The proposed Regulation on preventing the dissemination of terrorist content online: safeguards and risks for freedom of expression

Dr. Aleksandra Kuczerawy

Centre for IT & IP Law (CiTiP), KU Leuven, Belgium

For the Center for Democracy and Technology[1]

5 December 2018

[1] The research leading to this paper has received funding from the Center for Democracy and Technology (CDT).

Introduction

In September 2018, the European Commission released its proposal for a Regulation on preventing the dissemination of terrorist content online.[2] As stated in the Explanatory Memorandum, terrorists '*misuse the internet to groom and recruit supporters, to prepare and facilitate terrorist activity, to glorify in their atrocities and urge others to follow suit*'.[3] The misuse, as further explained, highlights the "particular societal responsibility" of internet platforms to protect their users from exposure to terrorist content. The Explanatory Memorandum notes that several hosting service providers have put in place certain voluntary measures to tackle terrorist content on their services. The measures, however, are not considered sufficient. The proposal, therefore, attempts to address the problem of terrorist content disseminated online by imposing certain obligations on hosting service providers.[4] As set out in Article 1, the proposal lays down: 1) rules on duties of care to be applied by hosting service providers to prevent the dissemination of terrorist content and ensure its swift removal; and 2) a set of measures to be put in place by Member States to identify terrorist content, to enable its swift removal by hosting service providers and to facilitate cooperation between stakeholders.[5] In particular, the proposal describes several methods to prevent dissemination of terrorist content online, for example, content referrals, removal orders, and proactive measures. Importantly, the proposal also includes a number of safeguards that attempt to prevent disproportionate interference with fundamental rights, in particular the right to freedom of expression and access to information.

To give the proposed Regulation "teeth", Member States are required to introduce effective, proportionate and dissuasive penalties for breaches of the obligations by hosting service providers. In case of a systematic failure to comply with removal orders within the timeframe of one hour, hosting service provider shall be subject to financial penalties of up to 4% of the global turnover of the last business year.[6]

The proposed Regulation is yet another step by the EU in a series of initiatives addressing the dissemination of illegal content online. For example, the proposal complements the provisions of the newly amended Audio-visual Media Services Directive, which requires video-sharing and possibly social media platforms to restrict access to harmful content (to

---

[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, Brussels, 12 September 2018, COM(2018) 640 final; https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-regulation-640_en.pdf.

[3] *Ibid*. p. 1.

[4] Article 2(1) of the proposal defines 'hosting service providers' as providers of information society services that store information provided by and at the request of the content provider and make the information available to third parties.

[5] The proposed Regulation (n 2), Article 1.1.

[6] The proposed Regulation (n 2), Article 18.

protect minors) as well as to content that incites violence or hatred or contains public provocation to commit a terrorist offence (to protect all citizens).[7] Other initiatives include the EU Internet Forum against Terrorism[8], the Code of Conduct on Countering Illegal Hate Speech Online[9], the Communication on Tackling Illegal Content Online[10], and the Recommendation on measures to effectively tackle illegal speech online[11]. The European Commission chose a binding instrument this time instead of a soft-law approach, in contrast to the previous initiatives. The initiatives show a steady shift from intermediary liability to intermediary responsibility. Hosting service providers are increasingly expected to take proactive measures to regulate content uploaded by their users.[12] The new approach represents a significant departure from the approach taken by E-Commerce Directive 2000/31/EC[13].

The aim of this paper is to analyse the proposed regulation for the prevention of online terrorist content from the perspective of the current legal framework, including the EU Charter of Fundamental Rights, and the accompanying CJEU case law. The analysis provided here is not intended to be exhaustive. Rather, the purpose of the paper is to highlight specific elements of the proposal that raise particular concerns from the perspective of the existing framework.

On 3 December 2018, the EU Council released its version of the proposal (general approach).[14] The paper focuses on the Commission's proposal but points out relevant changes in the Council's general approach.

---

[7] Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities PE/33/2018/REV/1, OJ L 303, 28.11.2018, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2018.303.01.0069.01.ENG&toc=OJ:L:2018:303:TOC.

[8] See the EU Internet Forum against Terrorism: Bringing together governments, Europol and technology companies to counter terrorist content and hate speech online, Brussels, 3 December 2015, http://europa.eu/rapid/press-release_IP-15-6243_en.htm.

[9] The Code of Conduct on Countering Illegal Hate Speech Online, http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf.

[10] European Commission, Commission Communication to the European Parliament, The Council, The Economic and Social Committee and The Committee of Regions, Tackling Illegal Content Online - Towards an enhanced responsibility of online platforms, Brussels, 28 September 2017, COM(2017) 555 final, https://ec.europa.eu/digital-single-market/en/news/communication-tackling-illegal-content-online-towards-enhanced-responsibility-online-platforms.

[11] European Commission, Recommendation on Measures to Effectively Tackle Illegal Content Online, Brussels, 1 March 2018, C(2018) 1177 final, https://www.isdc.ch/media/1585/8-commissionrecommendationonmeasurestoeffectivelytackleillegalcontentonline-1.pdf.

[12] See more in G. Frosio, The Death of 'No Monitoring Obligations': A Story of Untameable Monsters, 8 (2017) JIPITEC p. 199, https://www.jipitec.eu/issues/jipitec-8-3-2017/4621/JIPITEC_8_3_2017_199_Frosio.

[13] Directive 2000/31/EC of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), O.J. L 178, 17 July 2000.

[14] Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council

Relation to the E-Commerce Directive

The E-Commerce Directive (ECD) governs the liability exemptions for Internet intermediaries for content disseminated by their users. In principle, the Directive applies horizontally to various domains and any kind of illegal or infringing content. It provides liability exemptions for three groups of Internet intermediaries depending on the type of service they provide: mere conduit, caching, or hosting. Under Article 14 ECD, hosting providers can benefit from a liability exemption provided they act expeditiously to remove or disable access to information upon obtaining knowledge about its illegal character. The ECD introduces different levels of knowledge, "actual knowledge" required for criminal liability and "constructive knowledge" (awareness of the facts or circumstances from which the illegal activity or information is apparent) for civil liability. To benefit from the liability exemption, a service provider's conduct must be neutral. The CJEU has described neutrality as conduct that is '*technical, automatic and passive, pointing to a lack of knowledge or control of the data which it stores*'.[15] The CJEU further clarified the neutrality requirement in *L'Oréal v. eBay*, by stating that Article 14 of the Directive applies to hosting providers as long as they do not play an active role that would allow them to have knowledge or control of the stored data.[16] The liability exemption does not, however, affect the power of courts or administrative authorities to issue prohibitory injunctions in accordance with the national legal system. It also does not prevent Member States from establish specific procedures governing the removal or disabling of access to information. [17]

Under Article 15 ECD, Member States may not impose on providers of intermediary services a general obligation to monitor information they transmit or store. The same provision states that they cannot introduce a general obligation to actively look for facts or circumstances indicating illegal activity. The CJEU analysed Article 15 in the context of copyright infringing content on hosting services in *Sabam v. Netlog*.[18] The CJEU ruled that requiring a filtering mechanism that would oblige hosting providers to actively monitor almost all the data relating to all of its service users in order to prevent any future infringements would constitute imposing general monitoring, which is prohibited by Article 15 of the E-Commerce Directive.[19]

The proposed regulation on preventing the dissemination of terrorist content online refers to the ECD on several occasions. First, in Recital (5) the proposal specifies that the

---

on preventing the dissemination of terrorist content online – general approach, 14978/18, Interinstitutional File: 2018/0331(COD), 3 December 2018, https://www.parlament.gv.at/PAKT/EU/XXVI/EU/04/57/EU_45743/imfname_10862334.pdf.

[15] CJEU, *Google France and Google v. Louis Vuitton Malletier a.o.*, Joined Cases C-236/08 to C-238/08, 23 March 2010, paras. 113-114.

[16] CJEU, *L'Oréal v. eBay*, Case C324/09, 12 July 2011, paragraphs 112 – 116.

[17] Article 14.3 ECD.

[18] CJEU, *SABAM v. Netlog*, C-360/10, 16 February 2012.

[19] *Ibid.*, para. 38.

application of the Regulation should not affect the application of Article 14 ECD. In particular, the Recital explains, any measures taken by the hosting service provider in compliance with the Regulation '*should not in themselves lead to that service provider losing the benefit of the liability exemption*' in Article 14 ECD. Interestingly, the clarification extends to any proactive measures that may be taken by the hosting providers, which is significant. It illustrates the Commission's attempt to convince hosting service providers that taking proactive measures would not render them "active" hosts, who might lose the immunity afforded by Article 14 ECD.[20] The question remains, however, how the clarification sits with the existing jurisprudence of the CJEU. Carving out one type of content, specifically terrorist content, from the CJEU's interpretation of the neutrality requirement may seem arbitrary and unconvincing.

Article 6 of the proposed regulation provides that hosting service providers should take '*proactive measures to protect their services against the dissemination of terrorist content*'. Recital (19) adds that imposing *'specific proactive measures should not, in principle, lead to the imposition of a general obligation to monitor',* as prohibited by Article 15 ECD. After this optimistic note, things get confusing. The proposal explains that in light of the '*particularly grave risks associated with the dissemination of terrorist content',* the decisions adopted on the basis of the Regulation could, in fact, derogate from the prohibition set in Article 15 ECD. The derogation would apply to '*certain specific, targeted measures, the adoption of which is necessary for overriding public security reasons*'. It would seem, therefore, that the general monitoring obligation may very well be the intended outcome of the Regulation, in contradiction with Article 15 ECD (see more on proactive measures below).[21]

Recital (19) represents a major shift in the approach towards the obligations of online hosting services. Numerous guidelines and recommendations in the area of intermediary liability strongly advise against such an approach. For example, the Council of Europe Recommendation on the roles and responsibilities of internet intermediaries provides that State authorities '*should not directly or indirectly impose a general obligation on intermediaries to monitor content which they merely give access to, or which they transmit or store, be it by automated means or not*'.[22] The UN Special Rapporteur David Kaye, in his report on the promotion and protection of the right to freedom of opinion and expression, clarifies that States should refrain from establishing laws or arrangements that would

[20] A similar line of reasoning could be found in the EC Communication on Tackling Illegal Content (n 9), p. 10.
[21] The general approach of the Council does not amend the wording of Recital (19).
[22] Council of Europe, Recommendation CM/Rec(2018)2 of the Committee of Ministers to member states on the roles and responsibilities of internet intermediaries, Adopted by the Committee of Ministers on 7 March 2018 at the 1309th meeting of the Ministers' Deputies, https://rm.coe.int/1680790e14, p. 7.

require the "proactive" monitoring or filtering of content, as it would be both inconsistent with the right to privacy and likely to amount to pre-publication censorship.[23]

The definition of 'terrorist content'

The proposal targets 'terrorist content', which is defined broadly. According to Article 2(5) of the proposal, terrorist content is information 1) '*inciting or advocating, including by glorifying, the commission of terrorist offences, thereby causing a danger that such acts be committed*'; 2) '*encouraging the contribution to terrorist offences*'; and 3) '*promoting the activities of a terrorist group, in particular by encouraging the participation in or support to a terrorist group*'. Recital (9) lists also information that provides instructions for the commission of terrorist offences. The information can be contained in text, images, sound recordings and videos.

The definition of 'terrorist content' draws on the definition of 'terrorist offences' under Directive 2017/541 on combatting terrorism.[24] The list of terrorist offences includes, for example, attacks upon a person's life which may cause death, kidnapping or hostage-taking, or seizure of aircraft, ships or other means of public or goods transport.[25] The Directive refers to '*public provocation to commit a terrorist offence*', whether online or offline, which should be '*punishable as a criminal offence when committed intentionally*'.[26] Under the Directive, therefore, the distribution of information that causes a risk that a terrorist act may be committed and which advocate for such actions should be criminalized. The proposed regulation, however, does not specify that the content must amount to a criminal offence. It is unclear if the proposal intends to expand the scope of the targeted content to also include content that is not punishable under national law. If so, the broad definition risks curtailing the expression of extreme - but legal - content.[27] Some commentators therefore wonder whether the proposed regulation is in fact intended as an amendment to Directive 2017/541, which would require Member States to criminalize actions described in the

[23] United Nations, Report of the Special Rapporteur David Kaye on the promotion and protection of the right to freedom of opinion and expression, A/HRC/38/35, 6 April 2018, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement, p. 20.

[24] Directive 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA, OJ L 88, 31.3.2017, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32017L0541.

[25] For the full list see Article 3 of the Directive 2002/475 on combating terrorism.

[26] Article 5 of the Directive 2002/475 on combating terrorism.

[27] See more on the overbroad definitions of terrorist content in: D. Mijatović, Misuse of anti-terror legislation threatens freedom of expression. The Commissioner's Human Rights Comments, 4 December 2018, https://www.coe.int/en/web/commissioner/-/misuse-of-anti-terror-legislation-threatens-freedom-of-expression.

proposed regulation.[28] In any event, the difference between the definitions used in the two instruments undermines legal certainty and foreseeability for both instruments.

In the general approach of the Council, the definition of terrorist content is more closely aligned to the definition of 'terrorist offences' used in the Directive on combatting terrorism. Terrorist content is defined as '*material which may contribute to the commission of the intentional acts, as listed in Article 3(1)(a) to (i) of the Directive 2017/541*'. In addition, '*threatening to commit a terrorist offence*' is also listed as information that amounts to terrorist content. Interestingly, the Council proposes to add that terrorist offences refer to '*intentional acts*', but does not specify that 'terrorist content' must be punishable as a criminal offence.[29]

Recital (9) of the proposed regulation lists the factors that should be taken into account when assessing whether information amounts to 'terrorist content'. For example, the material produced by, attributable to or disseminated on behalf of an EU-listed terrorist organization or person constitutes an important factor. Identifying terrorist content on the basis of official EU lists of terrorist organizations would indeed facilitate the assessment. There are, however, other factors that may play a role in the assessment, for example, '*the nature and wording of the statements, the context in which the statements were made and their potential to lead to harmful consequences',* which are all indicative factors.[30] The fact that a content producer is not listed officially as a terrorist organization does not give any guarantee that the content is not terrorist in nature. Recital (9) further stipulates that the assessment must additionally take into account that '*expression of radical, polemic or controversial views in the public debate on sensitive political questions should not be considered terrorist content*'.[31] Moreover, content disseminated for educational, journalistic or research purposes should be adequately protected. Assessing whether content amounts to 'terrorist content' under the proposed regulation will certainly not be an easy task. It is particularly concerning in situations where the assessment will have to be conducted by the hosting service providers themselves. As private entities, they may not possess the same level of expertise as the competent authorities established to tackle the problem of terrorist content.

---

[28] See J. Barata, New EU Proposal on the Prevention of Terrorist Content Online, An Important Mutation of the E-Commerce Intermediaries' Regime, October 2018, http://cyberlaw.stanford.edu/files/publication/files/2018.10.11.Comment.Terrorism.pdf.

[29] It is also worth noting that the Council's general approach no longer refers to '*information that provides instructions for the commission of terrorist offences*' in Recital (9). Instead, it specifies that the definition of terrorist content includes '*content that provides guidance for the making and use of explosives, firearms or other weapons (…) or on other methods and techniques, including the selection of targets, for the purpose of committing terrorist offences*'.

[30] The Council's general approach does not amend the wording of Recital (9) describing the factors to be taken into account in the assessment.

[31] According to the ECtHR, the right to freedom of expression protects not only information and ideas that are favorably received or deemed inoffensive, but also those that offend, shock or disturb the State or any sector of the population. See ECtHR, *Handyside v. the United Kingdom*, 7 December 1976, para. 49.

The general approach of the Council changes the wording in Recital (9) from 'journalistic purposes' to 'counter-narrative' but retains 'educational' and 'research' purposes. The Council's proposal adds, moreover, that where '*the disseminated material is published under the editorial responsibility of the content provider, any decision as to the removal of such content should take into account the journalistic standards established by press or media regulation*' consistent with the law and the right to freedom of expression and the right to freedom and pluralism of the media as protected in Article 11 of the EU Charter. The change aims to take into account the journalistic standards established by press and media regulation, but it is doubtful that this change will actually facilitate the assessment process.

Scope

The proposed regulation targets services through which terrorist content is disseminated. According to Recital (10), the regulation would apply to hosting service providers which make information stored by their users available to third parties. Such services are covered irrespective of whether their activities are of a mere technical, automatic and passive nature. For the regulation to apply, therefore, the distinction between "passive" and "active" hosts is irrelevant. As a result, the proposal covers a broad range of services. The listed examples include '*social media platforms, video streaming services, video, image and audio sharing services, file sharing and other cloud services to the extent they make the information available to third parties and websites where users can make comments or post reviews*'. The proposal intentionally does not make any exceptions for small service providers. As observed in the Explanatory Memorandum, terrorists increasingly misuse smaller providers offering different types of hosting services globally.[32] The Impact Assessment states that micro, medium and small enterprises are particularly vulnerable facing illegal content which might be uploaded by their users.[33] Moreover, the proposal covers also hosting service providers established outside the Union but offering services within the Union, if substantial connection to the Union exists (as specified in Recital (11)).

While the explanation given for including small hosting services is logical, the decision will undoubtedly raise the costs of providing online services in the EU. The broad scope of material application, moreover, requires further reflection. There is a risk that the broad coverage would encompass not only hosts at the application level, but also hosts at the infrastructure level. The latter have no direct connection with the content provider. Therefore, hosts at infrastructure level may not be able to take action against specific piece

[32] Explanatory Memorandum to the Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online (n 2), p. 1

[33] 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 {COM(2018) 640 final} - {SEC(2018) 397 final} - {SWD(2018) 409 final}, Brussels, 12.9.2018 SWD(2018) 408 final, https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-swd-408_en.pdf, p. 14.

of content but would only be able to conduct a wholesale removal or blocking (e.g. not one comment but a whole blogsite) also affecting lawful content.[34] Such a disproportionate interference with otherwise lawful content fails to respect the right to freedom of expression and access to information protected by the Charter, but also expressed in the proposal, for example in Recitals (12), (17) and (19) as well as Articles 3, and 6. Recital (7) observes, moreover, that interference with the right freedom of expression and information should be strictly targeted. In particular, it must serve to prevent the dissemination of terrorist content, but without affecting the right to lawfully receive and impart information. It is hard to imagine how the broad scope of application, possibly including hosts at the infrastructure level, can ensure that the interference remains targeted and proportionate.

The general approach of the Council alleviates this concern to some extent. Recital (10) of the general approach excludes '*other services provided in other layers of the Internet infrastructure*', for example registries and registrars, DNS, payment or DDoS services. It also adds that '*only those services for which the content provider is the direct recipient are in scope*'. The wording added by the Council's proposal may help narrowing down the scope considerably, if it indeed limits the application to services which have a direct link with the content provider and therefore does not target hosting providers at the infrastructure level.

Content removal orders

The proposed regulation establishes several ways to prevent dissemination of terrorist content online. Article 4 grants power to the competent authorities to issue removal (or disabling) orders to hosting service providers. The hosting service providers shall follow the order within one hour from receiving it. This is a particularly short time frame, which does not allow for much consideration on the validity of the order. Instead, the provision demands compliance since the content has been already identified by the competent authorities. Article 4 also describes elements that the order must include. For example, a removal order shall contain information about redress available to the hosting service provider and to the content provider.

Mistakes resulting from inaccurate assessments may happen, but in such cases, a path for a relief should be available.[35] Including information about the right to redress is an important first step towards ensuring effective remedy. However, no further information is provided about how such redress should work. Is it the appeal, as mentioned in Article 4.9, a lack of which makes the removal order final? Neither Article 4 nor the further provisions on safeguards[36] further elaborate on this important element. Recital (8) indicates that the right

[34] One the matter of wholesale blocking see also ECtHR, *Ahmet Yıldırım v. Turkey,* 18 March 2013.
[35] For more on effective remedy in content removals see A. Kuczerawy, *Intermediary Liability and Freedom of Expression in the EU: From Concepts to Safeguards*, Intersentia, 2018, forthcoming.
[36] Section III, Article 10 provides an obligation to establish complaint mechanisms by the hosting providers, but only in reference to content referrals and proactive measures.

to effective remedy '*includes the possibility for hosting service providers and content providers to effectively contest the removal orders before the court of the Member State whose authorities issued the removal order*'. There is no mention, however, of an appeal to the competent authority that issued the order. While courts are indeed best placed to decide about rights in conflict, the time needed to reach a final decision may be substantial. The effectiveness of the remedy is questionable if it takes one hour to remove content, but months or years to put it back online.[37]

Mere mention of a redress mechanism and an appeal process does not constitute an effective remedy. It can only be assumed that any redress or appeal would be examined after the actual removal of content (which, as indicated earlier, must take place within one hour). Will it possible to appeal to the competent authority or only to the competent national court? How much time would it take to review the appeal against the removal order? What would be the procedure to put the removed content back online, in case of a successful appeal? All these questions remain unanswered. The absence of further precision in this regard is striking, as recital (13) states that the procedure and obligations resulting from legal orders should be harmonised.[38]

Content referrals

The removal orders issued by competent national authorities seem to address the common concern that hosting providers are not well-placed to decide about the illegal nature of content.[39] In Article 5 of the proposal, which provides another method of preventing dissemination of terrorist content online, the concern resurfaces. According to Article 5, the competent authority or the relevant Union body (Europol) may also send a "referral" to a hosting service provider. Such a referral should lead to expeditious, but voluntary, assessment of the content identified in the referral against the hosting provider's own terms and conditions. The hosting provider must, therefore, decide whether the content is in fact 'terrorist' in nature and whether to remove it or to disable access to it. To facilitate the

[37] See an example of about Facebook deleting a user's account because he posted a picture of a 19th-century painting of a woman's genitals ("L'Origine du Monde", an 1866 oil painting by Gustave Courbet). Deletion took place in 2011 and the case was resolved in 2018, by dismissal. See more in J. Schmid, S. Bouderbala, Facebook denies 'censoring' 19th-century vagina painting, 1 February 2018, https://phys.org/news/2018-02-facebook-denies-censoring-19th-century-vagina.html; and French court throws out Facebook nude art 'censorship' case, 15 March 2018, https://www.france24.com/en/20180315-french-court-facebook-nude-art-censorship-courbet.

[38] The general approach of the Council includes additional information about the redress or appeal mechanism in a new Recital (13a) and an updated Recital (25). The additional information, however, remains quite generic and does little to make the redress or appeal mechanism more effective in practice.

[39] See more on delegate private enforcement in: S. Stalla-Bourdillon, Chilling ISPs… when private regulators act without adequate public framework, Computer Law & Security Review , Vol. 26, 2010, pp. 290-297; and L. Belli and C. Sappa, The Intermediary Conundrum: Cyber-Regulators, Cyber-Police or Both?, , 8 (2017) JIPITEC p. 183, https://www.jipitec.eu/issues/jipitec-8-3-2017/4620/JIPITEC_8_3_2017_183_Belli_Sappa; and E. Coche, Privatised Enforcement and the Right to Freedom of Expression in a World Confronted With Terrorism Propaganda Online, Internet Policy Review 7(4), 3 December 2018, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3280904.

process, the referral shall contain sufficiently detailed information, including the reasons why the content is considered terrorist and, where necessary, additional information enabling the identification of the terrorist content referred. The actual decision about the content, however, has to be taken by the hosting provider. Article 5, therefore, essentially delegates one of the tasks of the competent authority to a private company. It is unclear why a hosting provider should be better placed (than the competent authority) to make an assessment about content which was not evidently terrorist so as to warrant a removal order. If the competent authorities were not able to assess the content then why is it assumed that the hosting providers would do a better job? If the referrals are meant to be a first step in the chain of events in response to terrorist content, ranging in magnitude from referrals (with voluntary decisions) to orders (establishing actual knowledge and eliminating the immunity offered by Article 14 ECD), then the proposal should be more clear on this point.

According to Article 5, assessment by the hosting providers in case of referrals should be done in reference to their terms and conditions. Conditioning removals on terms and conditions and not the applicable law, however, increases the risk of disproportionate interference with the right to freedom of expression and access to information. In this regard, it is important to consider that there may be a variety of content that the hosting providers prohibit on their platforms but which is not illegal. For example, platforms targeted at minors prohibit certain types of legal content because it is considered inappropriate, while platforms targeted at adults would permit the same types of content. Recital (9) of the proposal actually cautions that '*expression of radical, polemic or controversial views in the public debate on sensitive political questions*' may be legal and therefore '*should not be considered terrorist content*'. The assessment, however, is left to the private companies which may err on the side of caution and remove any sensitive or controversial content to avoid potential problems. Moreover, receiving a referral from a competent authority could lead to (actual or constructive) knowledge about the illegal content within the meaning of Article 14 ECD, which creates an incentive to remove the content at issue. Such 'voluntary' removals may have a chilling effect and unduly limit public debate. It is significant that avoiding removal of content which is not terrorist, '*in particular when implementing terms and conditions*', is mentioned in relation to the hosting providers' duties of care (Recital (12)). It is not mentioned, however, in relation to the referrals. Is removal of legal content considered as acceptable collateral damage in the attempt to prevent dissemination of terrorist content through the referrals? Moreover, why does the proposal consider it the States' task to help private companies with enforcing their internal rules and possibly eliminate legal content? Would the resulting interference with the right to freedom of expression be compatible with Article 52.1 of the EU Charter, as it is not provided for by law, but rather by the private rules established by hosting service providers?

Article 10 of the proposed regulation provides that hosting service providers should establish effective and accessible complaint mechanisms as a safeguard. The complaint mechanisms would be available to content providers whose content has been removed (or access to it disabled) as a result of a referral or of a proactive measure. The complaint mechanisms would allow content providers to submit a complaint against the action of the hosting service provider and request reinstatement of the content. Furthermore, Article 10 states that hosting service providers should promptly examine every complaint they receive. They should also reinstate any content the removal of which (or disabling of access to) was unjustified, without undue delay. Providing such a form of redress is praiseworthy. Its added value, however, remains uncertain. As indicated above, assessment of content will be conducted vis-à-vis terms and conditions, which may prohibit many types of legal content. Removal on the basis of a conflict with the terms and conditions will in many cases be considered justified by the service provider. Moreover, it should be highlighted that hosting service providers have little incentive to reinstate content that they have previously removed, even if the decision was based on a mistake. On many occasions it takes significant public outcry and negative media attention to reinstate the content.[40]

Proactive measures

The proposed regulation introduces a new type of measure in the fight against dissemination of terrorist content, namely 'proactive' measures. According to the proposal, proportionate proactive measures, including automated means, are an essential element in tackling terrorist content online. Article 6 provides that '*hosting service providers shall, where appropriate, take proactive measures to protect their services against the dissemination of terrorist content*'. It seems, therefore, that the initial decision whether and which proactive measures are appropriate is left to the hosting providers. The decision should be taken depending on '*the risks and level of exposure to terrorist content*'. The absence of removal orders and referrals addressed to a hosting provider is an indication of a low level of exposure to terrorist content (Recital (16)). Is it possible, in such a case, not to introduce any proactive measures? The hosting providers must also take into account the effects on the rights of third parties and the public interest of information. Recital (17) addresses further the possible effect on the rights of third parties by specifying that '*hosting service providers should ensure that users' right to freedom of expression and information - including to freely receive and impart information - is preserved*'. Moreover, the hosting providers should act

[40] For example, this happened in case of removal of the iconic 'Vietnam napalm girl' picture or the prehistoric 'Venus of Willendorf' figurine. See more in S. Levin, J. Carrie Wong, L. Harding, Facebook backs down from 'napalm girl' censorship and reinstates photo, 9 September 2016, https://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo; and Phys.org, Facebook apologises for censoring prehistoric Venus statue, 1 March 2018, https://phys.org/news/2018-03-facebook-apologises-censoring-prehistoric-venus.html. See also, Facebook 'sorry' for censoring nude breasts from iconic French liberty goddess painting, 19 March 2018, https://www.rt.com/news/421723-facebook-censor-delacroix-nudity/.

with due diligence to '*avoid any unintended and erroneous decision leading to removal of content that is not terrorist content*', in particular, when they use automated means.

The hosting providers, who received a removal order that has become final, have less freedom in deciding about proactive measures. The competent authorities can request such hosting service providers to submit a report on the specific proactive measures they have taken. The report should include all relevant information for the competent authority to assess whether the proactive measures are effective and proportionate, including information to evaluate the functioning of any automated tools used. The report should also include information about the human oversight and verification mechanisms employed. The purpose of the measures is (1) to prevent the re-upload of terrorist content which has previously been removed or to which access has been disabled[41], and (2) to detect, identify and expeditiously remove or disable access to terrorist content. The hosting service providers is expected to check against publicly or privately-held tools containing known terrorist content when implementing its proactive measures. Providers may also employ the use of '*reliable technical tools to identify new terrorist content'*, either using those available on the market or those developed by the hosting service provider.

If the competent authority considers the taken measures insufficient, it may request, through a dialogue with the hosting service provider, that the provider takes '*specific additional proactive measures*'. If no agreement can be reached, the competent authority has a power to impose '*specific additional (…) proactive measures*'.[42] The imposed measures should be necessary and proportionate. In particular, they should be selected taking into account *'the economic capacity of the hosting service provider and the effect of such measures on the fundamental rights of the users'*. As previously mentioned, the proposal explains that imposing such '*specific proactive measures should not, in principle, lead to the imposition of a general obligation to monitor'*, as provided in Article 15 ECD. The proposal adds, however, that '*in light of the particularly grave risks associated with the dissemination of terrorist content, the decisions adopted by the competent authorities could derogate'* from the prohibition established in Article 15, '*as regards certain specific, targeted measures'*.

The approach of Article 6 is particularly worrying, for several reasons. It should be clarified, first, that the measures described in Article 6 are not specific monitoring measures. The wording of Articles 6.3 and 6.4 could suggest so initially, as they use the term '*specific proactive measures*'. For the monitoring to truly be 'specific', it should target, for example,

[41] The general approach of the Council makes a small adjustment to the text of Article 6 and uses the wording '*to effectively address the reappearance*' instead of '*preventing re-uploads*'.

[42] The general approach of the Council adds to Article 6.4 that it '*shall be to the discretion of the competent authority referred to in Article 17(1)(c) to decide on the nature and the scope of the proactive measures, in accordance with the aim of this Regulation*'.

the activity of a specific individual or group of individuals on a platform.[43] The monitoring that Articles 6.3 and 6.4 would trigger, however, must apply to all the content to be effective. Article 6 merely refers to specific technologies, rather than specific monitoring, as allowed under the E-Commerce Directive.

Second, to effectively recognise content, the technical tool must examine the entirety of content on the platform. The requirement, therefore, amounts to installing upload filters by the service providers. According to CJEU, a requirement to install a filtering system capable of identifying specific types of content, for almost all information stored by the users, applied indiscriminately to all of them, as a preventive measure, and for unlimited period of time, amounts to a general monitoring obligation.[44] Arguably, the system described by the CJEU in *Sabam v. Netlog* aimed to identify copyright infringing content. A difference in the type of the targeted content, however, does not change the way the filtering system works. No matter how specific the content being targeted is, a service provider must still monitor all uploads to catch it. The general character of monitoring refers to the set of content that must be monitored (entirety) and not the subset of the content that is being searched for (specific). It is difficult to imagine how examining all the incoming content to identify terrorist content would not constitute a general monitoring obligation as prohibited by Article 15 ECD. Despite the attempts to diminish the meaning of Article 6, Recital (19) of the proposed regulation seems to accept that the proactive measures in fact amount to general monitoring. Recital (19) justifies the derogation with reference to the particularly grave risks associated with the dissemination of terrorist content. The E-Commerce Directive, however, does not foresee any exemptions to the prohibition in Article 15. The proposed measures, therefore, contradict the EU acquis.

Third, preventing re-uploads of previously identified terrorist content involves comparing all newly uploaded content with an existing database of content known to be terrorist. The proposal, however, requires the hosting providers to do more than mere prevention of re-uploads. Specifically, the proposal requires the hosting providers to also detect new terrorist content, which has not been identified yet. Performing the task by humans can be helpful, but it is not a scalable solution considering the amount of content uploads online. Only intelligent software will be able to analyse and make autonomous decisions whether an entirely new content qualifies as terrorist. The proposal, therefore, advocates for using artificial intelligence software. Such software must take into account numerous elements, including context, in order to avoid misqualification of content which is being disseminated, for example, for educational, journalistic or research purposes. At the moment, however, content recognition software continues to make mistakes. The Impact Assessment

[43] C. Angelopoulos, Study on Online Platforms and the Commission's New Proposal for a Directive on Copyright in the Digital Single Market, 2017, https://www.repository.cam.ac.uk/handle/1810/275826, p. 37.
[44] CJEU, *SABAM v. Netlog*, C-360/10, 16 February 2012.

document accompanying the proposal states quite frankly that despite progress in the field, cases of misidentification of visual content continue to occur.[45] The Impact Assessment admits, moreover, that language processing systems are considered '*immature for accurately identifying illegal hate speech or other violent speech'.*[46] The Impact Assessment also points to stories reporting misidentification and takedown of lawful speech.[47] Erroneous removals and therefore unjustified interference with the right to freedom of expression is hence extremely probable.

Adding to the problems of performance, the artificial intelligence content recognition software would have to base its decision-making process on the extremely broad definition of terrorist content (see above). It is doubtful whether artificial intelligence software will be able to properly assess incitement to terrorist offences while maintaining radical, polemic or controversial, but legal views intact. In addition, costs of developing and implementing AI software remain high.[48]

The use and implementation of the proposed proactive measures comes equipped with safeguards, described in Articles 9 and 10 of the proposed regulation.[49] Article 9 provides that using automated tools requires effective and appropriate safeguards to ensure that decisions on removal are accurate and well-founded. Such safeguards should consists, in particular, '*of human oversight and verifications where appropriate*'. In any event, human oversight and verifications should be applied where a detailed assessment of the relevant context is required in order to determine whether or not the content is to be considered terrorist content. Human oversight, however, may not always help with erroneous removals, especially as platforms are often reluctant to admit a mistake in identifying content.[50] After all, if content violates their terms and conditions, its removal is considered justified. The complaint mechanism described in Article 10 does little to alleviate this issue, as the adjudication of complaints will still take place on the basis of the hosting providers' terms and conditions and not on the basis of applicable law.

---

[45] European Commission, Impact Assessment (n 33), p. 13, referring to a story by K. O'Flaherty, YouTube keeps deleting evidence of Syrian chemical weapon attacks, 26 June 2018, https://www.wired.co.uk/article/chemical-weapons-in-syria-youtube-algorithm-delete-video. See also A. Fox Cahn, Why is Facebook censoring Rohingya accounts of the genocide? 2 October 2017, https://www.newsweek.com/why-facebook-censoring-rohingya-accounts-genocide-675526

[46] European Commission, Impact Assessment (n 33), p. 13

[47] European Commission, Impact Assessment (n 33), p. 13, referring to a story by R. Sabur, Facebook censors America's Declaration of Independence for 'hate speech', 5 July 2018, https://www.telegraph.co.uk/news/2018/07/05/facebook-censors-americas-declaration-independence-hate-speech/.

[48] See more in the Impact Assessment (n 33), p. 14.

[49] The Council's general approach makes no changes in Articles 9 and 10.

[50] For example on a removal decision made by an algorithm and confirmed by manual review see T. Dengel, WillowTree wants to help eliminate the gender pay gap–Facebook is working against us, 29 November 2018, https://willowtreeapps.com/ideas/willowtree-wants-to-help-eliminate-the-gender-pay-gap-facebook-is-working.

Conclusion

The EU fight against online terrorist content continues. After experimenting with a number of soft-law initiatives, the Commission has moved on to propose a binding instrument, in the form of a proposed regulation. The proposal introduces new obligations to prevent dissemination of terrorist content online for both Member States and for hosting service providers delivering services in the EU. The proposal, moreover, specifies several methods of preventing dissemination of online terrorist content, in particular: duties of care, strict removal orders with one hour time-frame, content referrals to be reviewed voluntarily, and extensive proactive measures.

The definition of terrorist content is broad and possibly encompasses content that may be radical, polemic or controversial, but not illegal. The methods to prevent the dissemination of terrorist content are far reaching. Removal orders do not require assessment of content by the hosting providers as the assessment has been conducted by the competent authorities. The proposed redress mechanism, however, does not provide enough details about the procedure to be able to mitigate the potential risks for overbroad removal orders. The vague phrasing does not ensure adequate protection against undue interference with the right to freedom of expression. Moreover, the short time-frame of one hour does not leave much room for clarification. Content referrals, on the other hand, delegate the task of properly assessing content from a public body to a private company, which may not have the necessary expertise to perform the task properly. Moreover, referrals require the assessment to be conducted vis-à-vis the hosting providers' terms and conditions rather than the applicable law, which could easily result in removal of content that is radical, polemic or controversial but not illegal. The proposed complaint mechanism does not prevent removal of content which is not illegal but merely violates terms and conditions.

The proactive measures envisaged by the proposed regulation would require the relevant hosting service providers to start using artificial intelligence software, which is still costly yet often delivers faulty results. The risk of erroneous removal of legal content is again very high. Moreover, the proactive measures imposed by competent authorities could lead to general monitoring obligations, which are currently prohibited by the E-Commerce Directive as interpreted by the CJEU. The proposal maintains that the imposition of proactive measures should in principle not have the effect of becoming a general monitoring obligation. At the same time, the proposal admits that a derogation of the prohibition against general monitoring obligations may indeed be possible. The approach of the proposed regulation to proactive measures is therefore incoherent with the existing EU acquis, in particular the E-Commerce Directive.

Taking all the issues into account, one must conclude that the proposal, in its current form, creates multiple risks for removal of legal content. The proposal, therefore, poses a serious

risk to fundamental rights protected by the EU Charter, in particular the right to freedom of expression and access to information. This is unfortunate, as the proposal refers to the right to freedom of expression and access to information in numerous instances. Indeed, the proposal includes a number of safeguards that aim to prevent the detrimental effect on the right. However, at this point, the safeguards are not sufficiently developed to mitigate the risks. It must be concluded, therefore, that the proposal does not strike a fair balance between the public interest objectives and the fundamental rights involved.