Peer Reviewed

AMSTERDAM LAW FORUM

Scientific

CENSOR THEM AT ANY COST? A SOCIAL AND LEGAL ASSESSMENT OF ENHANCED ACTION AGAINST TERRORIST CONTENT ONLINE

Naomi Klompmaker [*]

ABSTRACT

Governments, social media companies and European institutions are worried about the amount of texts, videos, images and sound files that are published and disseminated on the internet by terrorist groups and extremist individuals. Extremist and terrorist content on the internet may indeed contribute to radicalisation of home-grown terrorists and lone actors, but the relation between the internet and radicalisation is not evident. Terrorist content is increasingly removed from open social media platforms, but this hinders the possibility to provide effective counter narratives and results in the relocation of terrorists to unmonitored online environments. In 2018, the European Commission issued a proposal for regulation to prevent the dissemination of terrorist content online. This regulation introduces obligations for social media platforms to remove terrorist content. Yet, these measures are in conflict with the E-Commerce Directive and will have an impact on the freedom to conduct a business and the freedom of expression and information. It is argued that the costs of removing terrorist content in accordance with the proposed regulation are higher than the expected benefits.

Keywords: Terrorism, Extremism, Terrorist content, Radicalisation, Counter-radicalisation, Social media companies, Dark Net, E-commerce Directive, Liability, General monitoring obligations, Algorithms, Freedom to conduct a business, Freedom of expression and information

Introduction

In the last few years we witnessed many terrorist attacks in Europe, either organised by Jihadist terrorist organisations or committed by lone actors who are inspired by a Jihadist ideology. At the same time, groups that adhere to extreme right-wing ideologies are increasingly prepared to use violence to advance their political agenda. The terrorist attack on a mosque in New Zealand on 15 March 2019 is the most recent extreme-right terrorist attack and this ideology is also widespread among individuals and organised political groups in Europe. Moreover, this attack especially illustrates the role of the internet before, during and after a terrorist attack. The online behaviour of the perpetrator before the attack, the livestream of the terrorist attack and the subsequent online discussions and sharing of extreme-right views showed that the internet may function as a tool for radical individuals and terrorist perpetrators.

[*] Naomi Klompmaker is a recent graduate of the L.L.M in Transnational Legal Studies programme in the Vrije Universiteit Amsterdam. She received a BSc Criminology from the Erasmus University Rotterdam in 2016.

More than ever, we live in a globalised world that is facilitated by the internet. Traditional boundaries are fading, and we are able to obtain any kind of information, express opinions and create networks with people all over the world. It is inevitable that this creates risks for society because extremist ideologies may be formed in isolation of any physical contact and information about bomb-making can easily be found online. In addition, terrorist organisations are able to recruit new members or inspire individuals to carry out attacks. The dissemination of extremist videos, images, texts and livestreams are especially concerning with regards to the possible effects on potential violent radicals. The question is whether it is possible that the internet and online terrorist content functions as the indispensable link for individuals to commit terrorist attacks? In the last few years, European institutions and politicians are increasingly paying attention to this issue and are stressing the need for effective action to counter terrorist content online. Yet, the internet remains a difficult area to govern because technological and jurisdictional issues are obstructing governments to control and moderate online content on their own. That is why certain initiatives in Europe were introduced where private actors in the digital environment take the lead to counter illegal content on their platforms. The major social media companies Facebook, YouTube and Twitter already increased content filtering and they have adapted their removal policies with regards to content that depicts extreme violence, incites to violence or represents a jihadist ideology.

Nevertheless, terrorist content is currently still visible on both large and small social media platforms. It appears challenging to reduce the accessibility to all illegal content on the internet. It is therefore necessary to establish a more effective framework for public and private actors in the European Union to prevent that terrorists get a platform to incite to violence against others, while respecting the fundamental rights of companies and internet users and anticipating on any unintended consequence. However, in delegating responsibilities to online service providers in the area of content monitoring, European regulators and public authorities in the Member States are limited by The E-Commerce Directive.[1] That Directive provides that online service providers are in certain situations protected against liability for third-party content hosted on their services and prevents that public authorities oblige online service providers to monitor all content on their platforms.

In response to the concerns of the European politicians and institutions about terrorist content online and despite the limitations to enlarge the role of private actors, the European Commission issued in September 2018 a proposal for regulation 'to prevent the dissemination of terrorist content'[2]. This regulation introduces measures and specific obligations for internet companies and national authorities to target terrorist content on the internet. In this article I will first discuss the use of the internet by terrorist organisations and individuals with jihadist or extreme right-wing views. Then I will assess to what extent the internet and online terrorist content contributes to radicalisation and terrorism and consider the consequences of removing that content from social media platforms. Afterwards, I will make a legal assessment of the proposal for a regulation to prevent terrorist content in light of the E-Commerce Directive and the relevant fundamental rights.

[1] Directive 2000/31 of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] *OJ L178* (Directive on Electronic Commerce).

[2] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online, A contribution from the European Commission to the Leaders' meeting in Salzburg on 19-20 September 2018' COM (2018) 640 final, [hereinafter 'the proposal' or 'the regulation'].

I. Terrorists' Use of the Internet

The internet and social media have become highly popular communication tools used by individuals, organisations, and institutions in every segment of the society.[3] The internet and social media platforms can lead to a notable presence of certain groups, which leads to more attention for its objectives and explains the appeal to terrorist groups.[4] Hence, it is not surprising that terrorist organisations and their supporters also use it for various purposes. The internet is used by terrorist groups and their supporters to promote or support acts of terrorism through propaganda, financing, training and planning.[5] According to the 2018 Europol report, terrorist groups and their sympathisers are using the internet especially for propaganda and recruitment and this activity happens mostly on the 'surface web', the part of the internet that is open and accessible for everyone. Europol has investigated that more than 150 social media platforms were abused by terrorist groups and supporters, but also file-sharing sites and messaging services store, disseminate and advertise links to such terrorist content.[6] Next to this, Europol found that terrorist activity also takes place on the 'Dark Net'. This is a decentralised and anonymous network on the internet that is accessed by using special software.[7] Supposedly, the Dark Net is foremost used for the purpose of fundraising and advertising terrorist propaganda that is hosted on the surface web.[8]

This issue already attracted the attention of intergovernmental institutions, legal scholars, and internet companies, but it was the recent terrorist attacks in Europe that raised the concern that more effective responses are necessary.[9] While major social media platforms like Facebook, Twitter and YouTube already reduced the accessibility of terrorist content, start-up social media and companies with limited resources are increasingly abused by terrorists.[10] Terrorist organisation the Islamic State (IS) is considered to be effective in using the internet for spreading their narratives and gather support for their objectives.[11] In the last few years there was a mass dissemination of IS related information, high quality images, videos and audio recordings on social media and messaging apps. The online messages consist of messages that are calling for men, women and children to take part in the jihad. A small part of this content shows extreme violence such as beheading videos of western hostages and videos that depict jihadist fighters. Most content that IS produces is non-violent and depicts, for example, the civilian life in the caliphate, community projects and religious preaching. By choosing this kind of propaganda, the organisation aims to attract all kind of people that aspire a global Muslim community and

[3] J. Argomaniz, 'European Union Responses to Terrorist Use of the Internet', *Cooperation and Conflict* 2015-50(1), p. 250.

[4] L. Bertram, 'Terrorism, the Internet and the Social Media Advantage', *Journal for Deradicalization* 2016-7, p. 230.

[5] United Nations Office of Drug and Crime, 'The use of the Internet for terrorist purposes', 2012, at: https://www.unodc.org/documents/frontpage/Use_of_Internet_for_Terrorist_Purposes.pdf (accessed on 28 April 2019), p. 3.

[6] Europol, EU Terrorism Situation and Trend Report (TE-SAT) 2018, at: https://www.europol.europa.eu/activities-services/main-reports/european-union-terrorism-situation-and-trend-report-2018-tesat-2018 (accessed on 28 April 2019), p. 15.

[7] Software such as TOR (The Onion Router) or I2P (Invisible Internet Project). Gabriel Weimann, 'Terrorist Migration to the Dark Web', *Perspectives on Terrorism* 2016-10(3), p. 41.

[8] Europol 2018, supra note 6, p. 15.

[9] S. Jones, B. Toucas & M.B. Markusen, 'From the IRA to the Islamic State: The Evolving Terrorism Threat in Europe' *Center For Strategic & International Studies: Transnational Threats Program* 2018, pp. 10-14.

[10] Europol 2018, supra note 6, p. 15.

[11] Y. Veilleux-Lepage, 'Paradigmatic Shifts in Jihadism in Cyberspace: The Emerging Role of Unaffiliated Sympathizers in Islamic State's Social Media Strategy', *Journal of Terrorism Research* 2016-7(1), p. 41.

caliphate.[12] The effectiveness of their strategy lays in the fact that the content that is produced by the official IS media organisation is being disseminated by their supporters all over the world on the major social media platforms, by sharing, tweeting and liking.[13] Those sympathisers are not necessarily required to engage in terrorist acts, but can comfortably support IS from behind their computer.[14] As a result of the losses of IS of territory and resources, the IS media organisation encouraged their supporters to manifest themselves online by producing and disseminating user-generated content, for example by praising attacks, inciting to violence, and archiving older official terrorist content.[15] The type and amount of terrorist propaganda has also changed because of the actions of major social media companies like Facebook and YouTube. These companies are increasingly removing IS-related content and banning related social media accounts and channels from their platforms.[16] According to the Europol report, in 2017 it was difficult to find English content that is related to IS on Facebook and Instagram because these platforms detect and take it down more quickly, in particular videos that contain extreme violence or calls to violence. On the other hand, due to the focus on IS propaganda, content that is produced by al-Qaeda and other Jihadist accounts became more visible.[17] It is evident that IS supporters are also aware of the monitoring of law enforcement and social media platforms. Therefore, IS promoted the use of tools like the *TOR* browser, that makes it possible to anonymously visit specific IS websites or Dark Net websites. In addition, they relocated their communication to private chat groups, encrypted message applications such as Telegram, blogs, web fora and platforms with smaller capacities to counter terrorist content.[18] Yet, in 2016 a jihadi author urged the IS supporters to increase the use of popular platforms like Facebook, Twitter and Instagram again in order to reach a wider audience.[19]

With regards to the extreme right-wing groups or individuals, it has shown that they are also using the internet to spread their messages and reach out to others. The report of Rebecca Lewis: 'Alternative Influence: Broadcasting the Reactionary Right on YouTube' is particularly useful to determine how the far-right uses online social networking. In this report, she identified a media system on YouTube that she calls the Alternative Influence Network (AIN).[20] The followers of this network produce content and discuss subjects that range from mainstream conservatism to white nationalism. The common factor is their general opposition to feminism, social justice and left-wing politics and the common target groups are Jews, Muslims, migrants or government officials.[21] Yet, it is clear that the most extreme right-wing content is visible on online discussion fora where online content is not moderated. An example is the website *Stormfront.org*, that functions as a virtual community for various white power movements, where they disseminate

[12] Europol 2018, supra note 6, p. 32.

[13] Veilleux-Lepage 2016, supra note 11, p. 41.

[14] G. Weimann, 'New Terrorism and New Media', *Commons Lab of the Woodrow Wilson International Center for Scholars* 2014-2, p. 3.

[15] Idem, pp. 31, 32.

[16] Idem, p. 15; M. Conway and others, 'Disrupting Daesh: Measuring Takedown of Online Terrorist Material and Its Impacts', *Studies in Conflict and Terrorism* 2018, pp. 6, 12.

[17] Europol 2018, supra note 6, p. 31; Conway 2018, supra note 16, pp. 6, 12, 13.

[18] Europol 2018, supra note 6, pp. 31, 32.

[19] Ibid.; C. Bunzel, '"Come Back to Twitter": A Jihadi Warning Against Telegram', Jihadica, 18 July 2016, at: www.jihadica.com/come-back-to-twitter/ (accessed on 28 April 2019).

[20] R. Lewis, 'Alternative Influence: Broadcasting the Reactionary Right on YouTube', *Data & Society* 2018, pp. 3, 4.

[21] Idem, p. 8.

their ideologies, discuss with other users and eventually create a collective identity.[22] Stormfront is the largest and most active forum and is used by known right-wing terrorists, among others Anders Breivik who is convicted for the Oslo terrorist attacks.[23] Other examples are the online discussion fora *4chan* and *8chan*. On the latter, Brenton Tarrant, the perpetrator of the terrorist attack in New Zealand, announced the attack on two mosques on 15 March 2019, by sharing a manifesto and a link to the Facebook livestream of his attack. This livestream was probably shared to get maximum attention for the attack and while Facebook removed the video within one hour after a notification of the police, the video was repeatedly re-uploaded and disseminated by users on YouTube, Twitter and other platforms and messaging applications.[24] In the aftermath of the attack, users of *8chan* commentated the disseminated manifesto of the terrorist and the video of the attack, whereby some messages were directly praising the attack or inciting to violence.[25] Similarly, users on 4chan commentated the attack in a tram in Utrecht on 18 March 2019 where four people were killed. This attack was inspired by a radical Islamic ideology, and this provoked reactions of users on 4chan whereby some users marked this attack as the beginning of a racial war and were directly inciting to violence.[26]

I.2 Online radicalisation

The European Commission is concerned about terrorists' misuse of the internet to recruit supporters, prepare and facilitate terrorist acts, glorify those acts and generate fear in society. The Commission underlines the fact that the 'lone wolves' who committed recent terrorist attacks in Europe were radicalised and inspired by terrorist content online.[27] The Islamic State for example, possibly radicalised Muslims in Europe via the internet, by providing them comradeship and belonging.[28] Members of Islamic State encouraged those radicals to participate in the organisation and travel to the caliphate, or to carry out attacks in their own country.[29] The exposure to terrorist propaganda itself, without interaction with the terrorist organisation, may also contribute to the radicalisation of lone actors that carry out attacks.[30]

Similarly, with regards to extreme right-wing groups, social media facilitates the creation of a collective identity via networking, information-sharing and discussing. This gives extremist right-

[22] R. Scrivens, G. Davies and F. Richard, 'Measuring the Evolution of Radical Right-Wing Posting Behaviors Online' *Deviant Behavior* 2018, pp. 1, 2.

[23] On 22 July 2011, Anders Breivik exploded a bomb in Oslo, killing 6 people and injured hundreds, after that he shot and killed 69 people in the Workers Youth Summer camp on Utoya Island. See in general: H. Beirich, 'White Homicide Worldwide', *Southern Poverty Law Center*. (2014), at: https://www.splcenter.org/sites/default/files/d6_legacy_files/downloads/publication/white-homicide-worldwide.pdf (accessed on 28 April 2019).

[24] J. Schellevis & P. Houthuijs, 'Christchurch terrorist wist hoe hij de maximale aandacht zou krijgen', NOS, 15 March 2019, at: https://nos.nl/artikel/2276146-christchurch-terrorist-wist-hoe-hij-de-maximale-aandacht-zou-krijgen.html (accessed on 28 April 2019); 'Dader Christchurch zat op anoniem forum 8chan, wat is dat voor plek?', *Nederlandse Omroep Stichting* (The Netherlands, 16 March 2019) at: https://nos.nl/artikel/2276286-dader-christchurch-zat-op-anoniem-forum-8chan-wat-is-dat-voor-plek.html (accessed on 28 April 2019).

[25] H. Bahara & A. Kranenberg, '"Alsjeblieft tien doden, we hebben een terroristische aanslag nodig": hoe gebruikers van extreemrechtse fora reageerden op de aanslag in Utrecht', *De Volkskrant(The Netherlands,* 22 March 2019) at: https://www.volkskrant.nl/nieuws-achtergrond/alsjeblieft-tien-doden-we-hebben-een-terroristische-aanslag-nodig-hoe-gebruikers-van-extreemrechtse-fora-reageerden-op-de-aanslag-in-utrecht~bf72207c/ (accessed on 28 April 2019).

[26] Ibid.

[27] Commission proposal 2018, supra note 2, p. 1.

[28] Bertram 2016, supra note 4, p. 234.

[29] Ibid.

[30] Idem, p. 235.

wing groups the opportunity to recruit new members and organise extreme-right events.[31] This collective identity may also result in attacks of lone actors. For example, Brenton Tarrant, the aforementioned perpetrator of the terrorist attacks in New-Zealand, was no part of an extreme-right group, but developed his ideas on online discussion fora and was inspired by extreme right-wing online influencers and by the manifesto of the Norwegian far-right terrorist Anders Breivik.[32] There are various explanations why and how someone radicalises, but it is often seen as a process that starts with the search for a fundamental meaning and a return to a root ideology. Those individuals may search for like-minded individuals or groups and unite with those who have adopted a violent form of that ideology.[33] This can result in the polarisation of the social space and the construction of an ideological discourse that is characterised by a clear distinction between "us" and "the others". In order to make this distinction, the others are often being dehumanised in a process of scapegoating.[34] It is stated that the more individuals perceive that there is no alternative interpretation of their political concepts that are part of a specific ideology, the more this individual is radicalised.[35] However, in order to turn to violent behaviour, the radicalised individual has to adhere to a violent ideology that denies individual freedoms of the "others", and the radicalised individual has to perceive that there is no other option possible to reach his or her goals. This process accelerates when alternative options are rapidly decreasing and the ideological calls for violent actions are increasing.[36]

Some authors argue that the internet contributes to the radicalisation process because it acts as an 'echo' chamber.[37] When a certain narrative is repeated on different platforms while alternative views and narratives are rejected or not visible at all, this creates the impression that those views are widely supported. Individuals will focus on information that confirms their opinions and reject alternative views, and the more their opinions are confirmed, the sooner individuals will normalise it and adopt more extreme views.[38] According to a research about online discussion fora, the echo chamber effect exists among radical and ideologically homogeneous online groups, like the extreme-right forum Stormfront.org.[39] The author found that while more extreme people are likely to turn to such radical groups, participation therein further exacerbated their views.[40] However, others have argued that the echo chamber exists only in some circumstances and the effects for extremism and radicalisation cannot be generalised, therefore it is difficult to develop a model that targets echo chambers effectively, without unintended consequences.[41] In my opinion, it is convincing that more extreme opinions will be adopted on online discussion fora

[31] S. Alava, D. Frau-Meigs, & G. Hassan, 'Youth and Violent Extremism on Social Media Violent Extremism on Social Media: Mapping the Research', *UNESCO* 2017, p. 23.

[32] R. Evans, 'Shitposting, inspirational terrorism and the Christchurch mosque massacre', *Bellingcat*, (15 March 2019) at: https://www.bellingcat.com/news/rest-of-world/2019/03/15/shitposting-inspirational-terrorism-and-the-christchurch-mosque-massacre/ (accessed on 28 April 2019).

[33] Alava and others 2017, supra note 31, p. 12.

[34] Ibid.

[35] D. Koehler, 'The radical online: Individual radicalization processes and the role of the Internet', *Journal for Deradicalization* 2014-1, p. 125.

[36] Ibid.

[37] I. von Behr and others, 'Radicalisation in the Digital Era: The Use of the Internet in 15 Cases of Terrorism and Extremism', *RAND Europe* 2013, p. 27.

[38] Veilleux-Lepage 2016, supra note 11, p. 45; B. van Ginkel, 'Incitement to Terrorism: A Matter of Prevention or Repression?' *ICCT Research Paper* 2011*,* p. 6.

[39] M. Wojcieszak, ''Don't talk to me': Effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism', *New Media and Society* 2010-12(4), p. 643.

[40] Ibid.

[41] K. O'Hara & D. Stevens, 'Echo Chambers and Online Radicalism: Assessing the Internet's Complicity in Violent Extremism', *Policy and Internet* 2015-7(4), p. 403.

like *8chan* or *4Chan,* where certain online sub groups like */pol/* are overrepresented by white supremacists, neo-Nazis and other extreme right-wing ideologists.[42] Similarly, the encrypted message applications like *Telegram* or websites on the Dark Net provides extremist Islamic groups the opportunity for discussing, radicalising and recruitment without external interferences. In contrast, social media platforms like Facebook, YouTube and Twitter are somehow different. These social media platforms give the opportunity to reach a large audience in a short time period, where people may engage directly through discussing, sharing and recreating content.[43] This makes it also possible to react, discuss and interact with opposing groups and individuals with alternative opinions. For example, there are users on social media who undermine or ridicule the messages of terrorist organisations by producing memes or videos.[44] Similarly, when you look at the comment section under news articles that are shared on Facebook or Twitter, the comments and subsequent discussions show views and opinions that can be identified as part of extreme right-wing or extreme left-wing ideologies and everything in between.[45]

It is possible that such interactions on social media influence the ideologies or actions of both groups. This relates to the interplay between opposing extremist groups, what is identified as 'cumulative extremism'. Extremist groups may mutually radicalise based on the extremist ideologies or violent actions of the other group, what fosters recruitment on both sides and eventually leads to a spiral of violence.[46] It is observed that violent interactions on the street between opposing movement have escalated towards violence and contributed to a diminishing of social cohesion.[47] On the other hand, the interactions between opposing extremist groups has also led to a de-escalation, whereby the groups shifted to non-violent tactics.[48] It is especially interesting to see what the effects are of interactions between opposing ideologies on social media platforms.[49] At this moment, it is not exactly clear whether those interactions enhance the polarisation in society and escalating conflicts between individuals and groups towards violence or the just the opposite. One study that is conducted on internet users may shed light on this. This study found that when social media users are getting negative feedback in the form of dislikes or disapproving reactions on their comments in a political discussion, this led to an increase of their undesired behaviour.[50] However, that effect may also be explained by so-called 'trolling', what means that the user is deliberately posting provocative questions or viewpoints in order to provoke a reaction, promote the posting of divergent opinions and stir up the conflict about sensitive subjects within a community.[51] Nevertheless, in general it is likely that discussions and exposure to different opinions on social media platforms reduces the echo chamber effect.[52] It

[42] /pol/ stands for politically incorrect, see in general: https://en.wikipedia.org/wiki//pol/ or the subforum itself at: https://8ch.net/pol/index.html.
[43] Weimann 2014, supra note 14, pp. 2, 3.
[44] M. Lakomy, 'Cracks in the Online "Caliphate": How the Islamic State Is Losing Ground in the Battle for Cyberspace', *Perspectives on Terrorism* 2017-11, p. 46.
[45] As I have witnessed on Facebook, especially when the news article relates to immigration policy, sensitive political issues or terrorist attacks.
[46] J. Bartlett & J. Birdwell, 'Cumulative radicalisation between the far-right and Islamist groups in the UK: A review of evidence', *Demos* 2013, p. 5.
[47] J. Busher & G. Macklin, 'Interpreting "Cumulative Extremism": Six proposals for enhancing conceptual clarity', *Terrorism and Political Violence* 2015-27(5), p. 889.
[48] Idem, p. 886.
[49] Idem, pp. 895, 896.
[50] J. Cheng, C. Danescu-Niculescu-Mizil, & J. Leskovec, 'How Community Feedback Shapes User Behavior', *Association for the Advancement of Artificial Intelligence* 2014, p. 9.
[51] Ibid. In general: https://en.wikipedia.org/wiki/Internet_troll.
[52] O'Hara 2015, supra note 41, p. 410.

follows from research that exposure to moderate views or to mixed media decreased extremist attitudes among individuals. In contrast, the exclusive exposure to extremist right-wing discourses resulted in radicalisation of attitudes among individuals with different political backgrounds.[53] While it is true that the internet is used to search for information that confirms existing political beliefs, most people are gathering information from a variety of online and offline sources.[54] Moreover, in the search for confirmative information, it is stated that people are not ignoring all alternative information.[55]

Apart from the question if the internet and social media platforms exacerbates extremist views or not, does the removal of terrorist propaganda that incites to terrorism have the effect of preventing violent radicalisation and result in a decline of terrorist acts? With a view on the Combatting Terrorism Directive, this question may be answered in the positive. In recital 22 of that Directive it is stated that the removal of online content that constitutes public provocation to commit a terrorist offence is an effective means to prevent terrorism.[56] It is argued that public provocation or incitement to terrorist offences is the starting point for terrorism, because it aims to change the perception that people have of violent acts. This may eventually radicalise people into committing terrorist attacks.[57] Yet, it is not proven that there is a direct causal relation between extremist content online and the adoption of extremist ideologies and engagement in terrorism.[58] There is consensus among scholars that the internet is a source of information and communication for terrorists and a platform for extremist propaganda that facilitates radicalisation.[59] According to a study among terrorist offenders, 30% of the terrorist actors had accessed extremist ideological content online, and 29% had communicated with other radicals online.[60] However, it is important to underline that the internet has a mere enabling role, because it was not the deciding factor for radicalised actors to engage in a violent act. The motivation, intent and capabilities of terrorists are influenced by many factors, and online activity must be seen as one of them.[61] It was found in those studies that online learning and communicating complemented interactions in the offline world, and those face-to-face interactions remained crucial in the radicalisation processes of most terrorist actors.[62] Moreover, the influence of the internet on radicalisation and the use of the internet for communicating, planning and learning varied among different types of terrorists.[63] The authors concluded that the influence of the internet is generally dependent on the need of

[53] Alava and others 2017, supra note 31, p. 23.

[54] Council of Europe, 'Algorithms and Human Rights, Study on the human rights dimensions of automated data-processing techniques and possible regulatory implications' prepared by the committee of experts on internet intermediaries, *DGI [2017]12,* (2018), p. 18; E. Dubois & G. Blank, 'The echo chamber is overstated: the moderating effect of political interest and diverse media', *Information Communication and Society* 2018-21(5), pp. 741, 742.

[55] O'Hara 2015, supra note 41, p. 410.

[56] Directive 2017/541 of the European Parliament and of the Council of 15 March 2017 on combatting terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA [2017] *OJ L 88*, recital 22.

[57] Van Ginkel 2011, supra note 38, p. 3.

[58] M. Conway, 'Determining the role of the internet in violent extremism and terrorism: Six suggestions for progressing research', *Studies in Conflict and Terrorism* 2017-40(1), p. 77; Alava and others 2017, supra note 31, p. 46.

[59] P. Gill and others, 'What Are the Roles of the Internet in Terrorism? Measuring online behaviours of convicted UK terrorists'*, VOX-Pol* 2015, p. 31-33; Von Behr and others 2013, supra note 37, pp. 24-29.

[60] P. Gill and others, 'Terrorist Use of the Internet by the Numbers: Quantifying Behaviors, Patterns, and Processes', *Criminology and Public Policy* 2017-16(1), p. 107.

[61] Idem, pp. 110, 114.

[62] Idem, p. 114; Gill and others 2015, supra note 59, pp. 31-33.

[63] Gill and others 2017, supra note 60, p. 101.

the radicalised individual and the availability of co-offenders, extremist ideological content and expertise.[64] First, it was evident that lone actors use the internet more than members of a terrorist cell, in particular to learn about how they should conduct an attack.[65] Second, those who carried out an armed attack or used explosives, were likely to have learned online, in contrast with those who carried out more primitive types of attacks.[66] Third, the extremist right-wing terrorists were using the internet far more than Jihadist terrorists, possibly because the extreme right-wing extremists are more likely to be a lone actor.[67]

In conclusion, the radicalisation process and decision to use violence is a result of both online and offline interactions, and the extent of online activity differs among type of terrorists. In order to effectively counter radicalisation processes and prevent future terrorist attacks, it is recommended by scholars that policy and regulation should not target solely the location where radicalisation may take place, such as the online environment. Rather, there should be a better understanding of the drives and needs of different types of terrorists and the reason why and how they use the internet or the offline environment for radicalisation, learning and planning.[68]
Before turning to the analysis of the proposed regulation that aims to deal with online terrorist content, it is important to look into the possible detrimental effects of removing terrorist content from social media platforms. This will be discussed below.

I.3 Consequences of removing terrorist content on social media platforms

When social media companies are removing content that is extremist, radical or critical towards certain individuals or groups, whether it is inciting to violence or not, this possibly has the effect of preventing vulnerable individuals to come in contact with extreme ideologies and opinions. However, there is not enough evidence to explain the relation between extremist content, radicalisation and terrorist attacks. This is problematic because the removal of extremist content does have some detrimental consequences. Firstly, the removal of terrorist content will result in the relocation of communication to echo chambers like the aforementioned online discussion fora or encrypted message applications. Although some authors observed that the more extremist Islamic discourse already shifted to the Dark Net as a result of the monitoring on social media.[69] It can be expected that also the potential radical individual with divergent opinions will sooner connect with radicalised, extremist individuals and groups on such homogeneous online discussion fora, where extremist views will be exacerbated. This implies that terrorist groups and radical individuals are able to disseminate propaganda, provide and search for instructions, recruit new members and coordinate terrorist attacks, without interruption of governments and social media platforms.[70] This leads to the conclusion that there is not enough evidence to claim that the removal of extremist or terrorist content on social media platforms will prevent the radicalisation of individuals with divergent opinions or moderately extremist beliefs.

Secondly, the removal of terrorist and extremist content will make it impossible for public authorities, civil society, deradicalised individuals and religious leaders to counter extreme content by supplying alternative opinions. There are some initiatives that focus on this counter radicalisation strategy, for example the Radicalisation Awareness Network (RAN), that brings

[64] Idem, pp.107-109, 114.
[65] Idem, p. 110.
[66] Ibid.
[67] Ibid.
[68] Idem, p. 114.
[69] Weimann 2016, supra note 7, p. 41.
[70] Ibid.

practitioners from around Europe together to prevent radicalisation.[71] According to a report of the RAN, it is important to reduce the amount of terrorist content, but it will be impossible to prevent all terrorist material on the internet. It is therefore essential to provide for skills that people need to be sceptical of extremist content.[72] Moreover, reducing the accessibility of terrorist content will not prevent people to search for information and get attracted to the ideology that it presents. Therefore, it is necessary that alongside any counter narrative campaign, there should be a focus on protecting the rights of minorities, addressing the grievances of individuals or groups and providing them opportunities to participate fully in society.[73]

Counter narratives on social media will not result in full deradicalisation of individuals, but it dismantles extremist messages, provides alternative opinions and ideas that can influence the radical individual to reconsider extremist beliefs.[74] It is clear that the effectiveness of counternarratives depends on the specific content and how it is disseminated. First, the counter narrative should contradict the ideological themes that promote terrorism. Second, the source of the message should be credible to the audience members.[75] For example, a violent Islamic narrative should not be countered by governments, but rather by religious leaders that convey that violence and atrocities conflict with religious values.[76] The message should also not directly confront the extremist beliefs directly, but rather reframe the narrative.[77] According to Ravndl (2017), public counter messages that target the far-right should not be overly aggressive, derogatory or moralising, since public repression and stigmatisation of radical right actors and opinions will fuel anger and violence.[78] For example, when concerns about high immigration are dismissed by policymakers or the 'elite', the ideas of the extreme right-wing will be confirmed. The politically correct 'elite' is namely perceived as part of the underlying conflict and persons that adhere to the extreme-right ideologies will experience feelings of marginalisation. This enhances the polarisation in society and some far-right actors may turn to violent action to generate political change.[79] Rather, the counter narrative should focus on countering the idea of 'us' and 'them'.[80]

This section makes clear that radicalisation may take place on the internet. The echo chamber effect is clearly visible within homogeneous online groups, but it is not evident that this effect takes place on social media platforms. The open nature of social media rather provides the opportunity to engage with a wider community where individuals are exposed to a variety of information and opinions. There is not enough evidence to establish a causal relation between online terrorist or extremist content and violent radicalisation and terrorist attacks. This implies

[71] Radicalisation Awareness Network, at: https://ec.europa.eu/home-affairs/what-we-do/networks/radicalisation_awareness_network_en. (accessed on 28 April 2019).
[72] Radicalisation Awareness Network (RAN), 'Preventing Radicalisation to Terrorism and Violent Extremism: Approaches and Practices' 2018, a: https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/radicalisation_awareness_network/ran-best-practices/docs/ran_collection-approaches_and_practices_en.pdf (accessed on 28 April 2019), p. 469.
[73] Ibid.
[74] Idem, p. 474.
[75] K. Braddock & J. Horgan, 'Towards a Guide for Constructing and Disseminating Counternarratives to Reduce Support for Terrorism', *Studies in Conflict & Terrorism* 2016-39(5), p. 386.
[76] RAN 2018, supra note 72, p. 471.
[77] Idem, p. 475.
[78] Jacob Ravndal, 'Right-Wing Terrorism and Violence in Western Europe: A Comparative Analysis', *Series of dissertations submitted to the Faculty of Social Sciences*, University of Oslo 2017-667, p. 45.
[79] Idem, p. 33.
[80] RAN 2018, supra note 72, p. 471.

that it is unconvincing that removing terrorist content from social media platforms will prevent radicalisation and reduce terrorist attacks in Europe. Naturally, the issue of terrorist content on social media platforms requires appropriate action. In the next section I will describe how the most recent regulatory attempt in the European Union aims to do that.

II. The European Commission's Proposal for Regulation for Preventing the Dissemination of Terrorist Content Online.

The use of the internet by terrorists and the need for effective action to respond to this issue is also widely discussed in European politics. In March 2017, the German Minister of Justice and Consumer protection stated that the pressure on social networks must be increased and proposed to impose fines on social media companies when they do not remove illegal content fast enough.[81] Another example is the statement issued by European leaders Theresa May and Emmanuel Macron in June 2017 where they expressed to join forces in tackling online radicalisation, most importantly by exploring the possibility to create a new legal liability for internet companies if they fail to remove content.[82] Almost one year later, in April 2018, the French and German Interior Ministers wrote a letter to the European Commission in which they propose that internet companies should develop tools that automatically detect terrorist and other illegal content and remove it within one hour. They further add that if companies fail to remove terrorist content, they should be sanctioned.[83] Ultimately, in June 2018, the European Council issued their 'Conclusions, inviting the European Commission to present a legislative proposal to improve the detection and removal of content that incites hatred and the commitment of terrorist acts.[84]

The amount of political statements that emphasises the responsibility for online intermediaries in countering terrorist content is remarkable and is clearly putting pressure on the EU institutions to act. In pursuance of setting up a binding framework that will tackle terrorist content online in Europe, it is essential to gain support from the relevant stakeholders and find consensus on the methods that will be used. The many legal documents, statements and voluntary arrangements show the difficulties of creating legislation that intends to change the roles and responsibilities of public and private actors in law enforcement and security matters. The EU must come up with creative ways to fit and measure such legislation into the existing EU law and fundamental rights regime. The most recent attempt to deal with terrorist content in the European Union aims to do just that. The European Commission acknowledged that further action against terrorist content is needed, since the voluntary initiatives of internet companies and EU institutions fall short in addressing this problem.[85] Therefore, in September 2018, the European Commission submitted

[81]' Bekämpfung von Hasskriminalität und strafbaren Falschnachrichten – Bessere Rechtsdurchsetzung in sozialen Netzwerken' (14 March 2017) at: https://www.bmjv.de/SharedDocs/Artikel/DE/2017/03142017_GE_Rechtsdurchsetzung_Soziale_Netzwerke.html (accessed on 28 April 2019).

[82] 'UK and France announce joint campaign to tackle online radicalisation', at: https://www.gov.uk/government/news/uk-and-france-announce-joint-campaign-to-tackle-online-radicalisation (accessed on 28 April 2019).

[83] J. McNamee, 'Leak: France and Germany demand more censorship from Internet Companies, ' EDRi, 7 June 2018, at: https://edri.org/leak-france-germany-demand-more-censorship-from-internet-companies/ (accessed on 28 April 2019).

[84] European Council Conclusions of 28 June 2018, Press Release 421/18, 29 June 2018, p. 2, at: https://www.consilium.europa.eu/en/press/press-releases/2018/06/29/20180628-euco-conclusions-final/pdf. (accessed on 28 April 2019).

[85] Such as the European Commission, 'EU Internet Forum: Bringing together governments, Europol and technology companies to counter terrorist content and hate speech online', Press Release December 3,

a proposal for regulation on preventing the dissemination of terrorist content online.[86] This regulation introduces procedures and obligations for 'hosting service providers' and competent authorities in the Member States. To clarify, hosting service providers are defined as providers of information society services that store information that is provided by the user of that service and make that information available to third parties.[87] You should think of Facebook, YouTube and Instagram, but also smaller social media platforms. This regulation applies to all hosting service providers that enable legal or natural persons in the Union to use the services, even if they are not established in the Union.[88]

The first procedure that is introduced in this regulation is the removal order, which is laid down in article 4. This means that law enforcement or judicial authorities in the member state determine that certain online information constitutes as terrorist content and issue a decision that the hosting service provider must remove it within one hour.[89] The second procedure is the referral as laid down in article 5 of the regulation. This is similar to an existing procedure, the EU Internet Referral Unit, where social media platforms voluntary consider content against their terms and conditions that Europol referred to them.[90] The regulation adds that this referral may also be sent by the competent authority of the Member State.[91] The hosting service provider is required to develop operational and technical measures in order to assess this content immediately against their own terms and conditions and remove the content if it is in violation of the latter.[92] The third procedure in the regulation requires hosting service providers to take proactive measures. These consist of preventing the re-upload of terrorist content, checking content against databases of known terrorist content and employing automated technical tools to identify new terrorist content.[93] Article 6 of the regulation describes different situations in which proactive measures should be taken: on initiative of the hosting service provider, in co-operation with the competent authority, or imposed by the competent authority. These three different procedures and obligations for the hosting service provider will be discussed in light of the E-Commerce Directive and fundamental rights.

III. Legal Assessment of the Proposal in light of the E-Commerce Directive.

In drafting new regulation, there should be attention to compatibility of the measures with existing EU law. In this section, I will assess the proposed procedures in the terrorist content regulation in relation with the E-commerce Directive, in particular, the liability framework and monitoring obligations.

2015, at: http://europa.eu/rapid/press-release_IP-15-6243_en.htm (accessed on 28 April 2019);European Commission, 'European Commission and IT Companies announce Code of Conduct on illegal online hate speech', Press Release May 31, 2016, at: http://europa.eu/rapid/press-release_IP-16-1937_en.htm (accessed on 28 April 2019).

[86] Commission proposal 2018, supra note 2.

[87] Idem, art. 2 and recital 10 provides examples of hosting service providers: social media platforms, video streaming services, file sharing services, cloud services that make information available to third parties, and websites where users can communicate with each other.

[88] Idem, recital 10. However, according to art. 2 and recital 11, there should be a substantial connection to the Union. This exists when the service provider has an establishment in the Union, has a considerable number of users in the Union or targets their activities towards the Member States.

[89] Idem, art. 4(1), (2), recital 13.

[90] Idem, recital 15.

[91] Idem, art. 5(1), 5(4).

[92] Idem, art. 5(2), 5(5).

[93] Idem, recital 18.

III.1 The Liability Framework

The liability exemption framework is set out in article 14 of the E-commerce Directive. This article aims to protect service providers against liability for third-party content that has been published on their services. There are certain requirements and conditions before liability can be exempted, and it has been subject of various cases from the CJEU and the ECtHR.

The liability exemption in article 14(1) of the E-Commerce Directive only applies to the service provider that confines itself to providing that service 'neutrally' by a merely technical and automatic processing of the data provided by its customers. In other words, it should not have played an active role allowing it to have knowledge or control of the data stored.[94]

In the proposed regulation of the Commission, it is stated in recital 5 that the liability exemption framework in article 14 of the E-commerce Directive is not affected by application of the regulation. It is underlined that when hosting services take measures according to the regulation, including proactive measures, 'this should not in themselves lead to that service provider losing the benefit of the liability exemption provided for in that provision.'[95] The European Commission implies with recital 5 of the regulation that the role or the activities of the hosting service providers will not automatically be considered to be 'active' when they take the measures in accordance with the regulation and fall within the scope of article 14 of the E-Commerce Directive.

Article 14 E-Commerce Directive provides that hosting service providers are under certain conditions not liable for information stored at the request of a recipient of that service. The first condition as laid down in article 14(1)(a) is that the provider 'does not have actual knowledge of illegal activity or information, and as regards claims for damage, is not aware of facts or circumstances from which that illegal activity or information is apparent.' The second condition in article 14(1)(b) reads: 'that the provider, upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information.' This is also named the principle of 'notice and take down' and must be undertaken in observance of the principle of freedom of expression.[96]

In the *L'Oréal v. Ebay* case, the Court argued that the liability exemption would not apply when the service provider is aware of facts and circumstances on the basis of which 'a diligent economic operator' should have realised the illegality, but fails to expeditiously remove it.[97] The Court explains that service providers can become aware of such facts or circumstances in various situations. For example, as a result of an investigation undertaken on its own initiative or when they receive an external notification on illegal content on their platforms.[98] In the case that such a notification is insufficiently precise and inadequately substantiated, the failure to remove this content will not automatically result in liability. However, national courts should take the presence of a notification into account when determining if the service provider was actually aware of facts or circumstances of which a diligent economic operator should have realised the illegality.[99]

[94] E-Commerce Directive, supra note 1, recital 42; Case C-324/09, *L'Oréal v eBay International,* [2011] ECR I06011, at paras. 113, 116.

[95] Commission proposal 2018, supra note 2, recital 5.

[96] E-Commerce Directive, supra note 1, recital 46; Case C-324/09, *L'Oréal v eBay International* [2011] ECR I06011, Opinion of Advocate-General Jääskinen, paras. 155-56.

[97] *L'Oréal v eBay,* supra note 94, para. 120.

[98] Idem, paras. 121, 122.

[99] Idem, para. 122.

Hosting service providers may have actual knowledge of past or current illegal activity. In applying this to the proposed regulation, when the competent authority in a Member State issues a removal order with reasons why this content is considered to be terrorist content, this will give the hosting provider actual knowledge of the illegal content. The hosting service provider can choose to disable access to the content in that Member State alone, but then it risks liability in another Member State because it had actual knowledge based on the removal order. The hosting service provider will sooner choose to disable access in the European Union or remove the content entirely, also because it would take too much time to look if this content violates the law of other countries.[100]

In contrast, with regards to the procedure of a referral that is introduced in the proposed regulation, it is not so clear whether that leads to actual knowledge. This referral consists of content that doesn't fulfil the legal requirements of terrorist content, otherwise the competent authority would have issued a removal order. According to article 5 of the regulation, hosting service providers shall 'as a matter of priority' make an assessment of the referral and inform the competent authority 'expeditiously' about the outcome and timeframe of the actions taken. Non-compliance with these two obligations will result in penalties, despite the initially voluntary nature of the referral.[101] These penalties should not encourage hosting service providers to over-remove content. Yet, when the service provider decides not to remove the referred content because it does not constitute terrorist content, it remains possible for national authorities to establish liability when this content appears to be illegal. This referral is namely an external notification that may lead to actual knowledge of illegal activity or awareness of facts or circumstances on the basis of which 'a diligent economic operator' should have identified the illegality.[102] This uncertainty about liability and the threat of penalties will increase the propensity of the hosting service provider to remove the referred content.

It is established that actual knowledge also concerns future illegal activity, and the failure to remove this can lead to liability of the hosting service provider. Initially, this sounds rather impossible to accomplish, but there are situations in which the service operator has actual knowledge of illegal activity, but does not prevent that user to continue or repeat that same infringement in the future.[103] For example, it is possible to prevent this by using algorithms to detect and prevent the upload of previous identified infringements.[104] In the proposed regulation, this is expected through the implementation of the proactive measures as set out in article 6 of the regulation. These include the prevention of the reupload of terrorist content on their platforms, checking content against databases of known terrorist content and using reliable technical tools to detect, identify and remove new terrorist content. The main question is if the hosting service providers can avoid liability when they implement these measures in accordance with the proposed regulation.

In answering that question it is useful to look at the *Telekabel Wien* case, where the CJEU assessed an injunction on a service provider to block access to copyright infringing content.[105] The

[100] J. Barata, 'New EU Proposal on the Prevention of Terrorist Content Online, An Important Mutation of the E-Commerce Intermediaries' Regime', *The Center for Internet and Society* 2018, p. 7.

[101] Commission proposal 2018, supra note 2, art. 18(1)(c).

[102] *L'Oréal v. eBay*, supra note 94, paras. 120-122.

[103] Opinion of Advocate-General Jääskinen, supra note 96, paras. 165, 168.

[104] K. Klonick, 'The new governors: The people, rules, and processes governing online speech', *Harvard Law Review* 2018-131, pp. 1638-9.

[105] Case C-314/12, *UPC Telekabel Wien v. Constantin Film Verleih GmbH and Wega Filmproduktionsgesellschaft mbH*, [2014].

Court concluded that this injunction would give the service provider the possibility to benefit from liability exemption by proving that 'all reasonable measures' are taken.[106] However, two cases of the European Court of Human Rights (ECtHR) showed a different approach with regards to the duties of service providers. The ECtHR does not interpret EU law and therefore this case does not concern the framework in the E-commerce Directive. Yet, these cases are of importance because the ECtHR determines if imposing liability on an online service provider is an interference with the freedom of expression and information and whether that interference is compatible with the Convention. In the case *Delfi AS v. Estonia*, the ECtHR concluded that liability may be imposed on Internet news portals in conformity with article 10 of the Convention[107] if they fail to take measures to remove "clearly unlawful comments" without delay, even without the presence of a notice from the victim or from third parties.[108] The question that arises is how the internet portal should be aware of those clearly unlawful comments. The Court considered the news portal to be an 'active' service provider that had control over the comments section and could have expected to be liable for the unlawful comments.[109] One year later, the ECtHR ruled in the *MTE v. Hungary* case that operators of Internet portals were not liable for offensive, but "not clearly unlawful" user-generated content. The national court had violated the right to freedom of expression since the comments did not amount to hate speech or incitement to violence. It was concluded by the Court that is not necessary to monitor all comments before publication, because 'this amount to requiring excessive and impracticable forethought capable of undermining freedom of the right to impart information on the internet.'[110]

With a view on these cases, it can be deduced that the measures that hosting service providers will take in accordance with the proposed regulation should be effective in removing clearly unlawful comments, but they cannot be required to monitor all comments in order to detect those comments. However, the proactive measures consist of the implementation of automated systems to detect new terrorist content. It is evident that the implementation of monitoring activities will influence the assessment of their knowledge or awareness of illegal facts and circumstances.[111] Those automated systems are able to detect illegal content on their services, so when they do not remove this content, they can be held liable because they had actual knowledge or awareness based on their own research.[112] The promise in recital 5 of the proposed regulation that they will not be considered to be active hosts seems rather meaningless in this regard. Moreover, since hosting service providers must report to the competent authority about the proactive measures taken, this will incentivise them to remove doubtful terrorist content in order to avoid liability.

III.2 Monitoring Obligations

Case law of the CJEU has shown that imposed measures on service providers in the form of an injunction should be proportionate and strike a fair balance between the fundamental rights in question.[113] It is important with regards to proportionality, that the injunction does not impose

[106] Idem, paras. 52-55.

[107] Convention for the Protection of Human Rights and Fundamental Freedoms [1950], CETS No.005.

[108] *Delfi AS v. Estonia* App no.64569/09 (ECtHR, 16 June 2015), para. 159.

[109] Idem, paras. 115, 127.

[110] *Magyar Tartalomszolgáltatók Egyesülete (MTE) and Index.hu Zrt v Hungary* App no. 22947/13 (ECtHR, 2 February 2016), para. 82.

[111] P. Valcke, A. Kuczerawy & P. Ombelet, 'Did the Romans Get It Right? What Delfi, Google, eBay, and UPC TeleKabel Wien have in Common' in M. Taddeo & L. Floridi (eds), *The Responsibilities of Online Service Providers* (Law, Governance and Technology Series 31, Springer International Publishing 2017) p. 114.

[112] Opinion of Advocate-General Jääskinen, supra note 96, paras. 165, 168.

[113] *L'Oréal v. eBay*, supra note 94, paras. 141, 143.

impossible, disproportionate or illegal duties like a general obligation to monitor.[114] This general obligation to monitor is namely prohibited under article 15(1) E-commerce Directive. Following recital 47 of the E-commerce Directive, it is only allowed to impose monitoring obligations in a specific case, particularly when issued by national authorities in accordance with national legislation. This means that it is important to assess whether the imposed measures would oblige service providers to carry out specific or general monitoring. This is briefly addressed in the *L'Oréal v. Ebay* case, where the Court stated that it is contrary to article 15(1) to require service providers to actively monitor all the data of each of its users in order to prevent any future infringement of intellectual property rights via the providers website.[115] The CJEU clarified the specific and general monitoring obligations further in the *Scarlet Extended v SABAM* and *SABAM v. Netlog* cases. The Court had to assess injunctions of national courts that required service providers to implement filtering systems to prevent any future infringements of intellectual-property rights. The Court concluded that such preventive monitoring requires service providers to actively observe all the electronic communication on the service. This would include all users of that service and all information to be transmitted. This injunction has the effect of obliging service providers to carry out general monitoring, which is prohibited by article 15(1) E-commerce Directive.[116]

In comparing that reasoning with the proposed regulation, it is clear that the proactive measures of hosting service providers, including automated means in certain cases, raises questions of legality. Therefore, it is underlined in the impact assessment that accompanies the proposal that the scope of the proactive measures should be proportionate to the level of exposure and economic capacity of the service provider.[117] They are compliant with article 15(1) E-commerce Directive to the extent that these measures are only limited to companies that are exposed to terrorist content, are proportionate and necessary, and amount to specific monitoring obligations.[118]

The proposed regulation underlines in recital 16 that the proactive measures that are taken on initiative of the hosting service provider, will not imply a general monitoring obligation.[119] Indeed, it is not an obligation when hosting service providers are 'voluntarily' conducting such monitoring. The prohibition of a general monitoring obligation is more relevant with regards to the imposed proactive measures in the regulation. When the proactive measures that are taken on own initiative turn out to be insufficient, the competent authority will require the hosting service provider to take ''specific additional proactive measures.'' Initially, these specific measures will be identified in co-operation with each other, but eventually, the competent authority may impose them on the hosting service provider. It is explained in recital 19 of the regulation that these imposed proactive measures 'should not in principle lead to a general obligation to monitor' as prohibited in article 15(1) E-commerce Directive. However, it is clear that when the hosting service provider is automatically detecting and identifying new terrorist content, they are actively filtering all the data of all their users. This is exactly the kind of filtering system that is referred to

[114] Opinion of Advocate-General Jääskinen, supra note 96, paras. 180-82.

[115] *L'Oréal v. eBay*, supra note 94, para. 139.

[116] Case C-70/10, *Scarlet Extended SA v. SABAM* [2011], paras. 35-40; Case C-360/10, *SABAM v. Netlog NV* [2012], paras. 33-38.

[117] European Commission, Commission Staff Working Document, impact assessment, Accompanying the document Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online SWD (2018) 408 final, p. 104.

[118] Idem, p. 47.

[119] Commission proposal 2018, supra note 2, recital 16.

in the *SABAM* cases, which is contrary to article 15(1) E-commerce Directive. Therefore, the Commission explains in recital 19 of the regulation that:

> 'Considering the particularly grave risks associated with the dissemination of terrorist content, the decisions adopted by the competent authorities on the basis of this regulation could derogate from the approach established in Article 15(1) of Directive 2000/31/EC, as regards certain specific, targeted measures, the adoption of which is necessary for overriding public security reasons'. [120]

This recital underlines further that before making that decision, the competent authority should strike a fair balance between the public interest objectives and fundamental rights and provide appropriate justification. In conclusion, by introducing the proactive measures that are imposed on hosting service providers, especially when consisting of automated detection tools, the Commision derogates from article 15(1) of the E-Commerce Directive and the accompanying case law of the CJEU.

IV. Legal Assessment of the Regulation in light of the Fundamental Rights

The regulation on preventing the dissemination of terrorist content online is not only in conflict with the E-Commerce Directive, it is also likely that certain measures shall have an impact on fundamental rights. This section will address the compatibility of the removal order, referral and proactive measures with the freedom to conduct a business and the right to freedom of expression and information.

IV.1 Freedom to Conduct a Business

It is estimated that 30500 hosting service providers offer their services in Europe, varying of file storage and sharing services, online media sharing services, webhosting services, social networking and discussion fora and online market places.[121] Some of these hosting service providers, like *Telegram*, offer encrypted storage of data that cannot be accessed because of the design of the service.[122] It is likely that the obligations that follow from the proposed regulation will affect the freedom to conduct a business for hosting service providers.

With regards to the removal order, it is especially the one-hour timeframe to respond that is expected to impose high compliance costs on the service provider. Also, the referral could have a significant impact on the freedom to conduct a business. This is because they will potentially receive a high number of referrals that they have to assess as a matter of priority. According to the impact assessment, the burden on service providers is not expected to be excessive, since the competent authorities already carried out a legal assessment of the content before issuing the removal order and the costs to comply with the referral will be mitigated by ensuring a high quality of referrals.[123] With regards to the proactive measures of hosting service providers, it can be useful to look at the *Scarlet v. SABAM* case again. The Court stated that national authorities and courts should strike a fair balance between protecting intellectual property rights and other fundamental rights.[124] The Court considered that the injunction that was issued to the service provider required a general monitoring, was not limited in time, and was a preventive measure that targets present

[120] Idem, recital 19.

[121] Commission impact assessment 2018, supra note 117, p. 5.

[122] Idem, p. 14.

[123] Idem, pp. 103, 104.

[124] *Scarlet Extended SA v. SABAM*, supra note 116, paras. 43, 44.

and future infringements.[125] Since this would require service providers to implement a complicated, costly and permanent computer system at their own expenses, this resulted in a serious infringement of the freedom to conduct a business of the service provider, as enshrined in article 16 of the Charter.[126]

In applying this reasoning to the proactive measures in the regulation, it will be especially the preventing of re-uploads and the development of automatic tools to detect new terrorist material that will demand high investment costs for hosting service providers. It is argued in the impact assessment that when these technologies are eventually implemented, the costs per take down are expected to be lower compared to a removal order or notice.[127] In addition, the concerns about compliance costs are mitigated since the economic capacity and the extent of exposure to terrorist content will be taken into account when proactive measures are suggested or imposed by the competent authority.[128]

In my opinion, this does not alter the fact that these high investment costs are difficult to accomplish for small and medium size companies that are exposed to a high number of terrorist content. The Europol Terrorism report of 2018 showed that many online sympathisers of IS migrated to smaller platforms with less capacity to detect or remove terrorist content.[129] Together with the requirement that these automatic detection tools should be accompanied by human review, the costs will increase exponentially.[130] This creates a tension between the effective measures that are needed to prevent terrorist content and the optimal functioning of the digital single market. These measures will directly diminish the resources of companies to innovate and grow and it is unavoidable that this will make it more difficult for new hosting service providers to enter the market. This is contrary to one of the pillars of the Digital Single Market strategy, to create the right conditions for innovation, investment, fair competition and a level playing field for digital networks and services.[131] However, it is stated in the impact assessment that the occurrence of terrorist content online also has a negative effect on trust, innovation and growth in the Digital Single market.[132] It is expected by the Commission that the obligations on service providers will incentivise them to establish new technologies that facilitate compliance. Moreover, the establishment of such 'robust mechanisms' in removing terrorist content online will reinforce trust in their services, what will lead to an increase of investment and users.[133] An example that confirms that reasoning is the recent scandal involving YouTube where major brands like Pepsi and Walmart pulled back after their advertisements were displayed next to videos promoting extremist views or hate speech.[134] Still, it remains the question what the effects will be in practice, but in my opinion, it would be beneficial to provide technical or financial assistance to companies with smaller resources that face a high exposure to terrorist content. This will improve effective

[125] Idem, para. 47.

[126] Idem, para. 48.

[127] Commission impact assessment 2018, supra note 117, p. 105.

[128] Commission proposal 2018, supra note 2, art. 6 and recital 18; Commission impact assessment 2018, supra note 117, p. 47.

[129] Europol 2018, supra note 6, pp. 31-32.

[130] Commission impact assessment 2018, supra note 117, p.105.

[131] European Commission, 'Digital Single Market strategy for Europe' (Communication) COM (2015) 192 final, p. 3.

[132] Commission impact assessment 2018, supra note 117, p. 17.

[133] Idem, p. 37.

[134] O. Solon, 'Google's bad week: YouTube loses millions as advertising row reaches US', *The Guardian,* (25 March 2017) at: https://www.theguardian.com/technology/2017/mar/25/google-youtube-advertising-extremist-content-att-verizon (accessed on 28 April 2018).

action against terrorist content while creating a level playing field for digital services in the digital single market.

IV.2 The Right to Freedom of Expression and Information

Before turning to the proposed regulation and the possible impact on freedom of expression and information, it is useful to understand how hosting service providers are monitoring content on their platforms.

IV.2.1 Content Moderation in General

Due to a rapid increase of users all over the world and the need to rely on human moderators with diverse background, companies like YouTube, Facebook and Twitter have developed a system of rules based on which they moderate content. On the one hand, social media companies are preventing content from publication on their platform based on their *ex ante* content moderation policy.[135] This means that they can prevent that certain content is published on their platform, because it violates laws in the country where the content provider is located.[136] They can also prevent publication when certain content is violating the platform's own terms and conditions. The massive flow of content that is being uploaded on the platform is monitored through an algorithmic system that is translating the data into certain outputs, what makes it possible that the publication of content is automatically declined or that the content needs interpretation by a human moderator.[137] This type of content moderation may for example be used to prevent the re-upload of illegal content that has been identified before.

The removal of content or blocking access to content on the platforms also takes place through *ex post* content moderation, either reactively when external flagged content is reviewed against internal guidelines, or proactively by using algorithms and human moderators.[138] This type of content moderation is increasingly used to act upon hate speech or content that incites to extremist violence, and this is partly a result of pressure from governments and the EU on intermediaries to change their terms and conditions and enhance actions to prevent illegal content on their platform. According to Daniele Citron (2018), it can be beneficial to request companies to change the terms and conditions to target illegal content. For example, cruel beheading videos, bombing instructions, live streams of terrorist attacks or messages that directly incite to violence against target groups are not protected by the right to freedom of expression, and the removal of this content may contribute to a diminishing of terrorism or other violent acts.[139] Yet, Citron (2018) argues that while the initiatives of intermediaries to act against extremist content seem initially positive, it is observed that companies are increasingly prohibiting speech in their terms and conditions that goes further than incitement to violence against groups or incitement to terrorist offences. This tendency to expand speech policies beyond the original goal of the speech regulation is named 'censorship creep' and creates risks for global freedom of expression.[140] This censorship creep is exacerbated by vague definitions of terrorist, extremist or hate speech, but it is also a result of the lack of transparency and accountability of removal policies.[141]

[135] Klonick 2018, supra note 104, pp. 1635-7.

[136] Idem, pp.1636-7.

[137] General Assembly, Report of the Special Rapporteur on freedom of expression David Kaye, submitted to the 73rd session of the General Assembly, UN Doc. A/73/348, 29 August 2018, para 4.

[138] Klonick 2018, supra note 104, p. 1638, 1639.

[139] D.K. Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review* 2018-93(3), pp. 1049, 1050.

[140] Idem, p. 1051.

[141] Idem, pp. 1666-7.

It is clear that social media companies are in the position to monitor and control content on their platforms as real governors. Therefore, it is crucial that their content moderation policy and their terms and services are set and applied with respect for the fundamental rights of their users. With regards to the proposed regulation, hosting service providers are required to implement the definition of terrorist content in their terms and conditions. It will now be assessed to what extent that definition is able to sufficiently protect the right to freedom of expression.

IV.2.2 Applying the Definition of Terrorist Content

The extensive use of social media by the Islamic State, and the influence thereof on lone actor terrorist attacks, recruitment and foreign fighters have led to an expansion of offences based on opinions and ideas that are deemed to be dangerous to society.[142] This reflects a more proactive and preventive approach in counter-terrorism context, with the aim of intervening in the pre-stage of terrorism. Nevertheless, by expanding the limitations of the right to freedom of expression there is a violation of article 10 ECHR and article 11 of the Charter. According to the ECHR, freedom of expression should be interpreted broadly and it is also applicable to information or ideas that offend, shock or disturb the State or any sector of the population, without which there is no democratic society. [143] Based on Article 10.2 ECHR, this implies that a measure to block access to content through filtering or removal of content must be prescribed by law, pursuing a legitimate aim and be necessary in a democratic society. Furthermore, there must be a pressing social need to restrict this right and this restriction must be proportionate to the legitimate aim pursued.[144] With regards to necessity, it is stated by the ECHR that this does not imply the balancing of two conflicting principles, such as freedom of expression and protecting public security. It rather means that freedom of expression is subject to a number of exceptions that must be narrowly interpreted.[145]

With regards to the proposed regulation, the main question is whether the definition of terrorist content shall result in disproportionate or unnecessary interferences with the right of freedom of expression or not. In article 2 of the regulation, the Commission introduces four different categories of information that constitutes as terrorist content. The first category is: 'inciting or advocating, including by glorifying, the commission of terrorist offences, thereby causing a danger that such acts be committed'. The second category is: 'encouraging the contribution to terrorist offences' and the third: 'instructing on methods or techniques for the purpose of committing terrorist offences'.[146] These categories of offences are similar to the definitions of public provocation, recruitment and training in the Directive on Combatting Terrorism. However, they go beyond the scope of that Directive, because the requirement of intention is eliminated from this definition. Also, the Commission added a new category of terrorist content: information that is *'promoting the activities of a terrorist group, in particular by encouraging the participation in or support to a terrorist group'.*[147]

Firstly, this definition is broad because instead of focussing on content that directly incites to the commission of terrorist offences, it entails the notion of "glorification". Various intergovernmental organisations stressed that such terms are too vague in the context of restricting

[142] H. Duffy & K. Pitcher, 'Inciting Terrorism? Crimes of Expression and the Limits of the Law' *Grotius Centre Working Paper Series* 2018/076-HRL, p. 3.

[143] *Handyside v the United Kingdom* (1976) Series A no. 24, para. 49.

[144] Ibid.

[145] *Sunday Times v. United Kingdom* (1979) Series A no. 30, para. 65.

[146] Commission proposal 2018, supra note 2, art. 2, recital 9; Combatting Terrorism Directive, supra note 56, art. 5-8.

[147] Commission proposal 2018, supra note 2, art. 2, recital 9.

speech.[148] Secondly, this definition is problematic because it also entails content that is promoting, encouraging and supporting a terrorist group, without the element of incitement. This is problematic because in general, online extremist or terrorist propaganda should be distinguished from material that incites acts of terrorism. With a view to the right to freedom of expression and the permissible limitations, it can be deduced that the dissemination of terrorist propaganda itself is not restricted, but when it comprises incitement to violence that clearly forms a threat to security, an interference with that right will sooner be legitimate.[149] For example, the ECtHR concluded that the prosecution of someone who expressed respect for the leader of a terrorist organisation, without inciting to violence, constituted an unjustified interference with the freedom of expression.[150] Thirdly, in applying this definition, hosting service providers do not have to assess whether someone intentionally glorifies or promotes terrorist offences that creates a danger that terrorist offences may be committed. Rather, they should assess if that content is "causing a danger" that an offence may be committed. It can be concluded that the mental element of the content provider is not important in this regard. To the contrary, the assessment is solely focussed on the possible effects of the content. However, with respect to principles of legality and certainty, it is dissuaded to create laws like this, that prohibit supporting or encouraging terrorism by requiring that the conduct causes danger that offences may be committed.[151] In order to prevent disproportionate interferences with the right to freedom of expression, these offences must be clearly defined and pursue a double requirement of intent: the intention to distribute messages and the intention to incite the commission of a terrorist offence.[152] The special rapporteurs that reviewed the proposal for regulation also expressed the concern that the definition of terrorist content lacks clarity about whether terrorist content amounts to criminal conduct, because conduct and intent form the basis of legal responsibility. They were concerned that this definition will result in restricting lawful content such as reports, articles and messages from journalists, human right organisations and academics that discuss terrorist groups and counter-terrorism measures.[153]

It follows from case law of the ECtHR that in assessing whether the interference to the freedom of expression is proportionate, the content of the statement and the context in which they are made should be taken into account.[154] This is also visible in recital 9 of the proposed regulation, that lists important factors that may lead to the conclusion that it constitutes terrorist content: the wording, nature and context of the statements, the fact that the content can in any way be linked

[148] General Assembly, Report of the Secretary-General, The Protection of Human Rights and Fundamental Freedoms while Countering Terrorism, UN Doc. A/63/337, 28 August 2008, para 61; Duffy & Pitcher 2018, supra note 142, p. 7.

[149] Provided that it is prescribed by law and necessary conform article 10 ECHR; Duffy & Pitcher 2018, supra note 142, pp. 27, 28.

[150] *Yalçinkaya and others v Turkey* (ECtHR, 1 October 2013), para 34.

[151] UN, 'Joint letter to the relevant European Union organ in advancing full respect for human rights in regulation of preventing the dissemination of terrorist content online, mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression; the Special Rapporteur on the right to privacy and the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism', OL OTH 71/2018, 7 December 2018, p. 3.

[152] Human Rights Committee 102nd session, General Comment No. 34, CCPR/C/GC/34, 12 September 2011, para 46; Human Rights Council, 16th session, Report of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, Martin Scheinin, A/HRC/16/51, 22 December 2010, practice 8.

[153] UN Special Rapporteurs 2018, supra note 151, p. 3.

[154] *Ceylan v Turkey* (ECtHR, 1999-IV), para. 32.

to a terrorist organisation or person and the potential that this content will lead to harmful consequences.[155] Moreover, it is underlined in the regulation that terrorist content does not include the expression of radical, polemic or controversial views.[156] Yet, in my opinion it will be difficult to apply vague definitions as glorifying, promoting and advocating without the element of intention in a situation where contextual factors are decisive. Also, by focussing on the vague notion that it may cause danger, the boundaries of this definition will be stretched. The results may be different per case, depending on the conclusion of an algorithmic screening and the interpretation of a human content moderator.[157]

IV.2.3 Content Moderation in Accordance with the Regulation

The measures in the proposed regulation requires the application of *ex ante* and *ex post* content moderation, what implies that hosting service providers have to act both before and after the publication of content on their platforms. It is possible that in conducting these content moderation policies, the right to freedom of expression and information is affected. In the impact assessment it is acknowledged that there is a risk of removing lawful content, and in order to prevent that, hosting service providers must include provisions in their terms and conditions to prevent the dissemination of terrorist content and apply them in a diligent, proportionate and non-discriminatory manner.[158] In general, they have to remove content in the observance of the freedom of expression and information.[159]

The procedure of the referral is an example of *ex post* reactive content moderation of social media platforms. It is especially useful and effective for public authorities to relocate the responsibility to private actors in acting against illegal content on their platforms. State actors can avoid political debates about the specific conditions of restricting speech and judicial hearings about interferences with fundamental rights. Private content moderation also allows more censorship than law would allow.[160] Article 52(1) of the Charter reads that an interference with the rights in the Charter must be provided for by law, respect the essence of those rights and be necessary and proportionate.[161] This means that in the situation that the competent authority in a member state orders the removal of lawful content or content that does not define as terrorist content, this results in an interference of the right to freedom of expression or information. It is likely that certain online content balances on the edge of qualifying as illegal or terrorist content. Therefore, it is useful for authorities to place the responsibility to tackle this type of content with the hosting service provider by means of a referral.

It is acknowledged in the impact assessment that referrals might address content that is not illegal and because the referral is not legally binding, it cannot be challenged in a court. However, it is argued that the referrals from Europol are not likely to contain content that includes protected

[155] Commission proposal 2018, supra note 2, recital 9.
[156] Ibid.
[157] Citron 2018, supra note 139, pp. 1053-5.
[158] Commission proposal 2018, supra note 2, art. 3, recital 12.
[159] Idem, recital 12.
[160] Citron 2018, supra note 139, pp. 1061, 1062.
[161] Charter of Fundamental Rights of the European Union, Art. 51 (1): 'Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others.'

speech, because Europol can only act within its mandate.[162] Nevertheless, the regulation does not provide an explanation why especially the hosting service providers should make a decision based on their terms and service to remove the referred content. The competent authority in the Member State and Europol have access to exactly the same information as the hosting service provider and they are legally bound to protect the fundamental rights of the users. It is true that also private actors should avoid infringing on the human rights of others and should address adverse human rights impacts of their actions[163], but if a private actor issues a decision and interferes with freedom of expression, this is not seen as an interference that is prescribed by law. Therefore, it is compatible with article 52 (1) of the Charter. According to the special rapporteur that reviewed the proposed regulation, the procedure of a referral gives public authorities more possibilities to have content blocked or removed and the lack of clear redress mechanisms creates risks for abuse and arbitrariness.[164] Moreover, it is argued by various authors that internet intermediaries are not the right actor to decide on the illegality of noticed content, since the liability framework in the E-Commerce Directive incentivises them to systematically over-block and remove lawful or doubtful content on their services. This is also identified as private censorship or collateral censorship, and this clearly has consequences for freedom of expression and information.[165]

Also, the proactive measures that are set out in the regulation are concerning with regards to freedom of expression. These measures consist of checking content against databases with terrorist content and using automatic tools to prevent the reupload of known terrorist content and detect new terrorist content. The hosting service providers are able to conduct this content moderation by using algorithmic filtering systems, and this can be considered as a filtering system as meant in the *SABAM* cases. Yet, the Court stated in the *Scarlet extended v. SABAM* case that there should be a fair balance between protecting intellectual property rights and other fundamental rights.[166] The Court argued that because the filtering system cannot distinguish adequately between unlawful and lawful content, there is a risk of removing lawful content, and this undermines the right to freedom of expression and information of the users.[167]

Likewise, according to the report of the Council of Europe, the algorithms that internet companies are using to facilitate the detection of illegal content are unable to detect cultural, language or gender-based contexts and sensitivities and they cannot identify public interest in the content.[168] Despite the fact that article 10 ECHR also protects shocking, offensive or disturbing

[162] Commission impact assessment 2018, supra note 117, p. 104; This mandate of Europol is laid down in the Europol Regulation where it is written that Europol shall perform the following tasks as set out in art. 4(m): 'support Member States' actions in preventing and combating forms of crime listed in Annex I which are facilitated, promoted or committed using the internet, including, in cooperation with Member States, the making of referrals of internet content, by which such forms of crime are facilitated, promoted or committed, to the online service providers concerned for their voluntary consideration of the compatibility of the referred internet content with their own terms and conditions.'

[163] OHCHR, 'Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework', 2011, Geneva, United Nations, p. 13.

[164] UN Special Rapporteurs 2018, supra note 151, p. 8.

[165] J.M. Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *U.C. Davis Law Review* 2017-3, p. 1176; E. Coche, 'Privatised Enforcement and the Right to Freedom of Expression in a World Confronted with Terrorism Propaganda Online', *Internet Policy Review* 2018-7(4), pp. 6, 7; A. Kuczerawy, 'Intermediary liability & freedom of expression: Recent developments in the EU notice & action initiative' *Computer Law and Security Review* 2015-31(1), p. 48.

[166] *Scarlet Extended SA v. SABAM*], supra note 116, paras. 43, 44.

[167] Idem, paras. 52, 53.

[168] Council of Europe, supra note 54, p.20.

content, these algorithmic detection and removal tools will likely detect this content and this can result in over-removal.[169] The European Commission acknowledges in the impact assessment that the compatibility of the proactive measures with the freedom of expression and information would depend on the accuracy of automated detection tools and how well they can avoid 'false positives'. With respect to that, the Commission argues that the technology evolves fast, and recent developments have shown that the error rates for these false positives are decreasing.[170]

With regards to the proportionality of the interference, it follows from the impact assessment that these measures are accompanied with safeguards, which are expected to mitigate the impacts on freedom of expression and information.[171] Indeed, it is laid down in the regulation that hosting service providers should ensure the right to freedom of expression and information, they have to act with due diligence and implement safeguards such as human oversight and verifications where appropriate, to avoid erroneous decisions to remove content that is not terrorist content.[172] Nevertheless, apart from the fact that algorithmic systems cannot detect the context, they are also designed and used in a certain social or organisational context, and this can produce biases.[173] Algorithmic decision-making systems are based on correlation between types of data and are programmed to enhance efficiency. The algorithms predict for example which individuals are likely to commit a crime and this increases the risk of stereotyping and indirect discrimination.[174] Thus, when online platforms are removing online content or suspend user accounts in accordance with systems that incorporate such biased concepts, they disadvantage vulnerable groups in society by restricting their freedom of expression.[175] The biased algorithms and removal policies of social media companies are for example not evenly targeting jihadist content and extreme-right content and this will have the effect of hindering a pluralist public debate that is equally accessible and inclusive to everyone.[176]

Following article 21 of the Charter discrimination on any ground is prohibited, and it is therefore necessary to design the algorithmic systems in a right way, in order to prevent unjustified differential treatments in the first place.[177] This is especially important because algorithms are not able to counteract the learned biases, unlike human beings.[178] These concerns were also expressed by academics in their feedback on the inception impact assessment where they urged that there is need of specific safeguards for algorithmic decision-making to prevent that biases and discrimination lead to erroneous decisions. Algorithmic systems should not for example target all Arabic posts and legitimate speech on Islam. It is important to prevent that 'catch-all' moderation systems are build that are overly preventive for specific communities.[179] An illustrating example is the removal of content and accounts on Facebook of the Rohingya, a Muslim ethnic group in

[169] Ibid.

[170] Commission impact assessment 2018, supra note 117, p. 105.

[171] Idem, p. 106.

[172] Commission proposal 2018, supra note 2, art. 9, recital 17.

[173] Council of Europe 2018, supra note 54, pp. 37, 38.

[174] Ibid, pp. 10, 11.

[175] Special Rapporteur David Kaye 2018, supra note 137, para. 15; Council of Europe 2018, supra note 54, p. 27.

[176] "How the authorities are failing to tackle far-right terror online" *Prospect*, 27 March 2019 at https://www.prospectmagazine.co.uk/other/authorities-failing-far-right-terror-internet-online-white-supremecy-home-office-twitter-facebook, (accessed on 28 April 2019); J.M. Berger, '"The Alt-right Twitter Census" Defining and describing the audience for Alt-right content on Twitter', *VOX-Pol* 2018, p. 55.

[177] Council of Europe 2018, supra note 54, p. 28.

[178] Idem, p. 27.

[179] Commission impact assessment 2018, supra note 117, pp. 15, 67, 68.

Myanmar. The Rohingya refugees recorded the ethnic cleansing in their communities and shared it on Facebook. The algorithms of the content moderation policy of Facebook were designed to detect Muslim extremism and this content was considered to be dangerous.[180] Another example of a filtering system that led to erroneous removals is the case of YouTube. In preventing and acting upon extremism on their platform, they use technology that is especially focused on the Islamic State and their sympathisers. The algorithms in this system are flagging and removing extremist videos, sometimes without human review. This content moderation policy has led to the removal of videos of the conflict in Syria, while many of these videos were meant to document the atrocities and war crimes, that could have been used for future prosecutions.[181]

There are many examples of policies from social media platforms that have had detrimental consequences. Therefore, transparency of platforms' content and removal policies are needed to assess the problem of illegal content and the impacts of the measures that are taken.[182] There should be transparency about the variables that are used, the reasons for using or changing the algorithms, the procedure that the algorithms follow, and the type of data that is processed. Unfortunately, there doesn't exist a regulatory framework that ensures that the programming of algorithms themselves and the results they produce are sufficiently protecting fundamental rights and ethical principles.[183]

In conclusion, the Commission places with this regulation a large responsibility for private actors to control the online environment in order to prevent terrorist content online. Yet, as this section made clear, it is likely that the application of the measures in accordance with this regulation will have a considerable impact on the fundamental rights of companies and their users. Moreover, the safeguards that are forwarded in the regulation are not expected to mitigate this impact sufficiently.

VI. Conclusions

The use of the internet by terrorist groups and organisations is evident, but the effects of that content on their audience are not so clear. It can be concluded that the internet and social media facilitate the radicalisation of individuals by providing possibilities to develop a collective identity that may result in violent action, especially when content providers are inciting to violence. Yet, not all extremists and terrorists are influenced by online terrorist propaganda or are actively communicating with other extremists online. The radicalisation process and turn to violence is dependent on many factors. Therefore, the influence of online terrorist content on violent radicalisation cannot be generalised for every type of terrorist. In order to effectively counter radicalisation and terrorism, there should be more attention to the differences in radicalisation processes, learning and planning of attacks.

While social media platforms are already removing extremist and terrorist content from their platforms, the unintended effects related to the removal of all radical or extremist ideas from social media platforms should not be overlooked. When all terrorist and extremist content is

[180] A.F. Cahn, 'Why is Facebook censoring Rohingya accounts of the Genocide?', *Newsweek* (10 February 2017) at: https://www.newsweek.com/why-facebook-censoring-rohingya-accounts-genocide-675526 (accessed on 28 April 2019).

[181] M. Browne, 'YouTube removes videos showing atrocities in Syria', *New York Times* (New York, 22 August 2017) at: https://www.nytimes.com/2017/08/22/world/middleeast/syria-youtube-videos-isis.html (accessed on 28 April 2019).

[182] Commission impact assessment 2018, supra note 117, p. 15.

[183] Council of Europe 2018, supra note 54, p. 40.

removed from Facebook or YouTube, it will not be possible to follow and monitor radicalising individuals on social media and it will be likely that extremist and terrorist communication eventually relocates to closed online environments where the radicalisation process possibly accelerates. At the same time, the removal of extremist and terrorist content on social media platforms interferes with online deradicalisation strategies, because it takes away the opportunity to dismantle extremist messages, provide for counter narratives or start discussions with radicalising individuals.

In this article, I discussed that the proposed regulation to prevent the dissemination of terrorist content online is problematic with a view on the E-Commerce Directive. For European regulators and national authorities, it might be convenient to relocate the responsibility to target illegal content to social media companies. However, in my opinion, in delegating obligations to hosting service providers, the limitations set out in the E-Commerce Directive should be respected. While the European Commission underlines that the hosting service provider shall not be seen as an active service provider in the sense of the E-Commerce Directive, this is not convincing when they implement the proactive measures in accordance with the regulation. Full compliance with the procedures in this regulation shall give the hosting service provider actual knowledge or awareness of illegal activity and when this content is not removed this will result in liability. In addition, when authorities impose the implementation of automatic tools to detect new terrorist content, this constitutes a general obligation to monitor as prohibited under article 15 of the E-Commerce Directive. This is especially problematic because this article aims to protect the rights of the service provider and the rights of the internet users.

Moreover, the proposed regulation is clearly affecting the fundamental rights of hosting service providers and internet users. First, the right to conduct a business will be impacted by the high compliance costs. I find the mitigating factors that are forwarded in the regulation not convincing, especially not with regards to small or medium sized companies that face a high exposure of terrorist content. It will be beneficial to provide assistance in the form of financial resources or technical expertise to such smaller service providers in order to tackle terrorist content on their platforms. This is necessary with regards to the effective functioning of the Digital Single Market, because the penalties that are introduced in the regulation will decrease the competitiveness of these smaller companies, and this will give the major social media platform the possibility to dominate the market even more.

Second, this regulation will highly affect the freedom of expression of content providers and internet users in general. On the one hand, the definition of terrorist content is too vague. It is especially difficult for private actors to assess whether online content causes a danger that terrorist offences may be committed. Given the fact that intention is removed from this definition, this will result in the removal of legal content. On the other hand, the regulation enlarges the risk of liability for companies and together with the high penalties in case of non-compliance, this will result in collateral censorship. The consequence of over-removal is that the content provider perceives it as being silenced, and this will diminish their trust in the digital platforms, but also in society in general. Moreover, the mandatory proactive measures that consist of automatic tools to detect and remove content that conflicts with the companies' terms and conditions will result in erroneous removals and have discriminatory effects. At this moment, companies are using biased algorithms and biased data sets to detect illegal content. The examples of discriminatory private censorship showed that companies are not always in control of the content moderating algorithms. While it is likely that the regulation will result in the removal of violent Jihadist content, it will probably also result in the over-removal of radical ideas, which are actually protected by the right to freedom of expression. There is a risk that if those biased algorithms are not modified or replaced, this will leave those who express approval for extreme right-wing ideologies and violence untouched. It is disappointing that the European Commission does not pay enough attention to the clear risks of algorithmic decision making. In my view, companies should be clear about the design of the algorithm and they should reassure that the algorithm is applied and interpreted in a right way. It would also be beneficial to introduce supervision on the algorithmic systems that companies are implementing and interpreting when they act in accordance with the regulation.

In conclusion, the popularity of far-reaching, but unsubstantiated measures to counter terrorism remains persistent in the current political environment. The shortcomings of the proposed regulation as identified in this article are partly a result of the pressure on the European Commission from the side of different European bodies and politicians in the Member States, but it is also a result of a balance between different interests. In my view, the European legislators should take the protection of fundamental rights of users and companies seriously and especially in the area of public security it is important to avoid excessive, unnecessary and ineffective measures. There should be made a fair balance between fundamental rights and public security objectives and it is necessary that the responsibilities are equally distributed among public and private actors. Yet, the legislators should not disregard the detrimental consequences of algorithmic detection tools and risks of privatised law enforcement. While I find it positive that the European Union attempts to regulate the internet by acting upon harmful online content, the costs of the proposed regulation in its current form are far higher than the expected benefits. With a view on the legislative process of this regulation, the European Parliament adopted its position on the proposed regulation and some of their suggestions are beneficial with regards to the fundamental rights of companies and internet users.[184] It depends on the result of the trilogue negotiations of the European Parliament, the Council of the EU and the European Commission whether this regulation evolves as an effective and appropriate response to terrorist content online.[185]

[184] European Parliament legislative resolution of 17 April 2019 on the proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online (COM(2018)0640 – C8-0405/2018 – 2018/0331(COD)), At: http://www.europarl.europa.eu/RegData/seance_pleniere/textes_adoptes/provisoire/2019/04-17/0421/P8_TA-PROV(2019)0421_EN.pdf (accessed on 28 April 2019).

[185] The triologue negotiations are starting in september 2019. Follow the ongoing steps in this legislative procedure at: https://eur-lex.europa.eu/procedure/EN/2018_331 (accessed on 11 July 2019).