ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Coche, Eugénie

Article

Privatised enforcement and the right to freedom of expression in a world confronted with terrorism propaganda online

Internet Policy Review

Provided in Cooperation with:
Alexander von Humboldt Institute for Internet and Society (HIIG), Berlin

Suggested Citation: Coche, Eugénie (2018) : Privatised enforcement and the right to freedom of expression in a world confronted with terrorism propaganda online, Internet Policy Review, ISSN 2197-6775, Alexander von Humboldt Institute for Internet and Society, Berlin, Vol. 7, Iss. 4, pp. 1-17, https://doi.org/10.14763/2018.4.1382

This Version is available at:
https://hdl.handle.net/10419/214064

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

https://creativecommons.org/licenses/by/3.0/de/legalcode

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

INTERNET POLICY REVIEW
Journal on internet regulation

Volume 7 | Issue 4

Privatised enforcement and the right to freedom of expression in a world confronted with terrorism propaganda online

Eugénie Coche
Institute for Information Law (IViR), University of Amsterdam, The Netherlands

Published on 05 Nov 2018 | DOI: 10.14763/2018.4.1382

Abstract: The purpose of this paper is to explore the risks of privatised enforcement in the field of terrorism propaganda, stemming from the EU Code of conduct on countering illegal hate speech online. By shedding light on this Code, the author argues that implementation of it may undermine the rule of law and give rise to private censorship. In order to outweigh these risks, IT companies should improve their transparency, especially towards users whose content have been affected. Where automated means are used, the companies should always have in place some form of human intervention in order to contextualise posts. At the EU level, the Commission should provide IT companies with clearer guidelines regarding their liability exemption under the e-Commerce Directive. This would help prevent a race-to-the bottom where intermediaries choose to interpret and apply the most stringent national laws in order to secure at utmost their liability. The paper further articulates on the fine line that exists between 'terrorist content' and 'illegal hate speech' and the need for more detailed definitions.

Keywords: Code of conduct on countering illegal hate speech online, Privatised enforcement, Illegal hate speech, Terrorism propaganda, Freedom of expression

Article information

Received: 22 Mar 2018 **Reviewed:** 17 Aug 2018 **Published:** 05 Nov 2018
Licence: Creative Commons Attribution 3.0 Germany
Competing interests: The author has declared that no competing interests exist that have influenced the text.

URL:
http://policyreview.info/articles/analysis/privatised-enforcement-and-right-freedom-expression-world-confronted-terrorism

Citation: Coche, E. (2018). Privatised enforcement and the right to freedom of expression in a world confronted with terrorism propaganda online. *Internet Policy Review*, *7*(4). DOI: 10.14763/2018.4.1382

Acknowledgments: *I would like to thank prof. dr. Joris van Hoboken for his supervision and insightful comments during the writing of my paper. A special thank you goes to Mariana Simon Cartaya who generously proof-read this paper. Moreover, I would like to thank the peer-reviewers for their time spent on reading and evaluating this paper.*

INTRODUCTION

Terrorism is not a new issue (Ansart, 2011), but terrorism propaganda online is. As early as 2008 the EU Council officially recognised the internet as a medium used by terrorist recruiters for the dissemination of propaganda material (EU Council Framework Decision 2008/919/JHA). Several studies revealed the important role played by social media platforms, predominantly Twitter, in ISIS' [1] propaganda strategy (Badawy & Ferrara, 2017, p. 2). A 2015 report illustrated that members of ISIS, on average, posted 38 propaganda materials each day, ranging from videos to photographs or articles and on a diversity of platforms, including Facebook, Tumblr, Twitter or Surespot (Winter, 2015, p. 10). Countering this type of speech has challenged traditional law enforcement in many ways. In 2014, the EU Commission recognised that traditional law enforcement is insufficient to deal with evolving trends in radicalisation and that all of society ought to be involved in the countering of terrorism online (COM (2013) 941 final, para. 8).

On 31 May 2016, four IT companies (Facebook, Microsoft, Twitter and Youtube, 2016) adopted the EU Code of conduct against illegal hate speech online (hereinafter, the Code). This instrument places enforcement responsibilities into the hands of private companies and gives rise to the practice of 'privatised enforcement'. The dangers stemming from such practice can be illustrated by Twitter's latest biannual report (2017), in which it indicates that from July 2017 through December 2017, 274,460 accounts were suspended because of terrorism' related activities *in violation of the company's terms and services*. It also specifies on its webpage concerning removal requests that 'out of the 1,661 reports received from trusted reporters and other EU non-governmental organisations (NGOs), 19% resulted in content removal due to terms of service (TOS) violations and 10% in content being withheld in a particular country based on local law(s)'. In other words, more posts seem to have been removed because of non-compliance with the companies' policies than due to illegality. Consequently, when placing private companies at the frontline of law enforcement online, the risk may arise that our right to freedom of expression is merely guided by their terms of service, which may not always be in accordance with the level of protection guaranteed under human rights instruments, such as under Article 10 of the European Convention on Human Rights (hereinafter, ECHR) or Article 11 of the Charter of Fundamental Rights of the European Union. Moreover, taking into account the primary profit-making nature of platforms, it is questionable in how far delegation of such large-scale public functions, which are fundamental to the proper function of our democracy, may be at odd with their business objectives and thereby result in a conflict of interests. As was pointed out in an article which discussed the liability of Google when faced with removal of defamatory content: 'in order to *pursue its profit* (emphasis added), Google did not adopt precautionary measures that could have prevented the upload of illegal materials [...] Google is profiting from people uploading materials on the internet' (Sarter et al., p. 372). Taking into account the intermediaries' data-driven business model, placing them at the frontline of law enforcement may be dangerous from a legal point of view but also for democracy in general.

Whereas the privatised enforcement phenomenon has already received considerable academic attention, this paper specifically focuses on the risks stemming from the Code, in the field of illegal hate speech and, in particular, terrorism propaganda. Through identifying such risks and by taking into account subsequently adopted EU instruments, recommendations are made on how to better guarantee respect for fundamental human rights in the online environment. These findings are especially relevant as the EU Commission issued, on 12 September 2018, a proposal

for a Regulation on the prevention of terrorist content online. Besides the proposal's general requirement that hosting service providers should remove or disable access to terrorist content within one hour after receipt of a removal order, it also encourages the use of 'referrals', whose content should be assessed against the companies *own terms and conditions*. In that respect, it makes no reference to the law.

In order to draw a conclusion and make recommendations, the content of the Code and its relationship with privatised enforcement is first discussed. This section also delineates to what degree terrorism propaganda falls within the scope of the Code. Doing so is necessary, seeing as the Code merely focuses on the removal of 'illegal hate speech' whereas the countering of terrorism propaganda formed one of the main incentives for its adoption. This was made clear by EU Commissioner Vera Jourová who declared, when announcing the Code, that recent terror attacks have strengthened the need for it and that 'social media is unfortunately one of the tools that terrorist groups use to radicalise young people' (European Commission, 2016). In other words, it investigates whether and to what extent terrorist propaganda can be countered through hate speech tools. In the second section, different reasons behind privatised enforcement in the field of terrorism propaganda are presented. This is followed by a discussion on the dangers of such practice from a free speech perspective. In the subsequent section, recommendations to outweigh the identified risks are proposed, by taking into account subsequently adopted EU instruments building upon the Code, namely the communication and recommendation on tackling illegal content. The final section presents important developments that have taken place since the adoption of the Code.

PRIVATISED ENFORCEMENT THROUGH THE EU CODE OF CONDUCT ON COUNTERING ILLEGAL HATE SPEECH ONLINE

The Code is a self-regulatory initiative under which Twitter, Microsoft, YouTube and Facebook made a commitment to put in place a notice-and-take down system for the countering of illegal hate speech, the ambit of which is laid down in Framework Decision 2008/913/JHA. This non-binding instrument encourages companies to assess the legality of a post within 24 hours after being notified and to remove or block access to it in case of unlawfulness. Importantly, it explicitly stipulates that the notified posts have to be primarily reviewed against the company's rules and community guidelines and only *'where necessary'* (emphasis added) against national laws transposing the Framework Decision. Through these means, specifically encouraging the companies to 'take the lead' and initiative in tackling illegal hate speech online, the Code stimulates the occurrence of privatised enforcement.

This phenomenon was defined as a practice in which private companies undertake 'non-law based "voluntary" enforcement measures' (Council of Europe, 2014, p. 86). Legal scholars define this practice as: 'instances where private parties (voluntarily) undertake law-enforcement measures' (Angelopoulos et al., 2015, p. 6). These two definitions show that privatised enforcement has three key components: enforcement of the law; by a private party; and imposed voluntarily (in the sense that the enforcement measures flow from self-regulatory initiatives and are thus 'non-law based'). This is sometimes also referred to as 'intermediarization' (Farrand, 2013, p. 405) or 'delegated' enforcement, in the sense that the regulator's role is delegated to companies and private sector actors (ADF International, 2016, p. 1). This practice has already been encouraged in different fields of law such as copyright law (EDRi, 2014, pp. 2-14) or the

countering of 'fake news' on social media (OSCE, FOM.GAL/3/17, 2017, section 4(a)).

Whereas terrorism propaganda formed one of the main reasons for adopting the Code, such speech is not explicitly mentioned in it. The companies are merely required to counter 'illegal hate speech'. In the Commission's Communication on 'tackling illegal content online' (COM (2017), 555 final) a clear distinction is made between 'incitement to terrorism' and 'xenophobic and racist speech that publicly incites hatred and violence' (p. 2). The latter refers to the type of hate speech that is criminalised under Framework Decision 2008/913/JHA and which serves as legal basis for content removal under the Code. Concerning incitement to terrorism, the Communication refers to Article 5 of the Terrorism Directive (EU Directive 2017/541), which covers the 'public provocation to commit a terrorist offence'. Bearing this in mind, how can the Code thus contribute to the countering of terrorism propaganda?

An important distinction to be drawn between 'incitement to terrorism' and 'illegal hate speech' is that the former only covers incitement to violence (See Article 3(1), point (a) to (i) of the Terrorism Directive) while the latter also extends to incitement to hatred. The relation between these two was made clear by Vera Jourová who stated, in the context of terrorism propaganda, that 'there is growing evidence that online incitement to hatred leads to violence offline' (European Commission, 2015). In this respect, it is important to highlight that the United Nations General Assembly (2013) has determined that 'the likelihood for harm to occur' is a factor that should be taken into account when assessing whether incitement to hatred is present (para. 29). Although 'incitement' is by definition an inchoate crime, there is thus an implicit assumption that the speech has a reasonable probability to incite the intended actions and thereby cause harm. In the *Surek v. Turkey* case, this implicit relation between incitement to hatred, on the one hand, and actions, on the other, was made clear by the European Court of Human Rights (hereinafter, ECtHR) which noted that the speech was 'capable of inciting to further violence by instilling a deep-seated and irrational hatred' (§62).

In the context of terrorism, the Commission claimed, in June 2017, that 'countering illegal hate speech online' serves to counter radicalisation (COM (2017), 354 final, p. 3). The link between radicalisation through hate speech and terrorist acts was also made explicit by Julian King, Commissioner for the Security Union who declared that: 'there is a direct link between recent attacks in Europe and the online material used by terrorist groups like Da'esh to radicalise the vulnerable and to sow fear and division in our communities' (European Commission, 2017). This overlap between incitement to hatred and incitement to terrorism may be explained by the fact that terrorism relies on extremist ideologies. These were identified by Europol (2013) to include religious, ethno-nationalist and separatist ideologies as well as left-wing and anarchistic ones (pp. 16-30).

However, it is relevant to highlight the *Leroy v. France* case, which illustrates how the Code, and thereby illegal hate speech, would fall short in countering all types of terrorism propaganda. In this case, a cartoonist was accused of glorification to terrorism after having published, on the day of the 9/11 terrorist attacks, a drawing representing the American Twin Towers. The drawing was interpreted by the Court (§42) as a call for violence to and glorification of terrorism but was not perceived as a reflection of the cartoonist's anti-American ideologies. This type of speech, in which the underlying extremist ideologies are implicit within the speech – and therefore 'hidden'– will not easily be caught under the Code. Indeed, for 'illegal hate speech' to be present, some kind of discrimination must be expressed (Article 1(a) Framework Decision 2008/913/JHA). Such a discriminatory element is however not required for 'incitement to terrorism' as defined under the Terrorism Directive.

In light of the above, it can be inferred that incitement to terrorism and illegal hate speech complement each other in the fight against terrorism propaganda online. However, for removal of less obvious terrorism propaganda, where no discriminatory element or incitement to violence is present, new instruments should see the light of day. The recently proposed Regulation (COM(2018), 640 final) which adopts a very broad definition of 'terrorist content' extending beyond 'incitement to terrorism', may be one of these. Having regard to the complexity of these legal definitions, which carries the risk of misinterpretation by non-legal persons, it is important to find out what the impetus is for involving internet intermediaries in the countering of such type of speech.

DIFFERENT REASONS FOR PRIVATISED ENFORCEMENT IN THE FIELD OF TERRORISM PROPAGANDA

In 2015, the Commission highlighted, in a proposal for a directive, the importance of internet intermediaries in the fight against terrorism propaganda (COM(2015)625, para. 2). The instrument stressed the importance of the Internet as 'primary channel used by terrorists to disseminate propaganda, issue public threats, glorify horrendous terrorist acts such as beheadings, and claim responsibility for attacks' (para. 10). Consequently, without yet imposing any obligations on the part of Internet intermediaries, this proposal raised awareness on the prominent use of social media for terrorist purposes and on the need to 'tackle the evolving terrorist threats in a more effective way' (para. 14). It can thus be said that this proposal framed the path for heightened scrutiny with regards to the Internet intermediaries' role in the context of terrorism. On 15 March 2017, EU Directive 2017/541 was adopted under which internet intermediaries were encouraged to develop voluntary actions for the countering of terrorist content on their services (recital 22).

When it comes to the countering of illegal content online, the Commission has emphasised the favourable position of internet intermediaries. In its recent proposal for a regulation concerned with the online removal of terrorist content, it indicated their 'central role' in the dissemination of such material as well as their 'technological means and capabilities' justifying their 'particular societal responsibilities' (recital 3). Placing internet intermediaries at the frontline of law enforcement online was thus by no means a coincidence. Indeed, as opposed to public authorities, intermediaries have better technological means at their disposal to swiftly notice illegal content, identify infringing authors and, subsequently, block or remove allegedly illegal material. Moreover, taking into account the speed at which terrorist content is disseminated across online services, it seems primordial to involve the parties that are most prone to react quickly.

FREEDOM OF EXPRESSION RISKS STEMMING FROM PRIVATISED ENFORCEMENT

Whilst it was practical to involve internet intermediaries in the counter-terrorism process, to have them enforce the law online constitutes a real potential danger for their users' right to freedom of expression. Importantly, as was made clear in *Jersild v. Denmark* (§30), the right to freedom of expression is twofold in the sense that it does not only protect individuals' right to impart information but also the public's right to receive such information. As repeatedly held by

the ECtHR, '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-fulfillment' (*Hertel v. Switzerland, § 46; Animal Defenders International v. The United Kingdom, § 100*). This was also recognised at the international level by the Human Rights Committee (2011, para. 2). Importantly, the right is very broad in scope and also applies to ideas that 'offend, shock or disturb any sector of the population' (*Handyside v. UK*, § 49). Whereas the right is subject to limitations, such limitations are strict as these must meet different requirements, which will be discussed below, in order to be permissible. Taking into account the broad nature of this right, on the one hand, and the strict limitations, on the other, it is argued that the Code gives rise to the risk that the rule of law is undermined and that private censorship may arise.

A CHALLENGE TO THE RULE OF LAW

As specified by the Code, removal of a post shall be primarily based on the company's terms of service and only secondarily and *when necessary*, on national law. In an issue paper by the Council of Europe (2014) it was warned that such a practice would give rise to the risk that 'general terms and conditions of private-sector entities are not in accordance with international human rights standards' and therefore that the rule of law is threatened (p. 14 and 87). In that same paper, the rule of law was described as 'a principle of governance by which all persons, institutions and entities, public and private, including the state itself, are accountable to laws that are publicly promulgated, equally enforced, independently adjudicated and consistent with international human rights norms and standards' (p. 10). Concerns about the rule of law not being respected were also expressed by Vera Jourová who, when discussing the Code, stressed that: 'the rule of law applies online just as much as offline. We cannot accept a digital Wild West *[…]* If the tech companies don't deliver, we will do it' (European Commission, September 2017).

The ECtHR has developed a test for the rule of law, which requires that any interference on a person's right to freedom of expression must be based on a proper legal basis. This legal basis must be sufficiently precise, accessible to the public and provide sufficient safeguards (*Kruslin v. France*, §27-36). The interference must also serve one of the legitimate aims under Article 10(2) ECHR and be necessary in a democratic society (S*unday Times v. UK*, §45). In a fact-sheet concerning the implementation of the Code, the Commission specified that the Council Framework Decision 2008/913/JHA should form the legal basis for removal of illegal hate speech (2016, p. 3). However, by explicitly encouraging IT companies to prohibit 'hateful conduct' in their Community Guidelines, the Code (para.10) encourages them to go further than what is prescribed under the Decision. As mentioned previously, the right to freedom of expression also extends to shocking or offending ideas (*Handyside v. UK*, § 49).

Under Facebook's terms and conditions, illegal hate speech is phrased as content that amounts to a 'direct attack based on what we call protected characteristics - race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability'. The threshold for Facebook to remove a post appears to be lower than what is required under the Framework Decision as there is no requirement for *incitement* to be present. Unlike Facebook, Twitter and YouTube do emphasise in their community guidelines that where the primary purpose of an account is *incitement* to harm, based on discriminatory grounds, the account will be deleted. Concerning Microsoft, it advises its users not to 'incite other users to threaten, stalk, insult, victimise, or intimidate another person or group of people'. Here, no reference is made to the discriminatory nature of the incitement. It is thus clear that the different conditions for 'illegal hate speech', required under EU law, are not always reflected in the companies' policies. A post that would not necessarily amount to any criminal offence under EU law may thus still be removed because of non-compliance with the terms of service.

This disparity between law and terms of use is well illustrated by a letter of the German Ministry of Justice and consumer protection (*Bundesministerium der Justiz und für Verbraucherschutz*, 2016) written in response to a request of the German Parliament who wished to obtain information on how many of the 100,000 contents deleted by Facebook were actually illegal under German Federal law. The answer was that it was unknown. As is explicitly stated in the letter: 'there is no examination whether concrete individual cases of hate messages are illegal'. Consequently, the danger exists that the law may be downgraded to terms of service.

The risk for IT companies to incorrectly interpret and enforce illegal hate speech is even more emphasised when taking into account that under EU law, different factors such as the intent of the speaker, the likelihood for harm to occur and the context of the speech must be considered (*Surek v Turkey, §62; Gokceli v Turkey, § 38)*. Although in theory the European Commission (2017) has specified that such factors shall also be taken into account by the IT companies in their assessment of illegal hate speech, no reporting activities have yet taken place in which is demonstrated that such elements play a role in their assessment. The only way through which can be inferred that the companies do take these factors into account is by taking a look at their community guidelines. However, when reading those, Twitter merely seems to take into account the 'context of the larger conversation' and Facebook fails to mention 'the likelihood for harm to occur' or require 'incitement' to be present (Allan, 2017).

PRIVATE CENSORSHIP

Another major risk on our right to freedom of expression is that the privatised enforcement system encouraged under the Code would lead to private censorship. The UN Special Rapporteur on freedom of expression defined 'private censorship' as meaning that 'censorship measures are delegated to private entities' which includes situations in which intermediaries undertake censorship on behalf of the state (United Nations General Assembly, 2011, A/HRC/17/27, para. 45 jo. 75). This risk, which is intrinsically related to the notice-and-take down system as supported by the Code, must be seen in light of the e-commerce Directive which has, *inter alia*, created an exemption regime to the liability of hosting service providers **(**Directive 2000/31/EC). As pointed out by legal scholars (Sartor et al., 2010), this legal construction may lead to internet intermediaries becoming the gatekeepers of the internet as it 'presupposes authorising the provider to exercise the controls that may prevent its liability, i.e., empowering it to exclude all those contents that may generate liability' (p. 376).

Indeed, according to Article 14 (jo. recital 46) of this Directive, hosting providers may be exempted from liability when they 'expeditiously remove or disable access to illegal content' after having been notified of such content's presence. As was argued by legal scholar Aleksandra Kuczerawy (2015), such a mechanism implies a conflict of interests for the intermediary. To put this in her own words: 'they [the internet intermediaries] have to decide swiftly about removing or blocking content in order to exonerate themselves from possible liability, which basically makes them a judge in their own cause' (p. 48). Consequently, as was pointed out by her, they will have the incentive to be over-protective and to remove or disable access to content regardless of their illegality and, sometimes, even without carrying out a balancing of interest. This may in turn result in users' right to freedom of expression being impeded as 'any potential controversial information would then likely be prevented from reaching public accessibility' (Sartor et al., 2010, pp. 376-377).

During a public consultation on the e-Commerce Directive, the majority of stakeholders (including internet intermediaries) were of the opinion that over-removal of content is partly due to legal uncertainties surrounding the scope and terms of the liability exemption (European

Commission, 2012, SEC (2011)1641, pp. 43-46). As was argued by Lisl Brunner (2016), former policy director on matters of intermediary liability, such legal uncertainty has increased after the ECtHR ruling in the *Delfi* case as the Court did not clarify the fine line that exists between service providers which are of an active or passive nature. This distinction is of importance since the European Court of Justice (CJEU) has repeatedly confirmed that the liability exemption established under the e-Commerce Directive can only be enjoyed by hosting providers which are of a mere technical, automatic and passive nature (L'oreal SA and others v. Ebay International A.G. and others, paras. 111-116; Google France SRL and Others v. Louis Vuitton Malletier SA and others, paras. 114 and 120). Whereas the *Delfi* case concerned the alleged infringement of a publishers' right to freedom of expression under Article 10 ECHR, the cases decided by the CJEU related to the interpretation of EU law in the course of preliminary ruling procedures. However, both courts were in these cases confronted with liability issues and the interpretation of Article 14 of the e-Commerce Directive. Taking into account aforementioned risks, it is necessary to find out how these could at most be outweighed.

DIFFERENT WAYS TO BALANCE THE DANGERS OF PRIVATIZED ENFORCEMENT ON THE RIGHT TO FREEDOM OF EXPRESSION

One way to counterbalance the issue of overly broad terms of service through which the rule of law may be threatened would be to provide legal safeguards to end users. In this regard, legal scholars (Angelopoulos et al., 2015) stressed, in a study concerned with privatised enforcement and human rights limitations, the importance of IT companies to be transparent and accountable and to take into account due process principles (p. 57). This idea was also supported by the Commission in its communication on 'tackling illegal content online' and subsequent recommendation (COM(2017), 555 final, p. 14; C(2018), 1117 final, Chapter II (16-17) jo. preamble pt. 20).

Whilst the Code states that it promotes transparency, it only does so by encouraging publication of transparency reports. In the two latest periodical reviews, no attention was paid to the existence of transparency measures towards end users whose post had been notified and/or removed (European Commission, 2017; 2018). The main focus was whether the companies had provided feedback to *notifying* users. Whilst the Commission did stress, in its communication, the importance of transparency reports, it also stressed the importance of being transparent towards users whose post had been notified and that information shall be provided about received counter-notices (COM(2017), 555 final, p. 16). Intrinsically related to this point and as was put forward by Kuczerawy (2015, p. 51), the companies should have in place a system of counter-notices. This would help uphold due process principles in notice-and-actions procedures. The need for this was further supported by the Commission (COM(2017), 555 final, p. 17; C(2018)1117, Chapter II(13))

Another way to secure respect for the rule of law online would be through the States' positive obligations. The ECtHR has recognised that states play an important role in protecting the right to freedom of expression, which includes both negative obligations (to abstain from interfering with that right) and positive ones (to take action) (*Özgür Gündem v. Turkey*, para. 43; *Centro Europa 7 S.R.L. and Di Stefano v. Italy*; *Youth Initiative for Human Rights v. Serbia; Dink v. Turkey*, para. 137). The Council of Europe (2014) has already suggested that 'states have an obligation to ensure that general terms and conditions of private companies that are not in

accordance with international human rights standards must be held null and void' (p. 114). Legal scholars also supported this idea and stressed that 'States may be found to be in breach of their positive obligations for their failure to prevent violations of individuals' fundamental rights as a result of privatized law enforcement by online intermediaries' (Angelopoulos et al., 2015, p. 79). However, as these scholars mentioned, different criteria must be taken into account when establishing whether a breach of a state's positive obligation occurred (p. 79). Whilst analysis of such a breach goes beyond the scope of this paper (since it would require a case-by-case analysis), relying on state's positive obligations could help to foster the rule of law in the online environment. However, as was concluded in their study, discussions should find place in order to 'operationalize relevant positive obligations of States in the context of self-regulatory or privatised law enforcement measures by online intermediaries' (p. 79).

Concerning the countering of private censorship, IT companies should have more legal certainty about their liability exemption provided for under the e-Commerce Directive. This was encouraged by Kuczerawy (2015, p. 46) who claimed that legal uncertainties exist with regards to Article 14 of the Directive, such as the scope of the term 'service providers', the meaning of 'actual knowledge' or the term 'expeditiously' (pp. 50-51). As was made clear in a working paper of the European Commission (SEC (2011), 1641 final) the rules for notice-and-take-down procedures vary from one member state to another, making it unclear for internet intermediaries as to which rules should be followed (p. 25). Such fragmentation could result in a race to the bottom where intermediaries choose to interpret the rules of the countries with the most stringent laws in order to secure at utmost their liability exemption. Indeed, when taking into account that internet intermediaries could potentially be subject to the laws of all countries in which their content is accessible, the safest way for them to act would be to take a restrictive approach and treat the harshest laws as threshold for content removal. In other words, by 'lowering the standards of free speech on the internet to the lowest common regulatory denominator' (Mills, 2015, p. 19)

In 2012, the Commission announced an initiative on 'Notice-and-Action' procedures aimed at harmonising, at EU level, the rules on these procedures (COM(2011), 942 final, p. 15). From all the different parties involved in the public consultation, most of them supported the idea that the EU should clarify the functioning of notice-and-action-procedures and thereby adopt binding minimum rules (European Commission, 2012, p. 8). Such an initiative did, however, not yet lead to an EU binding legal instrument. In May 2017, members of the European Parliament expressed, in an open letter to the Commission's Vice-President, their wish for a notice-and-action directive. According to them, having in place an EU framework on notice-and-actions procedures would help to counter the issue that 'large internet platforms are independently taking their own actions to take down online content, without transparency or independent scrutiny' (Schaake, 2017).

Recently, in its communication and subsequent recommendation on tackling illegal content online, the Commission tried to clarify some vague aspects of the liability exemption for hosting providers (COM(2017), 555 final, pp. 13-14; (C(2018), 1117 final, preamble pt. 26). Concerning the term 'expeditiously', the Commission takes a flexible approach by specifying that the term must be analysed on a case-by-case basis. With respect to the risk of a race-to-the-bottom, the Commission's recommendation clarifies that the laws to be taken into account are those from the member state in which the hosting provider is established or these where the services are provided (C(2018), 1117 final, preamble pt. 14).

Another possible way to achieve a higher level of legal certainty would, yet again, be through

positive state obligations. Importantly, the ECtHR established in *Dink v. Turkey* (para. 137) that one of these obligations consists in ensuring that individuals can express themselves without fear. In light of this, legal scholars have held that such a positive obligation could include the duty to reduce internet intermediaries' fear of being held liable, which would be a 'promotional obligation' (Angelopoulos et al., 2015, pp. 32; 42).

DEVELOPMENTS SINCE THE ADOPTION OF THE EU CODE OF CONDUCT ON COUNTERING ILLEGAL HATE SPEECH ONLINE

Since implementation of the Code, the IT companies have been put under serious pressure to better counter online terrorism propaganda. After the terrorist attacks committed in Brussels, London and Manchester, the Commission (COM(2017), 354 final) declared that the Code had helped in the counter-radicalisation process but that it was insufficient (p. 3). In light of this lack of effectiveness, different measures have seen the light of day.

On the one hand, the EU has issued non-binding instruments such as the communication and the recommendation on tackling illegal content online and, on the other, it is in the process of adopting legislation aimed at tackling terrorist content online.

Regarding the communication and subsequently adopted recommendation, several actors have criticised these. Concerning the former, the European Federation of Journalists (2017) pointed out to a lack of guidance to platforms for respect of the right to freedom of expression. According to Jens-Henrik Jeppesen, representative and director for European Affairs, the communication 'describes a regime of privatised law enforcement that does not attempt to draw a bright line between content that violates platforms' terms of service (TOS) and content that breaks the law' (Jeppesen, 2017). Furthermore, Marietje Schaake, member of the European Parliament, warned that 'the good parts on enhancing transparency and accountability for the removal of illegal content are completely overshadowed by the parts that encourage automated measures by online platforms' (Schaake, 2017). It should be borne in mind that technological means are not (yet) able to contextualise posts, whereas 'context of content' is a factor that needs to be taken into account when assessing illegal hate speech. However, on that point, the recommendation seems to provide better safeguards as it suggests that human oversight and verification should be provided where there is no 'human in the loop' (when automated means are used). Despite the better safeguards, the recommendation still seems to magnify the risks of privatised enforcement. With regard to terrorist content, it states that Europol and the competent authorities shall request removal either 'by reference to the relevant applicable laws *or* (emphasis added) to the terms of service of the hosting service provider concerned' (preamble pt. 34). Furthermore, the wording suggests that companies have discretion as to whether or not to remove terrorist content after having been notified by the member states' competent authorities (Chapter III (34)). As opposed to the Code which permits removal within 24 hours, the recommendation adopts a one-hour removal timeframe (Chapter III (33)). As was argued by Emma Llansó (2018), director of the 'Free Expression Project' at the Washington-based Center for Democracy & Technology, the recommendation places too much focus on the speed of removals and the need for automatic filtering technologies instead of on safeguards for human rights. Moreover, speedy decision may impact due process norms such as the right to be heard, protected under Article 6 ECHR.

Furthermore, by clarifying that 'terrorist content' is not limited to the offences listed in the Terrorism Directive but may also extend to 'content produced by or attributable to terrorist groups' (Chapter I(4(h)), the Commission makes it rather unclear what type of content is targeted and subject to the one-hour removal. This vagueness is emphasised given the fine line that exists between 'terrorist content' and illegal hate speech which, in some cases, contributes to radicalisation. Clarification on what type of speech is entailed in 'radicalization' would thus be helpful, especially when taking into account that Julian King, Commissioner for the Security Union, identified radicalisation as being the core problem of terrorism (Commission, 2018).

Last but not least, by taking a flexible and case-by-case approach, both instruments do not seem to clarify the liability exemptions contained in the e-Commerce Directive. In December 2017, different Members of the European Parliament urged the Commission to 'take up the specific issue of notice and action procedures as a priority, independent from its work on addressing illegal content online' (Schaake et al., 2017).

Concerning legislative measures, the EU Commission issued, on 12 September 2018, a proposal for a regulation on the prevention of terrorist content online COM(2018), 640 final). Like the recommendation, it adopts a one-hour removal timeframe and specifies that referrals should be assessed *against the companies' terms of service*. Moreover, it adopts an even broader definition of 'terrorist content' than under the recommendation and threatens non-compliant hosting service providers with penalties (Article 18). It also adds further confusion to the liability exemption under the e-Commerce Directive as it states that derogation to the general prohibition to monitor under Article 15 of it may exceptionally arise (explanatory memorandum, p. 3).

A new Audiovisual Media Services Directive is also being adopted which would make private companies accountable for having hate speech videos or videos inciting to terrorism present on their services (COM(2016), 287 final). This general trend to hold intermediaries accountable for illegal content has emerged in different fields of law, such as in intellectual property law with the proposed Copyright Directive (Com (2016), 593 final, Article 13). This regulatory tendency can also be seen at national level. For example, Germany adopted the so-called 'network enforcement law', which threatens social media companies with fines of up to 50 million in case of non-removal of illegal content within a certain time. [2] In a legal review of this (draft) law, commissioned by the Office of the OSCE Representative on Freedom of the Media, Bernd Holznagel (2017) warned that: 'with the risk of high fines in mind, the networks will probably be more inclined to delete a post than to expose themselves to the risk of a penalty payment' (p. 23). He also noted that such regulations may encourage platforms to circumvent the laws' territorial scope through removal of German-language comments (24).

CONCLUSION

The present paper has aimed to demonstrate how the Code has clear implications on internet users' right to freedom of expression. The privatised enforcement system encouraged under it could result in private censorship as well as undermining of the rule of law.

By taking into account different developments since the adoption of the Code, this paper claims that, from an EU-perspective, a shift from the focus on 'speed' to 'legality' should take place. Whereas the Code adopted a 24-hour framework for removal of illegal content, the recommendation on tackling illegal content online and the recently proposed regulation

(COM(2018) 641 final) encourages removal of terrorist content within one hour. Such short time frame, paired with the unclear definition attributed to 'terrorist content', will undoubtedly magnify the risks of over-removal of content. Moreover, the EU should clarify the liability exemption under the e-Commerce Directive by giving clear guidance on what the terms contained therein entail. This would help prevent a race-to-the bottom where intermediaries choose to interpret and apply the most stringent national laws in order to secure at utmost their liability. Concerning the IT companies, these should increase their level of transparency when removing posts. Unlike what the recommendation encourages, more efforts should be made in terms of transparency towards the users whose posts have been notified. IT companies should always provide counter-notices and provide feedback. Human intervention should also be a *conditio sine qua non* in cases where there is no human in the loop and thus not only 'where appropriate' as stipulated in the recommendation.

Despite the shortcomings in the recently adopted EU instruments, these illustrate that some attention is being paid at the EU level for the protection of human rights in the digital environment. However, having regard to the recent proposal for a regulation on the prevention of terrorist content online, more attention is still needed in order to reconcile the practice of privatised enforcement with respect for individuals' fundamental human rights.

REFERENCES

ADF International. (2016). *Response to call for submissions by the UN Special Rapporteur on the Protection of the Right to freedom of Opinion and Expression*. Retrieved from www.ohchr.org/Documents/Issues/Expression/Telecommunications/ADF.docx

Allan, R. (2017). Hard Questions: Who Should Decide What is Hate Speech in an Online Global Community?. Retrieved from https://newsroom.fb.com/news/2017/06/hard-questions-hate-speech/

Angelopoulos, C., Brody, A., Hins, W., Hugenholtz, B., Leerssen, P., Margoni, T., McGonagle, T., van Daalen, O., & van Hoboken, J. (2015). Study of fundamental rights limitations for online enforcement through self-regulation. Amsterdam: Institute for Information Law IViR. Retrieved from https://www.ivir.nl/publicaties/download/1796

Ansart, G. (2011). *The invention of Modern State Terrorism during the French Revolution*. Retrieved from https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1031&context=revisioning

Animal Defenders International v. The United Kingdom (App no 48876/08) ECHR, 2013.

Badawy, A., & Ferrara, E. (2018). The Rise of Jihadist Propaganda on Social Networks. *Journal of Computational Social Science*, *1*(2), 453-470. doi:10.1007/s42001-018-0015-z Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2982256

Bundesministerium der Justiz und für Verbraucherschutz. (2016). *Ihre schriftliche Frage Nr. 10/19 vom 6. Oktober 2016*. Retrieved from http://andrej-hunko.de/start/download/doc_download/863-schriftliche-frage-zur-groessenordnung-der-100-000-von-facebook-geloeschten-internetinhalte

Brunner, L. (2016). The liability of an Online intermediary for Third-party content: The watchdog becomes the monitor: intermediary liability after Delfi v Estonia. *Human Rights Law Review*, *16*(1), 163-174. doi:10.1093/hrlr/ngv048

Centro Europa 7 S.R.L. and Di Stefano v. Italy (App no 38433/09) ECHR, 2012.

Council of Europe. (2014). *The rule of law on the internet and in the wider digital world* (Issue Paper). Strasbourg: Council of Europe.

Delfi AS v Estonia (App no 64569/09) ECHR, 2015.

Dink v. Turkey (App nos 2668/07, 6102/08, 30079/08, 7072/09 and 7124/09) ECHR, 2010.

EDRi. (2014). *Human Rights and Privatised law enforcement*. Retrieved from https://edri.org/wp-content/uploads/2014/02/EDRi_HumanRights_and_PrivLaw_web.pdf

EU Council Framework Decision 2008/919/JHA of 28 November 2008 Amending Framework Decision 2002/475/JHA on combatting terrorism.

EU 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.

EU Directive 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA.

European Commission. (2012). Commission Staff Working Document, online services, including e-commerce, in the Single Market Accompanying the document: Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions, A Coherent framework to boost confidence in the Digital Single Market of e-Commerce and other online services. SEC (2011), 1641 final.

European Commission. (2012). Commission Communication to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions. A coherent framework for building trust in the Digital Single Market for e-commerce and online services. COM (2011), 942 final.

European Commission. (2012). *Summary of the results of the Public Consultation on the future of electronic commerce in the Internal Market and the implementation of the Directive on electronic commerce (2000/31/EC)*. Retrieved from http://ec.europa.eu/information_society/newsroom/image/document/2017-4/consultation_summary_report_en_2010_42070.pdf

European Commission. (2014). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Preventing Radicalisation to Terrorism and Violent Extremism: Strenghtening the EU's response. COM (2013), 941 final.

European Commission. (2015). Proposal for a Directive of the European Parliament and of the Council on combatting Terrorism and replacing Council Framework Decision 2002/475/JHA on combatting terrorism. COM (2015), 625 final.

European Commission. (2015). EU Internet Forum: Bringing together governments, Europol and technology companies to counter terrorist content and hate speech online. Retrieved from http://europa.eu/rapid/press-release_IP-15-6243_en.htm

European Commission. (2016). Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market. Com (2016), 593 final.

European Commission. (2016). European Commission and IT Companies announce Code of Conduct on illegal online hate speech. Retrieved from http://europa.eu/rapid/press-release_IP-16-1937_en.htm

European Commission. (2016). *Code of Conduct – Illegal online hate speech Questions and Answers*. Retrieved from http://ec.europa.eu/newsroom/document.cfm?doc_id=41844

European Commission. (2016). Proposal for a Directive of the European Parliament and of the Council amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of Audiovisual Media Services in view of changing market realities. COM (2016), 287 final.

European Commission. (2017). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. tackling illegal content online, towards an enhanced responsibility of online

platforms. (COM (2017), 555 final.

European Commission. (2017). Fact Sheet Code of Conduct on countering illegal online hate speech 2nd monitoring. Retrieved from http://europa.eu/rapid/press-release_IP-17-3493_en.htm

European Commission. (2017). Security Union: Commission steps up efforts to tackle illegal content online. Retrieved from http://europa.eu/rapid/press-release_IP-17-3493_en.htm

European Commission. (2017). Fighting Terrorism Online: Internet Forum Pushes for automatic detection of terrorist propaganda. Retrieved from http://europa.eu/rapid/press-release_IP-17-5105_en.htm

European Commission. (2017). *Code of Conduct on Countering Hate Speech online: One year after*. Retrieved from http://ec.europa.eu/newsroom/document.cfm?doc_id=45032

European Commission. (2017). Communication from the Commission to the European Parliament, the European Council and the Council. Eighth progress report towards an effective and genuine Security Union. COM (2017), 354 final.

European Commission. (2018). Commission Recommendation of 1.3.2018 On measures to effectively tackle illegal content online. C (2018), 1117 final.

European Commission. (2018). Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online. COM (2018), 640 final.

European Commission. (2018). Security Union: Commission follows up on terrorist radicalisation. Retrieved from http://europa.eu/rapid/press-release_IP-18-381_en.htm

European Commission. (2018). *Code of Conduct on Countering Illegal Hate Speech online: Results of the third monitoring exercise*. Retrieved from http://ec.europa.eu/newsroom/just/document.cfm?doc_id=49286

European Federation of Journalists. (2017). The European Commission will not legislate on illegal content online. Retrieved from https://europeanjournalists.org/blog/2017/09/29/the-european-commission-will-not-legislate-on-illegal-content-online/

Europol. (2013). *TE-SAT 2013 EU Terrorism Situation and Trend Report* (Report). The Netherlands: European Union Agency for Law Enforcement Cooperation. doi:10.2813/00041

Facebook, Microsoft, Twitter and Youtube. (2016). *Code of Conduct Countering Illegal Hate Speech Online*. Retrieved from http://www.statewatch.org/news/2017/sep/eu-com-illegal-content-online-code-of-conduct.pdf

Facebook. (2018). Community Standards Hate Speech. Retrieved from https://www.facebook.com/communitystandards/hate_speech

Farrand, B. (2013). Regulatory Capitalism, Decentered enforcement and its legal Consequences for Digital Expression: The Use of Copyright law to restrict freedom of speech online. *Journal of Information Technology & Politics*, *10*(4), 404-422. doi:10.1080/19331681.2013.843922

Gokceli v Turkey (App no 27215/95 and 36194/97) ECHR, 2003.

Handyside v UK (App no 5393/72) ECHR, 1976.

Hertel v. Switzerland (App no 59/1997/843/1049) ECHR, 1998.

Holznagel, B. (2017). Legal Review of the Draft Law on Better Law Enforcement in Social Networks. Vienna: Organization for Security and Co-operation in Europe. Retrieved from https://www.osce.org/fom/333541?download=true

Human Rights Committee. (2011, September) *General Comment No. 34* (CCPR/C/GC/34).

Jeppesen, J. (2017). Tackling illegal content online: the EC continues push for privatized law enforcement. Retrieved from https://cdt.org/blog/tackling-illegal-content-online-the-ec-continues-push-for-privatised-law-enforcement/

Jersild v Denmark (App no 15890/89) ECHR 1994.

Judgment of 12 July 2011, *L'oreal SA and others v Ebay International A.G. and others*, C-324/09, EU:C:2011:474.

Judgment of 23 March 2010, *Google France SRL and Others v Louis Vuitton Malletier SA and others,* Joined cases C-236/08 to C-238/08, EU:C:2010:159.

Kruslin v France (App no 11801/85) ECHR, 1990.

Kuczerawy, A. (2015). Intermediary liability & Freedom of Expression: Recent developments in the EU Notice & Action initiative. *Computer Law & Security Review*, *31*(1), 46-56. doi: 10.1016/j.clsr.2014.11.004

Llansó, E. (2018). EC Recommendation on Tackling illegal content online doubles down on push for Privatized law enforcement. Retrieved from https://cdt.org/blog/ec-recommendation-on-tackling-illegal-content-online-doubles-down-on-push-for-privatized-law-enforcement/

Leroy v France (App no 36109/03) ECHR, 2008.

Microsoft. (2018). Microsoft Community Frequently Asked Questions. Retrieved from https://answers.microsoft.com/en-us/page/faq?auth=1#faqCodeConduct3

Mills, A. (2015). The law applicable to cross-border defamation on social media: whose laws govern free speech in 'Facebookistan'? *Journal of Media Law*, *7*(1), 1-35. doi: 10.1080/17577632.2015.105594

OSCE, United Nations. (2017, March) *Joint Declaration on Freedom of expression and "Fake News", Disinformation and Propaganda* (FOM.GAL/3/17).

Özgür Gündem v. Turkey (App No. 23144/99) ECHR, 2000.

Sartor G., Viola De Azevedo Cunha M. (2010). The Italian Google-Case; Privacy, Freedom of Speech and Responsibility of Providers for User-Generated Contents. *International Journal of Law and Information Technology*, *18*(4), *pp. 356-378*. doi: 10.1093/ijlit/eaq010

Schaake, M. (2017). Open letter – MEPs want notice and action Directive. Retrieved from https://marietjeschaake.eu/en/meps-want-notice-and-action-directive

Schaake, M. (2017). No room for upload-filters in the EU. Retrieved from https://marietjeschaake.eu/en/no-room-for-upload-filters-in-the-eu

Schaake, M et al. (2017). Letter to the Commission on notice and action procedures. Retrieved from https://marietjeschaake.eu/en/letter-to-the-commission-on-notice-and-action-procedures

Sunday Times v UK (App no 6538/74) ECHR, 1979.

Surek v Turkey (No 1) (App no 26682/95) ECHR, 1999.

Twitter. (2017). Government TOS reports – July to December 2017. Retrieved from https://transparency.twitter.com/en/gov-tos-reports.html#government-tos-reports-jul-dec-2017

Twitter. (2017). Removal requests - July to December 2017. Retrieved from https://transparency.twitter.com/en/removal-requests.html

Twitter. (2018). Hateful Conduct Policy. Retrieved from https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy

United Nations General Assembly, Human Rights Council. (2011, May) *Report of the Special Rapporteur on the Promotion and Protection of the Right to freedom of Opinion and Expression* (A/HRC/17/27).

United Nations General Assemby, Human Rights Council. (2013, January) *Annual Report of the United Nations High Commissioner for Human Rights* (A/HRC/22/17/Add.4).

Winter, C. (2015). Documenting the Virtual "Caliphate". London: Quilliam Foundation. Retrieved from http://www.quilliaminternational.com/wp-content/uploads/2015/10/FINAL-documenting-the-virtual-caliphate.pdf

Youth Initiative for Human Rights v. Serbia (App no. 48135/06) ECHR, 2013.

YouTube. (2018). Hate Speech Policy. Retrieved from https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy

FOOTNOTES

1. The Islamic State of Iraq and the Levant

2. Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz - NetzDG) (only available in German), < https://www.buzer.de/s1.htm?g=NetzDG&f=1 >; Bundesministerium der Justiz und für Verbraucherschutz, 'Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act, NetzDG) - Basic Information', < https://www.bmjv.de/DE/Themen/FokusThemen/NetzDG/_documents/NetzDG_englisch.html>