyle-routine activity theory», namely that criminal activities take place when a motivated offender, a suitable victim, and the absence of an adequate defense converge in time and space.[86]

[26] First, the nature of the web allows asynchronous communication and permanence, so a certain post or comment is kept over a period of time. In addition, the pervasiveness and proximity, albeit in a virtual space, means that victims are exposed to hatred even without interacting directly with perpetrators.[87] Exposure tends to increase with both the number of sites or platforms visited and the time spent online.[88] Another element is the suitability to be a victim: in other words, someone who is able to satisfy the desires of the motivated offender. In practice, one tends to be more exposed if one reacts to a content of hatred (*e.g.*, by responding under a comment) or if one exposes a topic or argument (*e.g.*, of a political nature) attracting the attention and anger of those who are willing to express hatred.[89] The last element of the theory concerns the absence of supervision to act as a deterrent, so that the offender does not commit the crime against the designated victim.[90]

[27] Another factor refers to the dynamics of the «pack»: people with similar experiences meet, interact and share the same world views, but instead of opening up and embracing different ideas they continue to feed the same arguments in a spiral that tends to lead to extreme positions.[91] Added to this are the algorithms that customize the experience on the site or platform to reflect the user's tastes and interests: it has been noted that these mechanisms often tend to create a «Filter Bubble»[92] around users, therein they are increasingly exposed only to content that is related to their preferences and to what they already agree. This dynamic could therefore lead dissatisfied users to join extreme-right groups or other radicalized groups and strengthen their biases.[93]

2.5.3. Exposure and sharing of hate content

[28] One of the biggest concerns about exposure to hate content is the potential for involvement and radicalization. Studies have shown a link between exposure to online violence and violent acts, including mass violence and even terrorism.[94] There are numerous cases that have had

86 Costello/Hawdon, Handbook, p. 7 (quoting Cohen/Felson, Social change and crime rate trends: A routine activity approach, in: American Sociological Review 44, Boston [USA] 1978, pp. 588–608).

87 Costello/Hawdon, Handbook, p. 7.

88 Costello/Hawdon, Handbook, p. 7.

89 Reyns/Fissel, Handbook, p. 14.

90 The hypothesis that there is a lack of supervision does not fully reflect reality: verification and filtering mechanisms exist, but their effectiveness may be questioned; see also Bruder Kleinschmidt, An International Comparison of ISP's Liabilities for Unlawful Third Party Content, in: International Journal of Law and Information Technology 18/4, Oxford (UK) 2010, p. 335 ff.; Reyns/Fissel, Handbook, p. 14.

91 Danielle Keats Citron, Hate Crimes in Cyberspace, Cambridge (USA) 2014, p. 57 ff.; Harel, Hate Speech, p. 314 ff.

92 Eli Pariser, The Filter Bubble, How the New Personalized Web is Changing What We Read and How We Think, New York (USA) 2012, p. 47 ff.

93 Pariser, p. 9.

94 Costello/Hawdon, Handbook, p. 11.

perpetrators who have been, in some way, involved or influenced by the widespread of hatred online.[95] On the other hand, there is also another aspect, namely that of content sharing. The courts recognize the punishability not only of new hate content, but also the dissemination of existing content to a wide audience that re-ignites the effects of the original expression.[96] On participative platforms this can be done by an act of appreciation (the «like»), by a comment below the original post or by a re-publication (the «share» or «retweet») on the same platform or on third party sites. It is therefore possible, with extreme ease, to take advantage of the available features to bring other users into contact and contribute to the dissemination of the content.[97]

[29] The Federal Supreme Court recently upheld the ruling of a Zurich district court, according to which a «like» is not only an act of appreciation, but also an instrument of promotion, since activating it also allows the circle of contacts to become aware of that specific liked content.[98] In this sense, whoever puts a «like» on a content of hate that can be criminally prosecuted commits the same offense as the author of the original post, since it contributes to spreading the message with the same intent.[99]

[30] In my opinion a «like» does not have the same impact as a «share», because it is a vague gesture, operated for other reasons and not necessarily related to the post, whose content perhaps is not even known.[100] Moreover, just with reference to the algorithms, there is no evidence that a «like» of a user is notified to all his contacts, nor is it certain that it appears among the posts of the so-called wall.[101] This is also true, albeit in a minor way, for activating a «share» and commenting. A «like» cannot be compared to the concept of propagation within the meaning of Art. 261bis and 173 Crim. Code, nor can one who «likes» be considered an accomplice, since the punishable act (posting and making available the content of hate) would have been carried out even without the «help» of a third party.[102]

[31] The foregoing should be not applicable to a retweet, since it is a republication of the original text, and all the constituent elements of a new infringement, already given in the first act, are thereby realized. Therefore, whoever makes a retweet is punishable as the principal author, unless the medium (Twitter) is entitled to immunity under Art. 28 para. 1 Crim. Code, so that only the author is the sole responsible.[103] So for example, the district court of Zurich established that a retweet is a Twitter-typical distribution and the retweeter may thus not be qualified as author according to the Swiss Criminal Code. Nonetheless a retweet might constitute a personality right infringement under Civil law.[104]

95 Keats Citron, p. 27 ff. and p. 177 ff. (On 4chan's «recruiting» technique).

96 BGE 118 IV 153 consid. 4a; Selman/Simmler, p. 261 f.

97 Musy, p. 12.

98 District Court of Zurich GG160246 of May 29, 2017.

99 Judgement 6B_1114/2018 of January 29, 2020 consid. 2 f.

100 See also Musy, p. 13; cfr. Daniel A. Freitag, Straf- und Strafprozessrecht / Liken von rechtsextremen Inhalten auf Facebook, in: Digitalisierung – Gesellschaft – Recht, Zurich/St. Gallen (CH) 2019, p. 356 f.

101 Pariser, 32 ff.

102 Freitag, p. 353 ff.; Rafael Studer, Straflosigkeit des Likens – Exemplifikation anhand ehrverletzender Tatsachenbehauptungen auf Facebook, in: recht, Bern (CH) 2018, p. 183.

103 Musy, p. 13 f.; BSK-StGB-Riklin, Art. 173, N 4; see more at 3.2.

104 Cfr. District Court of Zurich GG150250 of January 26, 2016 (not convincing, since the constituent elements are given).

3. The role of participative platforms

3.1. Hints on Internet governance

[32] Many users (still) believe that rules set offline do not apply online, and vice-versa.[105] The history of the Internet contradicts this belief, since from the very beginning laws and agreements have been enacted – between states, but also between different types of organizations and interest groups –, which on the one hand protect principles such as open access and the democratization of the Internet, and on the other hand try to limit its excesses and abuses.[106] There is therefore a complex that makes Internet governance, implemented by different actors (in a «multi-stakeholder» regime), very fragmented.[107] This does not, of course, preclude the role of the States.[108]

[33] While among the various stakeholders involved in the governance there is an immediate convergence of interests towards certain contents (e.g. against pornography, copyright infringement or the protection of minors), this is not always the case in regards to expressions of hatred, as restrictions might lead to a «chilling effect» on the freedom of expression.[109] The data show that two thirds of the world's Internet traffic flows to platforms managed by Facebook and Google, which in some way has relevance on the governance of Internet.[110] Also, the main participatory platforms are established in the United States, where there is a milder control by authorities and where laws and justice tend not to impose stringent forms of censorship on content, in order not to violate the First Amendment and to comply with § 230 Communications Decency Act (CDA).[111]

[34] In general, the cross-border nature of participatory platforms makes the application of existing rules complex.[112] Irrespective of the location of the platforms, the criminal prosecution of hate speech on the Internet under Swiss law takes place with the application of the general rules on territoriality. Art. 3 and 8 Crim. Code are therefore applicable, so that an action is punishable where the plaintiff is located, but also where the action is concluded or, as recent case law has shown, where the victims against whom the hate speech is directed are located.[113]

[35] On the forms of cooperation with law enforcement authorities, *e.g.*, in the case of users hiding behind anonymity, the platforms generally agree to release information about their users only to

105 Lyle et al., Combatting Cybercrime, p. 127 (citing the report of the UK Communications Committee to the Parliament of 2014).

106 Christopher Kuner, Data Protection Law and International Jurisdiction on the Internet (Parts 1&2), in: International Journal of Law and Information Technology 18/2&3, Oxford (UK) 2010, p. 176 ff., p. 227 ff.; Saxon R. Shaw, There is No Silver Bullet: Solutions to Internet Jurisdiction, in: International Journal of Law and Information Technology 25, Oxford (UK) 2017, p. 284; Wagner, p. 4 (addressing the problem of Internet jurisdiction).

107 David Equey, La responsabilité pénale des fournisseurs de services Internet, diss., Bern (CH) 2016, N 58; Kate Klonick, The New Governors: the People, Rules, and Processes Governing Online Speech, in: Harvard Law Review 131, Cambridge (USA) 2018; p. 1655; Shaw, p. 285.

108 Grimmelmann, p. 515 ff.; Wagner, p. 10 ff.

109 Instead of many, see *Reno v. American Civil Liberties Union*, 512 U.S. 844 (1995); Grimmelmann, p. 178 ff.

110 Martin Amstrong, Referral Traffic – Google or Facebook?, in: Statista May 24, 2017, https://www.statista.com/chart/9555/referral-traffic---google-or-facebook/ (visited on October 14, 2019).

111 Section 230 of the Communications Decency Act of 1996 (47 U.S.C. 230); for more details see also Costello/Hawdon, Handbook, p. 1; Grimmelmann, p. 185 ff.

112 Statement of the Federal Council 17.3734 of December 1, 2017.

113 BGE 125 IV 177 consid. 3; cfr. BGE 105 IV 326 consid. 3.

US law enforcement agencies with a legitimate court order or subpoena, otherwise by means of international letter rogatory issued by a Swiss court.[114]

[36] In civil matters, the jurisdiction of a particular Swiss court must be examined in accordance with Art. 10 e 129 Federal Act on Private International Law (PILA). Art. 129 PILA provides for a forum at the domicile or seat of the defendant, as well as a forum at the place of the act or result. The place of the result is the place where the direct and immediate injury to the legally protected interest occurred.[115] In the case of a civil violation committed on the Internet (*e.g.*, of Art. 28 Civ. Code), these principles would allow the victim to file an action before the courts of each State where the disputed content is accessible. To avoid such a result, it is required, in addition to the accessibility to the content, a certain proven connection with Switzerland.[116]

3.2. Participative platforms as media

3.2.1. Immunity of the media

[37] Largely due to the role of § 230 CDA, which grants Internet intermediaries extensive immunity in terms of liability for user-generated content, no participative platform in the United States has so far incurred in sanctions for violation of content posted by users on their sites deemed abusive.[117] Switzerland also recognizes a privilege granted to the media under Art. 28 para. 1 Crim. Code: if an offense is committed or completed in the form of publication in the media, then only the author is liable to prosecution.[118] Those who are part of the chain of diffusion typical of the medium (booksellers, kiosk-shopkeepers, newsagents, delivery men, sign layers) and who make accessible to the public contents of hate liable to criminal prosecution, are not punishable.[119] The normative idea, assimilable to the United States' one, is that it is not intended to paralyze freedom of expression through the press by pursuing individually each person involved in the publication in question.[120] At the same time, there is a willingness to prevent the media from exercising preemptive censorship of content, for fear of incurring in possible sanctions.[121] Scholars and district courts recognize participatory platforms such as Facebook and Twitter as media in accordance with Art. 28 para. 1 Crim. Code.[122]

[38] However, users may not avail themselves of immunity for infringements committed under Art. 173 line 1 para. 2 and Art. 261bis para. 4 Crim. Code, *i.e.*, when hate content is redistributed

114 Facebook's Help Center: Information for Law Enforcement Authorities, https://www.facebook.com/safety/groups/law/guidelines/ (visited on April 11, 2020); Twitter's Help Center: Law enforcement guidelines, https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support (visited on April 11, 2020).

115 Julien Francey, La responsabilité délictuelle des fournisseurs d'hébergement et d'accès Internet, Zurich/Basel/Geneva (CH) 2017, N 134 ff, 321 ff.

116 Judgement 5A_812/2015 of September 6, 2015; Case *CICAD vs. Switzerland*, Judgement of the ECtHR (17676/09) of June 7, 2016.

117 Klonick, p. 1655; David Talbot/Jeff Fossett, Exploring the Role of Algorithms in Online Harmful Speech, in: Berkman Klein Center Collection 08/19, Cambridge (USA) 2017, https://medium.com/berkman-klein-center/exploring-the-role-of-algorithms-in-online-harmful-speech-1b804936f279 (visited on October 14, 2019)

118 Freitag, p. 349; BSK-StGB-Riklin, Art. 173, N 4.

119 District Court of Zurich GG150250 of January 26, 2016; Musy, p. 14; Selman/Simmler, p. 262 ff.

120 BGE 128 IV 53 consid. 5e; BSK-StGB-Zeller, Art. 28, N 11 f.

121 BSK-StGB-Zeller, Art. 28, N 12; see also Keats Citron, p. 66 ff. (on the information cascade in the United States).

122 District Court of Zurich GG150250 of January 26, 2016 consid. 2; GG160246 of May 29, 2017; Musy, p. 13; Selman/Simmler, p. 262 ff.; cfr. judgement 6B_1114/2018 of January 29, 2020 consid. 3.

or when the expression of hatred has obvious discriminatory purposes. Similarly, other offences involving the constituent element of the availability of hate content to the public can also be prosecuted (*e.g.*, Art. 259 Crim. Code).[123]

3.2.2. Media as defenders of freedom of expression

[39] The U.S. § 230 CDA is a legal instrument with a fundamental contradiction: on one hand it wants to encourage the removal of certain content, on the other hand it wants to avoid the cluttered removal and the dangers of collateral censorship for users.[124] Participative platforms have to be at the same time both «good Samaritans» and intervene on controversial contents, and defenders of freedom of expression and avoid any restriction in this sense.[125]

[40] The assumption, which has not yet been refuted, is that platforms, on a voluntary basis and without consequences, decide whether and how to moderate the content, reinforced by three concepts.[126] The first concept that has been affirmed by the United States' case law is that platforms are comparable to public places or actors of the state (and therefore constitutionally obliged to guarantee the exercise of freedom of expression).[127] In a way, this concept has not fully convinced, although a recent ruling has compared social media to being entities that perform quasi-public functions.[128] The second concept is that participative platforms fall within the category of traditional media, *i.e.*, television and radio, and in this sense they must give a proper and fair voice to all parties.[129] The last concept, by analogy with the First Amendment, wants the platforms themselves to play an important role in guaranteeing freedom of expression in the world and must therefore be protected. The example is that of the printed press and the right of reply, which can be refused under the freedom of the press clause and considered an unconstitutional intrusion into the editorial function exercised by the platforms.[130]

[41] The fear that participative platforms may opt to exercise collateral censorship and thus restrict freedom of expression has meant that, at least so far, none of the above concepts have in any way called into question, at least legally, their immunity under § 230 CDA.[131] Similarly, in Switzerland, as long as the platforms do not exercise any control over the content published and distributed on their sites, these run no risk of criminal prosecution.[132] Switzerland also does not consider it appropriate to adopt a measure that sanctions (fines) platforms that do not comply with a simple request for blocking or deletion of content deemed abusive by a user, as recently introduced in Germany, because it may conflict with the freedom of expression.[133] More generally,

123 BGE 125 IV 177 consid. 5b.; 126 IV 176 consid. 2; judgement 6B_645/2007 of May 2, 2008 consid. 6.

124 Klonick, p. 1608.

125 Clark et al., p. 19.

126 Klonick, p. 1606.

127 See *Marsh v. Alabama*, 326 U.S. 501 (1946); cfr. *Logan Valley in Hudgens v. National Labor Relations Board*, 424 U.S. 507 (1976).

128 Klonick, p. 1609 (citing *Packingham v. North Carolina*, 137 S.Ct. 1730 [2017], «*the decision makes access to private online platforms a First Amendment right*»).

129 Grimmelmann, p. 185.

130 Klonick, p. 1613 (citing *Miami Herald Publishing Co. v. Tornillo*, 418 U.S. 241 [1974]).

131 Waldron, p. 29; O'Regan, p. 408.

132 Equey, N 62 ff., 81; Musy, p. 16.

133 Statement of the Federal Council 17.3734 of December 1, 2017; Report on cybercriminality issued by the Federal Department of Police and Justice (FDJP) on June 2013 (hereinafter FDJP Cybercrime, p. . . .); see also

the media, and thus the platforms, have a cascading liability under Art. 28 para. 2 and 3 Crim. Code, i.e. only in the event that the author of the content cannot be identified, the authorities could then consider prosecuting the media directly, *i.e.*, prosecuting the editor or publisher.[134] With regard to participatory platforms, no such case history is known.[135]

[42] Participative platforms may be involved in civil litigation if the nature of the injury results in a violation of personality and there is a direct causal connection between the content and the platform, typically under Art. 28 Civ. Code.[136] It has to be noted that denialism in Switzerland is prosecuted as crime, because it is considered a violation against public order, a falsehood that violates the interests of all citizens (Art. 261bis para. 4 Crim. Code), while in the United States denial of the Holocaust can only be challenged in civil courts, with a limited chance of obtaining a favorable judgement.[137] The United States leaves it to individuals to deal with hate speech and deliberately limit interference and excessive state power over its own citizens.[138]

3.2.3. Excursus: The Christchurch massacre

[43] In March 2019, Brenton Tarrant, an Australian extreme right-wing sympathizer, in an attack on a mosque during a religious service in Christchurch, New Zealand, took 50 lives while live-streaming his story on Facebook. Tarrant also posted on 8chan 78 pages of his manifesto, including a link to the streaming of the forthcoming massacre. Many lawmakers have learned that hate content posted online by Tarrant had escaped Facebook filters and law enforcement attention for several hours, and that it was only removed from Menlo Park's social media after an explicit request from the New Zealand authorities.[139] Despite the fact that the streaming was deleted from the platform, 6 million people saw what was happening in real time and to this day it is still possible to trace both the footage, taken from other sites and platforms.[140]

[44] The non-immediate reaction to what was happening on Facebook has set in motion a series of measures: the best known is the introduction of a law against «abhorrent violent material» (2019)[141] in Australia, passed to criminalize not only the perpetrators of violence, but also to punish sites or platforms that host violent content and incite hatred if they do not act within

---

Katrin Bennhold, Germany Acts to Tame Facebook, Learning From Its Own History of Hate, in: NYT May 19, 2018, https://www.nytimes.com/2018/05/19/technology/facebook-deletion-center-germany.html (visited on May 13, 2019).

134 Cfr. BGE 121 IV 109.

135 BGE 121 IV 109 consid. 5e; BSK-StGB-Zeller, Art. 28, N 12.

136 Judgement 5A_437/2016 of December 2, 2016; cfr. Judgement 5A_195/2016; 5A_975/2015 of July 4, 2016 consid. 5.1; BSK-StGB-Zeller, Art. 28, N 10.

137 BGE 126 IV 20 consid. 1e; 121 IV 76 consid. 2; see also Niggli, N 1485; Schauer, Hate Speech, p. 130 ff.; Suk, Hate Speech, p. 145 ff.

138 Robert J.Boeckmann/Carolyn Turpin-Petrosino, Understanding the Harm of Hate Crime, in: Journal of Social Issues 58/2, Washington (USA) 2002, p. 211 ff.

139 Ian Bogost, Social Media Are a Mass Shooter's Best Friend, in: The Atlantic March 15, 2019, https://www.theatlantic.com/technology/archive/2019/03/how-terrorism-new-zealand-spread-social-media/585040/ (visited on October 5, 2019).

140 Bogost, Ibid.

141 Australia's Criminal Code Amendment (Sharing of Abhorrent Violent Material) – Bill 2019 Act 45, https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=s1201 (visited on May 6, 2019).

a reasonable time to report and remove it from public view.[142] The Australian government has essentially sought to counter the damage of hate speech through the threat of fines and imprisonment by pressure on Facebook-like platforms to be more responsible, and has worked to identify and block entire sites that host even a fraction of content deemed illegal.[143] However, the authorities responsible for implementing the directives, which provide for a notice issued to sites and platforms relying on complaints, are not entirely convinced of the effectiveness of the system, which is reminiscent of the one against child pornography.[144] In their opinion, illegal images and videos of minors are a target that leaves no doubt, but this is not the case when it comes to content that intertwines politics and violence, on the borders of what is permissible under the criteria of freedom of expression.[145]

[45] For critics, moreover, Australian law still leaves too much power to participative platforms to decide what to take down without having to disclose their reasons: according to them, sites and platforms would be encouraged to operate a pre-emptive censorship to avoid sanctions. Nevertheless, the legislator's intentions are to detect violent content uploaded on the web in inappropriate contexts.[146]

[46] A few weeks after the entry into force of the amendment to the Australian Criminal Law, the House Committee to the Judiciary sat in a hearing to explore and understand whether there were grounds for a series of more restrictive legislative steps against sites and platforms with extremist and nationalist connotations on the United States soil. This, of course, in the knowledge that the outlook for change and, above all, for a ban that has hitherto been guaranteed by the First Amendment is unlikely at the moment.[147] But supporters of a now-projected change in the constitutional interpretation of freedom of expression have multiplied. In fact, other episodes of mass shootings in Pittsburgh (October 27, 2010) and El Paso (August 3, 2019), which have once again fueled the debate on whether to ban controversial sites, are not long overdue.[148]

3.3. Participative platforms as «gatekeepers» and «good samaritans»

3.3.1. An overview

[47] Contents of hatred have been the commonplace since the dawn of the Internet: first on discussion forums, then in comments under all sorts of news or blogs,[149] and again on participative platforms and now also through messaging apps. Often these expressions are formulated

142 Damien Cave, Australia Passes Law to Punish Social Media Companies for Violent Posts, in: NYT April 3, 2019, https://www.nytimes.com/2019/04/03/world/australia/social-media-law.html (visited on April 10, 2019).

143 Cave, Ibid.

144 Wagner, p. 108 ff.

145 Thomas J. Holt, Regulating Cybercrime through Law Enforcement and Industry Mechanism, in: Annals of The American Academy of Political & Social Science 679, Philadelphia (USA) 2018, p. 145 ff.

146 Cave, Ibid.

147 US Committee on the Judiciary, April 9, 2019, https://judiciary.house.gov/news/press-releases/chairman-nadler-opening-statement-hearing-hate-crimes-and-rise-white-nationalism (visited on April 10, 2019).

148 Wong, 8chan: the far-right website linked to the rise in hate crimes, in: The Guardian August 5, 2019, https://www.theguardian.com/technology/2019/aug/04/mass-shootings-el-paso-texas-dayton-ohio-8chan-far-right-website (visited on December 29, 2019); see also List of Mass Shooting in the United States, https://en.wikipedia.org/wiki/List_of_mass_shootings_in_the_United_States_in_2019 (visited on January 4, 2020).

149 See case *MTE & Index.hu vs. Hungary*, Judgement of the ECtHR (22947/13) of February 2, 2016.

anonymously or under pseudonyms.[150] When offenders use a platform to disseminate their hate content, they not only harm the suitable victims, but also violate the terms of use of that platform and, as seen above and depending on the jurisdiction, they also become liable to criminal sanctions.[151]

[48] The platforms, due to their immunity as media, have no legal obligation to intervene spontaneously against the publication and dissemination of content of hatred, but might be required to remove content deemed abusive, usually when requested by authorities.[152] However, if there is a willingness to contain behaviors that in some ways are considered problematic, the legislative and policy machine of states does not delay in setting in motion, so that also participative platforms have to comply and activate continuous filtering mechanism on their contents.[153]

[49] By the end of the 1990s, many websites had to take effective direct measures to prevent publication or access in particular to pornographic content, content unsuitable for minors and content subject to copyright.[154] In the United States, obscene material can be regulated without contradicting constitutional principles because, according to the Supreme Court, it does not evoke any social value and its offensiveness substantiates a restriction, so that a website has a duty to remove any contents not suited for minors.[155] Since 2003 also Switzerland began a fight against harsh pornography and child pornography and exploitation online, prohibiting any form of diffusion of such contents.[156]

[50] The U.S. Digital Millennium Copyright Act of 1998 (DMCA) had introduced the concept of «safe harbors» for service providers and a series of articles to regulate copyright in the light of the then emerging technologies, including a prohibition to circumvent the protections applied to protect files subject to copyright (Section 1201).[157] Therefore, allowing users to share any content, images, music, videos, music, software or electronic games between themselves is considered copyright infringement.[158] Switzerland recently enacted the new Art. 39d Federal Act on Copyright Law that enforces participative platforms to ensure that copyright infringing materials, once removed, can't be made available again on their infrastructures.[159]

[51] Further restrictions have been enacted in Europe with the introduction of the General Data Protection Regulation (GDPR): the right to privacy, which is very much felt in the «old continent», has also in some way forced giants such as Facebook or Google to take measures and

150 Keats Citron, p. 2.

151 Francey, N 321–324; Selman/simmler, p. 263 f.; see also Bob Faris et al., Understanding Harmful Speech Online, in: Berkman Klein Center for Internet & Society Research Publications, Cambridge (USA) 2016, https://cyber.harvard.edu/publications/2016/UnderstandingHarmfulSpeech (visited on September 1, 2019), p. 8.

152 Equey, N 89 ff.; Freitag, p. 349; BSK-StGB-Zeller, Art. 28, N 11 f.; cfr. BGE 121 IV 109.

153 Mira Burri, Controlling new media (without the law), in: Swiss National Centre of Competence in Research 71, Bern (CH) 2011, p. 327 ff.; Klonick, p. 1602 ff.

154 Francey, N 84; Grimmelmann, p. 178 ff. (citing, among others, *Reno v. American Civil Liberties Union*, 512 U.S. 844 [1995]); see also European Commission (EC), Tackling Illegal Content Online, in: Communication from the Commission of the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions, September 28, 2017, COM(2017) 555, p. 3.

155 Grimmelmann, p 178 ff.

156 BBl 2017 6327; cfr. BBl 2001 5483 f.; see also judgement 6S.558/2001 of December 20, 2001.

157 Burri, p. 336; Clark et al., p. 12; Wagner, p. 72.

158 Giancarlo F. Frosio, *Why keep a dog and bark yourself? From intermediary liability to responsibility*, in: International Journal of Law and Information Technology 26, Oxford (UK) 2017, p. 16 f.; see also *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004 (2002).

159 Entered into force on April 1, 2020; see also BBI 2018 591.

adapt their directives to avoid sanctions and fines in the millions.[160] Restrictions, especially in the wake of the Cambridge Analytics scandal, are in a way also being introduced in the United States: it is therefore no surprise that Facebook has agreed with the Federal Trade Commission (FTC) a $5 billion fine for allowing third parties access to data and attempting to influence the polling and electoral decisions of around 50 million Americans.[161] Following the decision of the European Court of Justice (EUCJ) to repeal the Safe Harbor regime,[162] Switzerland negotiated a new treaty (the Swiss-US Privacy Shield) to improve the enforcement mechanisms of the privacy rules adopted within the borders (ex Art. 6 para. 1 Federal Act on Data Protection) by American companies.[163]

3.3.2. Gatekeeping

3.3.2.1. Content removal requested by Courts or authorities

[52] Participative platforms adapt their terms of use in accordance with new statutory requirements or case law to avoid possible negative consequences on their business operations and on users' satisfaction.[164] In general, participative platforms spontaneously act as gatekeepers, meaning that they are prone to respond punctually to court decisions or requests by authorities, if contents at issue violate their terms of use or violate some law or regulation.[165] In Switzerland, platforms are granted immunity under Art. 28 Crim. Code in relation to hate speech, but these are still called upon to intervene if a criminal activity is committed on their sites.[166] Direct involvement of the participative platforms is expected, for example, when courts request the deletion or blocking of hate content that is still accessible or visible to users.[167] However, requests addressed to platforms located in the United States may be rebounded by U.S. courts because these are based on grounds contrary to constitutional principles, which also have priority over any judicial cooperation treaties.[168] Otherwise the platforms usually follow the request.[169] Outside the judicial context, Swiss authorities have limited powers for enforcing content removal and therefore these have to expect a voluntary based cooperation from platforms. In some cases, however, special

160 Clark et al., p. 13; Grimmelmann, p. 288 ff.; see also Daphne Keller, Dolphins in the Net: Internet Content Filters and the Advocate General's Glawischnig-Piesczek v. Facebook Ireland Opinion, in: Stanford Center for Internet and Society, Stanford (USA) 2019, p. 36 ff. (on the *Glawischnig-Piesczek v. Facebook Ireland Ltd.*, C 18/18 [2019] CJEU).

161 Kang, *FTC Approves Facebook Fine of About $5 Billion*, in: NYT July 12, 2019, av. at https://www.nytimes.com/2019/07/12/technology/facebook-ftc-fine.html (visited on July 12, 2019); see also California Consumer Privacy Act of 2018 (CCPA) entered into force on January 1, 2020.

162 Case *Schrems vs. Data Protection Commissioner*, Judgment CJEU C-362/14 of October 6, 2015.

163 BBl 2017 5999; It should be noted that in *Schrems* the CJEU ruled that the EU-US Safe Harbor mechanism was inadequate to protect the transfer of data between Europe and the United States. In order to overcome the problems, including practical ones, that such a decision would have caused from a commercial point of view, a new treaty was enacted between the EU and the US, the EU-US Privacy Shield, and between Switzerland and the US, the Swiss-US Privacy Shield, https://www.privacyshield.gov/welcome (visited on October 24, 2019).

164 Mark Zuckerberg, The Internet Needs New Rules, in: The Washington Post, March 30, 2019, https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html (visited on April 10, 2019).

165 *E.g.*, Facebook Germany: Report on transparence, https://about.fb.com/de/news/2019/07/dritter-netzdg-transparenzbericht/, (visited on January 2, 2020); see also Keats Citron, p. 111 ff.

166 Francey, N 136 ff.

167 Francey, N 324; Judgement 5A_792/2011 of January 13, 2011 consid. C; see also Statement of the Federal Council 17.3277 of May 2, 2017.

168 BGE 141 IV 108; 143 IV 21; see also Burri, p. 336; Clark et al., p. 11 ff.; Francey, N 324.

169 *E.g.*, BGE 138 II 346 (case Google Street View and violation of privacy).

status is granted to law enforcement authorities («trusted flaggers») which allows them to report abusive content according to the terms of use of the participatory platform via preferential channels, which is then quickly followed up.[170] Upon receipt of the notification, the platform then takes action, whether it be the removal or blocking, or similar result, of the contested material.[171]

[53] According to the Federal Council, the voluntary cooperation model has proven its effectiveness so far, but it reserves the right to adapt the regulatory framework if circumstances change and the legal assets concerned are no longer protected.[172] Switzerland has decided not to align itself with a restrictive regulatory model such as the one introduced in Germany with NetzDG[173], which envisages penalizing providers if they do not intervene immediately in the event of a removal request (which can also be made by users).

[54] In the opinion of the Federal Council, the risk is that platforms, fearful of incurring sanctions, would drastically limit the possibility for their users to publish content and this would have consequences with regard to fundamental rights, including freedom of expression.[174] However, it considered the obligation to elect a domicile for serving notice in Switzerland to be particularly useful, so that the notification and action procedure for blocking or removing content deemed offensive would be more effective and faster.[175]

[55] The European Council tends to be less condescending to US platforms, particularly on data protection issues, but also on hate speech.[176] The EUCJ in its preliminary ruling in the *Glawischnig-Piesczek v. Facebook Ireland Ltd. case*, held that platforms such as Facebook must not only remove defamatory content, but also content considered «equivalent» and this on a global basis.[177] Similarly, the Delhi High Court in *Ramdev v. Facebook* also issued an injunction for defamatory material published in India to be removed and blocked on a global scale by certain participatory platforms, including Facebook and Twitter. An appeal has been lodged against this decision.[178]

3.3.2.2. Pre-emptive measures

[56] Preventive liability is that which requires participative platforms to be responsible for the removal, or in the case of a stay-down notification, but also to avoid not only the initial publication, but also subsequent ones. Clark et al. cite the controversial EU Copyright Directive[179], which requires that no site may display more than one snippet of content and that none of the

170 Statement of the Federal Council 19.3787 of June 20, 2019, at N 5.

171 Burri, p. 336; Clark et al., p. 12.

172 Statement of the Federal Council 19.3787 of December 1, 2017.

173 German Network Enforcement Act (Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken), as of September 1, 2017.

174 Report of the Committee on Legal Affairs 18.3306 of April 15, 2019, N 1.2.

175 Report of the Committee on Legal Affairs 18.3306 of April 15, 2019, N 1.1.

176 Clark et al., p. 12 f.; Titley et al.; p. 63.

177 Case *Glawischnig-Piesczek v. Facebook Ireland Ltd.*, Judgement CJEU C 18/18 (2019) of October 3, 2019.

178 An Analysis of Swami Ramdev v. Facebook – The Existential Risk.of Global Take Down Orders, in: SFLC.in November 15, 2019, https://sflc.in/detailed-analysis-swami-ramdev-v-facebook-judgment (visited on April 16, 2020).

179 Directive (EU) 2019/790 of the European Parliament and of the Council of April 17, 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, http://data.europa.eu/eli/dir/2019/790/oj.

users of a platform infringe copyright.[180] Indeed, doubts arise as to the effective implementation of this directive, in particular as to whether, in order to comply with Art. 15 and Art. 17 of the Directive, filters should be put in place to pre-screen what is published and, in the case of false positives, a complaint mechanism should also be provided. On the other hand, advocates of the new law believe that filters are not necessary, and would indeed be in violation of other EU laws. Controversial also remains the question that each Member State has two years to incorporate the directive into its own legislation, but the directive in this sense is only to be interpreted as a guideline and not as a default policy.[181] In relation to hate speech, this model does not appear to have been questioned, although the EU's tendency to be particularly prolific in enacting new laws may come as a surprise .[182]

[57] In Switzerland, measures of this kind in the area of hate speech are not intended to be introduced. For the Federal Council, the most effective measure so far has been voluntary cooperation.[183] However, in the area of child and youth protection, the new Art. 46a Telecommunications Act (TCA)[184] under revision stipulates that the Federal Office of Police (FedPol) may, upon simple notification, force a telecommunications service provider to remove any pornographic material ex Art. 197 para. 4 and 5 Crim. Code.[185] However, the application of preventive measures is particularly burdensome and not always effective.[186] The Federal Council also believes that the imposition of such mechanisms would be problematic from the point of view of freedom of expression, since it could lead platforms to adopt forms of systematic censorship for fear of incurring sanctions.[187]

3.3.3. Acting as good Samaritans

3.3.3.1. The interests at stake

[58] Participative platforms have every interest in offering features and tools to facilitate exchanges, interactions and understanding in a space considered accessible by anyone and anywhere in the world. Usually platforms want to give their users the best possible experience.[188] Their business model also requires users to stay connected for as long as possible and the platforms are therefore obliged to constantly develop technology that attracts and maintains users' attention: in fact, the more time they spend on the platform, the more potentially they are exposed to advertising and sponsored content.[189] As far as security is concerned, platforms try to

180 Clark et al., p. 12 f.; see also Reda, What's in the EU Copyright Directive, in: Julia Reda, February 13, 2019, https://juliareda.eu/2019/02/eu-copyright-final-text/ (visited on May 7, 2019).

181 Clark et al., p. 16; see also Keller, p. 4 ff.

182 See, *e.g.*, Holt, p. 145 ff.

183 Statement of the Federal Council 19.3787 of December 1, 2017; see also Report of the Federal Office of Communications, Legal Basis for Social Media: Update Status Report in Reply to Postulate Amherd of May, 10, 2017, p. 1 ff. (hereinafter Social Media Report 2013, p. . . .).

184 CC 784.10.

185 BBl 2017 6327; see also Francey, N 347 ff.

186 Francey, N 449 ff.

187 Statement of the Federal Council 19.3787 of December 1, 2017; Francey, N 491 ff.

188 Perset, p. 15 ff.; Talbot/Fossett.

189 Wagner, p. 44 and p. 105 ff.

preserve the security of all users in every reasonable way,[190] avoiding as much as possible abuse or harmful access to material that is presumed not to be appreciated by the majority of individuals accessing the platform, and this with a dual purpose linked on the one hand to the ideals of a free Internet, and on the other to prevent users and advertisers from fleeing for a loss of trust.[191] Platforms therefore adopt forms of control and implement methods of content removal.[192] In particular, they have created voluntary and spontaneous systems of self-regulation which, despite the immunity they are granted, characterize them as «good Samaritans» within the meaning of § 230 CDA.[193] Relying on their own values and adapting their terms of use accordingly, they strive to maintain, as far as possible, a hospitable space made available to users.[194]

[59] Some very general and summary considerations can therefore be made about platforms:

- their very existence depends on a critical mass of millions and millions of users;
- they have extensive power over data and the ability to influence users;[195]
- they present themselves as neutral, but they act with reference to the United States' legal framework (with its particularities) and with a tendentially progressive and liberal vision;[196]
- they are autonomous in deciding what, how much and whom to show content;
- thanks to treaties, laws and amendments, they are generally considered neutral vectors with respect to the content published by their users;
- the data scandal and the influence on American policy have undermined the trust placed in them by users, who are now more attentive to their data and privacy.[197]

[60] Taking the above into account, the platforms have consequently also adapted their terms of use and undertaken efforts to spontaneously counteract the abuses that occur on their sites. To a certain extent, the platforms have enacted their own self-regulation, which also aims to combat hate speech.[198]

3.3.3.2. Self-regulation

[61] Participative platforms usually adopt their own methods for content removal, often on the basis of internal policy choices and the type of vision given to the company. The basic idea under § 230 DCA is that internet intermediaries should moderate content according to their strengths and abilities without incurring sanctions or liability. The provision, in a certain sense inaccurately worded, allows platforms to determine on their own the level at which they can regulate content,

190 Keats Citron, p. 226 ff.

191 Wagner, p. 105 ff.

192 Titley et al., p. 62.

193 Klonick, pp. 1606–1608.

194 Duncombe, p. 5; Klonick, p. 1615 ff.; Wagner, p. 44, 105 ff.

195 Grimmelmann, p. 273; Keats Citron, p. 72; Wagner, p. 113.

196 Payne, Handbook, p. 3 ff.

197 Costello/Hawdon, Handbook, p. 10.

198 *E.g.*, Facebook's Community Standards, https://www.facebook.com/communitystandards/ (visited on December 30, 2019); Twitter's Help Center: Hateful conduct policy, https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy (visited on December 30, 2019).

without in fact being held responsible even if this level is not reached.[199] In these cases, the contents are removed or modified without intervention by any authority.[200] However, these changes often take place after the terms of use have been updated, and only when the content does not comply with these new rules it is then removed.[201] Twitter's decision not to allow any more politically related advertising tweets is a recent example in this sense.[202] Another example was the reorientation of Tumblr, a place known for adult content, after its acquisition by Verizon.[203]

[62] Participative platforms are not inclined to adopt a permanent and pervasive monitoring system on the content posted by individual users, except in cases provided for by law (and already mentioned).[204] The tendency is therefore not to intervene systematically and through excessively restrictive filtering systems on individual content, but rather to leave it directly to the users themselves to report, through specially designed functions, any forms of abuse.[205] The control mechanisms are however present, both with tools based on artificial intelligence, and through human moderators who, however, generally intervene only subsidiarily.[206] Basically, the aim is to avoid the application of forms of pre-emptive censorship, but also the risk of removing expressions that filters indicate as abusive, but which are falsely deemed so.[207] An example in this sense concerns Facebook's attempt to filter content labeled nudity, which, in its first iteration, sometimes considered images of breast-feeding women to be non-compliant.[208]

[63] If the report by a user or users is deemed relevant and the content is in violation of the terms of use, then the platform intervenes by blocking access or removing the content.[209]

[64] Consequences of an intervention from a participative platform range from content removal to temporary suspension of the user's account to a permanent deletion a ban from accessing the services.[210] Neither Facebook nor Twitter mention in their guidelines the possibility to report a user who publishes and spreads expressions of hatred directly to criminal prosecution. Also there is no mention whether a user's account can be restored.

199 Clark et al., p. 17 ff.

200 Christiana Fountoulakis/Julien Francey, La diligence d'un hébergeur sur Internet et la réparation du préjudice, in: Medialex, Bern (CH) 2014, p. 179 f.; BSK-StGB-Zeller, Art. 28, N 8.

201 An extensive overview of these self-regulating policies on hate speech is available at Facebook's Community Standards, https://www.facebook.com/communitystandards/hate_speech (visited on December 30, 2019).

202 See Facebook vs. Twitter on Political Ads: What Zuckerberg said, how Dorsey responded, Politico October 30, 2019, https://www.politico.com/news/2019/10/30/facebook-twitter-political-ads-062297 (visited on December 30, 2019).

203 Clark et al., p. 19.

204 Equey, N 89 ff.; Frosio, p. 16 ff.

205 Francey, N 175; cfr. Judgement of the Cantonal Court of Canton Vaud 2009/279 of March 26, 2009 consid. 6.

206 Richard Allan, Hard Questions: Who Should Decide What Is Hate Speech in an Online Global Community?, in: About Facebook, June 27, 2017, https://about.fb.com/news/2017/06/hard-questions-hate-speech/ (visited on January 3, 2020).

207 Allan, Ibid.

208 Clark et al., p. 18.

209 So also Twitter, in: Rules and policies («We will review and take action against reports of accounts targeting an individual or group of people [...], whether within Tweets or Direct Messages»).

210 *E.g.*, Enforcement options on Facebook, https://www.facebook.com/communitystandards/; Twitter: https://help.twitter.com/en/rules-and-policies/enforcement-options.

3.4. The outlook

[65] Already in 2013, in response to the Amherd postulate of September 29, 2011, the Federal Council had assigned the task to the various departments concerned with the topic of identifying comprehensive answers to the rising questions in connection with participative platforms. In the 40-page report, which covered topics such as «fake news», cyberbullying, hate speech, copyright infringement and more, it was concluded that it was not necessary to legislate specifically, but only to formulate pragmatic proposals based on existing laws and solutions already adopted in specific cases.[211]. In Switzerland, as also pointed out by the Federal Council, the fight against hate speech perpetrated on participatory platforms is primarily against the offender and not against the medium.[212] However, the involvement of the platforms is necessary in certain cases provided for by law and when authorities, through court orders, submit a request for removal or blocking of hate content.[213] So far, cooperation on a voluntary basis has been satisfactory, also taking into account issues such as extra-territoriality and extended immunity of the platforms, which is generally justified by constitutional guarantees concerning freedom of expression.[214] In Switzerland the phenomenon of hate speech remains under observation by authorities, which have expressed their willingness to adapt the legal framework to better protect the legal assets concerned.[215]

[66] In Europe and in other countries worldwide, however, there is an attempt to make the platforms more responsible, both through the promulgation of new, more incisive laws and through judicial decisions which, in some way, call for them to play a more active role in the fight against hate speech.[216] Nonetheless, the platforms lean towards a less intrusive approach, which allows them to manage the issue of abusive content through internal provisions and with the help of users, but always keeping in mind both the pursuit of business and economic goals and the more ideal aspects inherent to the principles that characterize the Internet, namely those of a free, peaceful space where everyone can express themselves.[217] It is therefore understandable that if touched on in their interests, they will challenge any attempt to limit their legitimate interests, often also referring to constitutional rights such as economic and media freedom.[218]

[67] New legislation and some court decisions have not yet had any concrete implications on how participative platforms have to adapt their actions towards the phenomenon. Nor are there any signs that there has been a slowdown in hate speech through the use of participatory platforms. The number of users using the platforms on a daily basis is constantly increasing. At the same time, so are abuses. In this scenario, platforms therefore play an important role, which calls for a certain amount of responsibility.[219]

211 Social Media Report 2013, p. 2.

212 Report of the Committee on Legal Affairs 18.3306 of April 15, 2019, N 1.2.

213 Statement of the Federal Council 19.3787 of June 20, 2019, at N 5; see also Francey, N 136 ff.

214 FDJP Cybercrime, p. 71 ff.

215 Statement of the Federal Council 17.3277 of May 2, 2017; see also Musy, p. 16.

216 Clark et al., p. 12 ff.; Titley et al., p. 63; see also Holt, p. 145 ff.

217 Perset, p. 15 ff.

218 Francey, 179 ff.

219 Musy, p. 20.

4. Conclusions

[68] The phenomenon of hate speech is complex and has increased with the advent of the Internet. There are various regulations that allow criminal prosecution and civil actions against authors who express hatred through participative platforms, whether directed and manifested against persons belonging to protected groups or against individuals.

[69] Under Swiss law, the liability of the platforms is limited to certain civil areas, in particular in cases of violation of personality, privacy or copyright, while in criminal matters it is generally excluded under Art. 28 Crim. Code. Nonetheless, the authorities must be able to counteract the effects of hate speech, not only by prosecuting the author, but also by removing the content concerned. This implies that authorities and platforms must find forms of collaboration necessary to restore the legal situation.

[70] Switzerland relies on the voluntary cooperation of platforms in removing or blocking hate content. At the same time, participating platforms themselves have adopted regulations and terms of use which, if violated, may lead to blocking or removal of abusive content. And this occurs also upon notification by other users.

[71] The role of participatory platforms in combating hate speech is therefore crucial. On the one hand they are the vector on which certain expressions are manifested, on the other hand they are also called to play a role ranging from being gatekeepers (when solicited by authorities) to being good Samaritans (when acting voluntarily). In addition to this, they find themselves in a constant conflict between freedom of expression and its restrictions. The balance, as is very often said, lies in the middle, but it must adapt according to the context and the moment.

[72] So far, the models of self-regulation and informal cooperation with the authorities, as adopted in Switzerland, seem to be effective. The system of spontaneous reporting by users, who are very close and in touch with reality, enables a series of procedures that allow platforms to react almost immediately against abusive content. In contrast, with an overtly excessive state interference, platforms could introduce more invasive forms of censorship and control mechanisms out of fear of the legal consequences in case of non compliance, thus generating a chilling effect on freedom of expression.