7

PUBLICUM

A Democratic dénouement? The EU vs terrorist content online

Gavin Robinson
University of Luxembourg, Esch-sur-Alzette, Luxembourg. E-mail: gavin.robinson@uni.lu

Abstract
This paper explores the manifold aspects of the draft Terrorist Content Online ('TCO') Regulation which are of clear import for the democratic future of the internet insofar as they will reshape the boundaries of what is acceptable behavior online and set out tools and procedures for deciding what must be (potentially, pre-emptively) excised from the exchanges taking place there. The piece places particular emphasis on the proposal's ramifications for legal certainty and freedom of expression in cyberspace. In doing so, it aims to raise the most salient policy concerns, legal difficulties and technological quandaries which ripple out from the EU legislator's laudable goal of tackling violent and terrorist content online. The article is structured as follows: after a discussion of the foundational concept of "terrorist content" which applies across the board in the draft text (II), we distinguish its headline provisions tightening up service providers' compliance with orders to remove or disable access to terrorist content (III) from those aspects of the proposal which aim to responsibilise providers to act unassisted against terrorist content, including through the hotly-debated use of proactive measures (IV). A few concluding remarks on the future direction of this live file are offered to close (V).

Keywords
Terrorist content; Service provider; Provider responsability.

Um desfecho democrático? A União Europeia vs o conteúdo terrorista online

Resumo
Este artigo explora os múltiplos aspectos do projeto de Regulamento de Conteúdo Terrorista Online ('TCO'), que são de clara importância para o futuro democrático da Internet, na medida em que reformularão os limites do comportamento aceitável online e definirão ferramentas e procedimentos para decidir o que deve ser (potencialmente, preventivamente) excluído das trocas que ocorrem ali. O estudo coloca ênfase particular nas implicações da proposta para segurança jurídica e liberdade de expressão no ciberespaço. Ao fazê-lo, visa levantar as mais importantes preocupações políticas, dificuldades jurídicas e dilemas tecnológicos que surgem do louvável objetivo do legislador

da União Europeia de combater conteúdos violentos e terroristas online. O artigo está estruturado da seguinte forma: após uma discussão do conceito fundamental de "conteúdo terrorista" que se aplica de forma geral no texto da minuta (II), distinguimos suas disposições principais que reforçam a conformidade dos provedores de serviços com as ordens para remover ou desativar o acesso ao conteúdo terrorista (III) dos aspectos da proposta que visam responsabilizar os provedores a agir autonomamente contra conteúdo terrorista, inclusive por meio do controverso uso de medidas proativas (IV). Algumas observações finais sobre a direção futura deste arquivo ao vivo são oferecidas na conclusão (V).

Palavras-chave
Conteúdo terrorista; Provedor de serviço; Responsabilidade do provedor.

Table of Contents | Sumário

Introduction; 1. What is terrorist content? 2. Removal orders; 3. The promise/spectre of proactive measures; Conclusions; References.

Introduction

Within 24 hours of the Christchurch mosque shootings on Friday 15th March 2019, Facebook had reportedly removed 1.5m uploads of the footage. By the following Tuesday morning, the company claims that more than 800 distinct edits of the footage had been posted to its platform. A YouTube spokesman stated that "(t)he volume of related videos uploaded to YouTube in the 24 hours after the attack was unprecedented both in scale and speed – at times as fast as a new upload every second".[1] Two months on, the New Zealand prime minister and French president jointly issued 'the Christchurch Call', a voluntary initiative signed initially by ten nations from six continents, eight major tech companies and the European Commission in a first step of its kind at the global level to tackle terrorist and extremist violence online.[2]

Although it is non-binding and far less prescriptive, the Call echoed the EU Commission's draft Regulation on preventing the dissemination of terrorist content online, released in September 2018,[3] swiftly supported by the Member States at the Council in December 2018, and at the time of writing entering trialogue negotiations with the European Parliament after the latter adopted the report of its LIBE Committee in April 2019.

On the one hand, the proposal builds upon EU-level initiatives to foster the voluntary cooperation of service providers in stopping the dissemination of terrorist content online;

[1] 'Facebook and YouTube defend response to Christchurch videos', *Guardian,* 19th March 2019.

[2] 'Leaders and tech firms pledge to tackle extremist violence online', *Guardian,* 15th May 2019.

[3] COM(2018) 640 final; see also the accompanying Impact Assessment: SWD(2018) 408 final.

significant co-regulatory efforts in particular have been made by EU, national and industry actors in recent years.[4] On the other, the draft Regulation parallels ongoing national developments which go a step further in imposing obligations – underpinned by considerable fines – on service providers to hastily remove illegal content and prevent re-uploading, such as the German *Netzwerkdurchsetzungsgesetz* (NetzDG), passed in June 2017.[5] At EU level, the proposal may be seen as part of a broader assault on the tech giants, whether via a series of antitrust fines against Google or fast-growing policies on "fake news", the regulation of audiovisual media services, and copyright. Specifically, in the criminal justice policy area, it also shares features with the responsibilisation of private digital conduits which is evidenced in the Commission's earlier proposal on cross-border access to electronic evidence (usually held by IT service providers) within the EU.[6]

The stance taken by the Parliament's LIBE Committee differs in many respects from the common position of the EU Council and the Commission's proposal. Although several of these differences are raised here where appropriate, space naturally precludes an exhaustive comparative analysis of all amendments entered by the Parliament.[7] Especially since the positions of the Council and Commission are so similar, changes made by the Parliament which are deemed liable be "partly rolled back in the next phase of the policy-making process"[8] are given particular attention.

Subject to that provision, this paper explores the manifold aspects of the draft Terrorist Content Online ('TCO') Regulation which are of clear import for the democratic future of the internet insofar as they will reshape the boundaries of what is acceptable behavior online and set

[4] Most notably, 2015 saw two key developments. First, the European Commission launched the EU Internet Forum to bring together the internet industry and Member States, as well as Europol, the Radicalisation Awareness Network and the European Strategic Communications Network in order to tackle the spread of terrorist content online. Second, Europol itself established internally the so-called EU Internet Referral Unit (IRU) to actively scan the internet for terrorist content and refer it to host platforms. In September 2018, the Commission reported that "over 50,000 decisions for referrals across over 80 platforms in more than 10 languages have been made since 2015" (Impact Assessment, *ibid*, 7, 12 and 140), whilst five Member States have since set up their own IRUs (12).

[5] See also the recent UK's *Online Harms White Paper,* published on 8th April 2019 and heralding an ambitious approach to countering "illegal and unacceptable content and activity" online; Executive Summary, para 2.

[6] See Mitsilegas, V. (2018), 'The privatisation of mutual trust in Europe's area of criminal justice The case of e-evidence', *Maastricht Journal of European and Comparative Law,* 263-265; Ligeti, K. & Robinson, G., 'Transnational Enforcement of Production Orders for Electronic Evidence. Beyond Mutual Recognition?' in Kert, R. & Lehner. A. (eds.), *Vielfalt des Strafrechts im internationalen Kontext: Festschrift für Frank Höpfel zum 65. Geburtstag,* NWV, 2018, 625–644.

[7] In this vein see the collaborative efforts directed by Van Hoboken, J. (2019), 'The Proposed EU Terrorism Content Regulation: Analysis and Recommendations with Respect to Freedom of Expression Implications', Working paper of the *Transatlantic Working Group on Content Moderation Online and Freedom of Expression*, 3rd May 2019.

[8] 'Terrorist Content Regulation: Successful "damage control" by LIBE Committee', *EDRi,* 6th April 2019.

out tools and procedures for deciding what must be (potentially, pre-emptively) excised from the exchanges taking place there. The piece places particular emphasis on the proposal's ramifications for legal certainty and freedom of expression in cyberspace. In doing so, it aims to raise the most salient policy concerns, legal difficulties and technological quandaries which ripple out from the EU legislator's laudable goal of tackling violent and terrorist content online.

The article is structured as follows: after a discussion of the foundational concept of "terrorist content" which applies across the board in the draft text (II), we distinguish its headline provisions tightening up service providers' compliance with orders to remove or disable access to terrorist content (III) from those aspects of the proposal which aim to impute responsibilities on providers to act unassisted against terrorist content, including through the hotly-debated use of proactive measures (IV). A few concluding remarks on the future direction of this live file are offered to close (V).

1. What is terrorist content?

In standing EU law, Member States are already obliged to "take the necessary measures to ensure the prompt removal of online content constituting a public provocation to commit a terrorist offence" (Art. 21, Directive (EU) 2017/541 on combating terrorism). Article 5 of the same Directive defines the criminal offence of public provocation to commit a terrorist offence non-exhaustively as follows: "the distribution, or otherwise making available by any means, whether online or offline, of a message to the public with the intent to incite the commission of (a terrorist offence), where such conduct, directly or indirectly, such as by the glorification of terrorist acts, advocates the commission of terrorist offences, thereby causing a danger that one or more such offences may be committed, is punishable as a criminal offence when committed intentionally".[9]

In contrast, article 2(5) of the draft TCO Regulation as proposed by the Commission defined terrorist content as "one or more of the following information: (a) inciting or advocating, including by glorifying, the commission of terrorist offences,[10] thereby causing a danger that such acts be

[9] Notably, a November 2018 reform inserted congruous provisions into the Audiovisual Media Services Directive ('AVMSD') from 2010 committing Member States to "ensure that video-sharing platform providers under their jurisdiction take appropriate measures to protect (…) (c) the general public from programmes, user-generated videos and audiovisual commercial communications containing content the dissemination of which constitutes an activity which is a criminal offence under Union law, namely public provocation to commit a terrorist offence as set out in Article 5 of Directive (EU) 2017/541 (…)"; new Art. 28b(1), amended AVMSD; consolidated version available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02010L0013-20181218&from=EN> accessed 3rd April 2019.

[10] Themselves defined in Article 2(4) of the proposal by reference to the extensive list of offences in Article 3(1) of Directive (EU) 2017/541.

committed; (b) encouraging the contribution to terrorist offences; (c) promoting the activities of a terrorist group, in particular by encouraging the participation in or support to a terrorist group within the meaning of Article 2(3) of Directive (EU) 2017/541; (d) instructing on methods or techniques for the purpose of committing terrorist offences."

The many differences between the two composite provisions are striking; in particular, the Regulation's (b), (c) and (d) would be novelties in EU law, whilst intent is missing altogether from the new proposal's definition of terrorist content. This is all the more notable in light of the fact that the deadline for implementation of the Directive on combatting terrorism passed on 8 September 2018 – less than a fortnight before the draft TCO Regulation was unveiled – with the Commission stating that 15 Member States had notified it of measures giving effect to the standing Art. 21 obligation in their domestic systems.[11]

Arguably, an evaluation of national measures taken pursuant to Art. 21 would be essential to gauging the added value of the Directive, as well as "its impact on fundamental rights and freedoms, including on non-discrimination" (as mandated by Article 29 of that Directive). The distant deadline for the Commission's added value report, however, is 8 September 2021. In opting to respond to Member State pressure with a proposal for a Regulation[12] before any such evaluation of the Terrorism Directive has occurred, the Commission may also have been swayed by its experience of chasing up the slow and patchy implementation of Art. 25 of Directive 2011/92/EU, which provided that Member States *shall* take the necessary measures to ensure the prompt removal of, and *may* take measures to block access to, web pages containing or disseminating child pornography.[13]

To return to the three new categories of terrorist content in the draft Regulation, the Commission paints their inclusion as a step to provide clarity to hosting service providers (HSPs)[14]

[11] Impact Assessment, op cit., 9 and 116.

[12] An amendment of the Terrorism Directive, for instance to broaden the activities/material covered therein from incitement to terrorism only to other offences (as proposed in the draft Regulation), was discarded by the Commission. It argued that "the focus on criminalisation of terrorist offences as opposed to purely preventative measures, the geographical limitations and limitations in terms of safeguards and other flanking measures (which would not have been possible under this legal basis) (…) would have had limited impact on the objective of preventing the dissemination of terrorist content" (Impact Assessment, op cit., 25).

[13] See COM(2016) 872 final, where the Commission (at 12) stated that continued work was still required to ensure the "complete and correct implementation" of Art. 25 across the Member States – a full three years after the deadline of 18 December 2013, with the Commission having opened infringement proceedings against a total of 15 Member States.

[14] The Regulation would apply to "hosting service providers" (HSPs), defined as "a provider of information society services consisting in the storage of information provided by and at the request of the content provider and in making the information stored available to third parties" (Art. 2(1)). Recital 10 of the draft gives as examples: "social media platforms, video streaming services, video, image and audio sharing services, file sharing and other cloud services to the extent that they make the information available to third parties and websites where users can make comments or post reviews". Another prerequisite is that HSPs offer services in the Union, irrespective of their place of main establishment (Art. 1(2)).

and competent authorities and as a basis for more effective preventative action,[15] given the variable nature of content used for radicalisation purposes. In this context, real-life cases are referred to:[16] a Danish schoolgirl found guilty in 2017 of attempted terrorism having tried to make bombs to be used in attacks against her former school (radicalised via internet and chat contacts within a few months); Daesh's tactics to "groom" young children (using cartoons); the attack on the Thalys train in 2015 (provoked in the preceding moments by a "call to arms" on a YouTube audio file). Yet it is worth repeating that the examples of terrorist content cited all ostensibly feature an element that was not taken up by the definition in the draft TCO Regulation: intent. This has raised concerns that any communication of terrorist-related content may risk automatic deletion, irrespective of the context of its use (i.e. for confrontation, reporting, research or historical purposes).[17] Without offering an explicit justification, however, the draft not only "draws on" (as stated in recital 9 of the proposal) but also goes beyond the wording of the takedown obligation in the Directive on combating terrorism.[18]

Curiously, the definition adopted in the Council's general approach (which largely tracked the Commission's proposal[19]) inserted a tautological reference[20] to the intentionality of the acts criminalized by the Terrorism Directive, whilst any reference to intent in relation to the sharing of such content was absent. It was unsurprising to see the Parliament take issue with these wordings in its amendments by both inserting a requirement of intent across the board and systematically tying each part of the definition back to the more developed provisions already in force in the Terrorism Directive.[21]

Would the inclusion of intent – a quintessential feature of the criminal law – in the definition of terrorist content anchoring a piece of internal market legislation constitute a victory for freedom of expression? Are HSPs in any sense equipped to assess the intentions of content providers, thereby adding a further level of complexity to their task of identifying content *per se* as terrorist, meaning *inter alia* "the nature and wording of the statements, the context in which the statements were made and their potential to lead to harmful consequences, thereby affecting the security and

[15] See recital 9 of the draft Regulation.
[16] Cf. Impact Assessment, op cit., 17.
[17] EDRi, "EU's flawed arguments on terrorist content give big tech more power", 24 October 2018, < https://edri.org/eus-flawed-arguments-on-terrorist-content-give-big-tech-more-power/> accessed 3 April 2019.
[18] Art. 5, Terrorism Directive, op cit.
[19] Although it added a new article 2(5)(aa): "threatening to commit a terrorist offence".
[20] A point also made by Van Hoboken et al, *supra,* 6.
[21] See also arguing that "the difference between the definitions used in the two instruments undermines legal certainty and foreseeability for both instruments", Kuczerawy, A. (2018), 'The proposed Regulation on preventing the dissemination of terrorist content online: safeguards and risks for freedom of expression', available at <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3296864> accessed 3 April 2019, 7.

safety of persons"?[22] Faced with this heavy responsibility, HSPs as well as competent authorities are suitably reminded in the Commission proposal of the importance of freedom of expression and information, "one of the values on which the Union is founded", but only – rather weakly – in a recital.[23]

The answer to such questions will play out differently in the contexts of state-issued removal orders (addressed in III.) or largely autonomous content management by HSPs (addressed in IV.).[24]

2. Removal orders

The proposal's headline-grabbing one-hour window for takedown can be traced back to a March 2018 Commission Recommendation,[25] and the driving force behind it is clear: platforms' responses times are seen as too slow and too unreliable.

Under the proposal, national competent authorities[26] may issue removal orders to HSPs who must comply within one hour of reception (Art. 4(2)). The recipient may decide whether to remove the relevant content or to block access thereto – but must do one of the two.[27] Removal orders sent to HSPs shall contain, *inter alia*: a statement of reasons explaining why the content is considered to be terrorist content by reference to the definitions used in the draft Regulation;

[22] Recital 9 taking up relevant ECtHR case law. Moreover, recital 9 mentions that "(t)he fact that material was produced by, is attributed to or disseminated on behalf of an EU-listed terrorist organisation or person constitutes an important factor in this assessment".

[23] Recital 7.

[24] A "Referrals" system included in Article 5 of the Commission proposal, providing for the codification (complete with attendant fines) of "voluntary" consideration by HSPs of content referred to them by national competent authorities or Europol has been effectively neutered in the Parliament's position due to the high risks to freedom of expression posed by the opaqueness of the procedure and the lack of due process (only a private complaints mechanism had been foreseen for content providers). Since it is deemed unlikely to survive in future negotiations, the referrals proposal will not be addressed in this paper. For a discussion of the risks of "private censorship" of online content deemed illicit by reference to companies' own terms and conditions, see Coche, E. (2018), 'Privatised enforcement and the right to freedom of expression in a world confronted with terrorism propaganda online', *Internet Policy Review* 7(4), esp. 5-8.

[25] The Commission's non-binding March 2018 Recommendation on measures to effectively tackle illegal content online, which encouraged Member States to develop their response to all types of illegal content inter alia through systems of notices to HSPs, informing content providers and counter-notices, transparency and safeguards, and the cooperation of HSPs with national competent authorities, with "trusted flaggers", and amongst themselves. Indeed, most of the core provisions of the draft Regulation have been fleshed out from the Commission's earlier specific recommendations relating to terrorist content, which featured *inter alia* a ban on terrorist content in HSPs' terms of service, referrals, proactive measures, cooperation and the one-hour rule: "Recommendation 35. Hosting service providers should assess and, where appropriate, remove or disable access to content identified in referrals, as a general rule, within one hour from the moment at which they received the referral" (Recommendation (EU) 2018/334, 14).

[26] Controversially, Member States are free to assign this task to either administrative, law enforcement or judicial authorities; discussed *infra.* See recital 13: "Member States should remain free as to the choice of the competent authorities allowing them to designate administrative, law enforcement or judicial authorities with that task". Insisting on the need for judicial oversight in all cases, see Van Hoboken et al, *supra* at 6.

[27] Recital 13.

information enabling the identification of the content referred (typically a URL); and information about redress available to both the HSP and to the content provider.

From the content providers' perspective, the draft Regulation provides that HSPs are in turn to supply them with "information" regarding the removal or blocking of content (Art. 11(1)) and (upon request of the content provider) reasons for such action (Art. 11(2)). However, these obligations may be suspended for up to four weeks where the competent authority decides for reasons of public security, such as the prevention, investigation, detection and prosecution of terrorist offences, that no information on removal or blocking should be disclosed (Art. 11).

The draft Regulation covers the main practical matters on the procedure for removal/blocking orders, e.g. nomination of HSP points of contact or legal representatives, framework for interaction between HSPs and competent authorities, provisions on language etc., but controversially[28] leaves open two key intertwined aspects: the choice of competent authority and the establishment of appeal mechanisms.

The Commission itself notes that in the majority of national systems a takedown order may come only from a judicial body within criminal proceedings. Some Member States, however, provide for administrative orders – subject to appeal before a court – and in a few Member States even law enforcement authorities can issue removal orders and refer content to service providers.[29] Fearing overreach from assorted authorities, the European Parliament promptly inserted a new provision into Article 2 "Definitions" in order to limit "competent authority" to mean "a single designated authority in the Member State, or an independent administrative authority, with the relevant expertise".[30] Meanwhile, another Parliament addition (Article 8a) provides that "Member States shall ensure that a content provider or a hosting service provider can appeal a removal order as referred to in Article 4(9) by seeking redress in front of the relevant judicial authority in the Member State in which the content provider is located or in which the main establishment of the hosting service provider or legal representative designated by the hosting service provider pursuant to Article 16 resides or is established", and elsewhere shores up provisions in the aim of making such appeals effective.

The tying of appeals to a content provider's or HSP's host jurisdiction reflects one of the Parliament's greatest concerns over the file: the ostensibly cross-border effect of removal orders issued by any given national competent authority. Perhaps alarmingly, the Commission proposal had made scant reference to the territorial scope of a removal order issued by a competent

[28] In this regard, see the concerns raised by a coalition of 31 civil society organisations in a letter sent to EU Member States' Home Affairs Ministers on 4 December 2018, available at https://edri.org/files/counterterrorism/20181204- CivilSociety_letter_TERREG.pdf (accessed 3 April 2019).
[29] Impact Assessment, op cit., 9-10.
[30] Amendment 31.

authority, before the Council effectively declared open season in version of Article 15(1), which now ended by stating that "(a)ny Member State shall have jurisdiction for the purposes of Articles 4 and 5, irrespective of where the hosting service provider has its main establishment or has designated a legal representative". The Parliament has responded by inserting a full article on cross-border cooperation, witha carve-out for the host jurisdiction (duly informed of a live removal order) to trigger fundamental rights protection for HSP, representative or content provider.[31]

In contrast to these deep differences over scoping and procedural issues, however, the one-hour rule – which is far tighter than existing provisions in national law mandating removal within 24 or 48 hours,[32] and justified by the Commission by reference to the speed at which terrorist content is claimed to spread across online services[33] - has remained largely untouched. And as the rule remains, so too does the central question: even if narrowed down to national effects, packed in judicial oversight and reinforced by appeals – is there a real need for such a rigid rule?

The question will have to be answered from an EU law perspective as well as a policy one. The immediate policy imperative to act at national level is of course constituted by the increasing use of HSPs to disseminate terrorist content. In terms of EU law, meanwhile, the main driver for Union-level action (on an Art. 114 TFEU legal basis) is the hindrance to the effective exercise of freedom of establishment and freedom to provide services across the Union which may result from the legal uncertainty caused by a deepening "fragmentation" of national responses to the removal or disabling of access to illegal content in general (as already envisaged in the e-commerce Directive)[34] and terrorist content in particular (as more recently mandated by the Terrorism Directive). Here too, there is fertile ground for debate over the coming months and beyond as to the true necessity of EU action – not only from a freedom of expression perspective but also on economic grounds.

On the HSP side, for example, the Commission cites input from "companies" to theeffect that diverging legislation is a serious concern.[35] Although it seems undeniable that the functioning

[31] Amendment 53.
[32] Impact Assessment, op cit., 116.
[33] See also the "evidence summary" on terrorist content online (Annex 9, Impact Assessment, op cit., 138), referring to one example of UK Home Office analysis showing that a third of Daesh links are disseminated within the first hour. The preference for such a short window for removal or blocking goes hand-in-hand with the Commission's choice to eschew an "all illegal content" approach – deemed "unnecessary and disproportionate" – and focus instead on terrorist content as a priority, at least for now; Impact Assessment, op cit., 23. Recently, the sluggish response of flaggers once more entered the public gaze in the wake of Christchurch. In the aftermath, a Facebook vice-president stated that the first user report came in 29 minutes after the broadcast started and 12 minutes after it ended; 'Facebook and YouTube defend response….', *supra.*
[34] Art. 14(3) of Directive 2000/31/EC, : "This article shall not affect the possibility for a court or administrative authority, in accordance with Member States' legal systems, of requiring the service provider to terminate or prevent an infringement, nor does it affect the possibility for Member States of establishing procedures governing the removal or disabling of access to information".
[35] Impact Assessment, op cit., 10.

of the "country of origin" regulatory principle would likely be undermined by the continued proliferation of national removal order systems with differing assessments of terrorist content (e.g. compliance with regulatory requirements in one Member State ensuring access to all others, only for one or several of those other Member States to remove or block access to content), the added value of the draft Regulation in this specific respect would seem to depend on levels of cohesiveness between competent authorities which are high enough to produce similarity between the decisions they take.[36] This triggers workability concerns, particularly given the likely varying nature of competent authorities and the open-ended definition of terrorist content they are to wield. Ultimately, the institutions have the duty to carefully weigh the benefits of regulating against the risk of inadvertently bolstering the abilities of untrusted national administrations' abilities to use divergent, open definitions of terrorist offences *per se* in national criminal law to actively censor legitimate debate.

3. The promise/spectre of proactives measures

Perhaps the most novel section of the draft TCO Regulation entails a duty on HSPs to take proactive measures to protect their services against the dissemination of terrorist content (Art. 6). HSPs are to report to the competent authority[37] on "the specific proactive measures […] taken, including by using automated tools, with a view to (i) preventing the re-upload of content which has previously been removed or to which access has been disabled because it is considered to be terrorist content; (ii) detecting, identifying and expeditiously removing or disabling access to terrorist content".[38]

Should the competent authority deem such measures to be insufficient, the HSP is bound to cooperate with it in order to establish "key objectives and benchmarks as well as timelines for their implementation" (Art. 6(3)), and, where no agreement can be reached, the taking of specific proactive measures can be imposed on HSPs by the competent authority (Art. 6(4)). Penalties for HSPs which fail to report on proactive measures, or fail to adopt such measures following a decision imposing them, are to be set out at national level (Art. 18(1)(d)).

The Parliament has of course reacted to this remarkable change of gear from the Commission and Council, as noted *infra.* Structuring the discussion that ensues, two main interrelated concerns can be highlighted in relation to the drive toward proactive measures: their fraught relationship

[36] See e.g. Art. 13(1) of the draft Regulation: "Competent authorities in Member States shall inform, coordinate and cooperate with each other and, where appropriate, with relevant Union bodies such as Europol with regard to removal orders and referrals to avoid duplication, enhance coordination and avoid interference with investigations in different Member States".
[37] The "competent authority" may be different to that in charge of referrals and removal orders, see Art. 17.
[38] Cf. Art. 6(2) of the Commission's proposal COM(2018) 640.

with current EU law on the one hand (i) and the question of whether they can in fact perform as expected – or might create more problems than they solve – on the other (ii).

i) Compliance (and coherence) with EU law

To take the EU law perspective first, in developing the proposal the Commission considered several options with regard to the use of automated tools: deferring to companies' own risk assessment; mandatory uptake of measures limited to the prevention of re-uploading of known terrorist content; and mandatory uptake of "appropriate proactive measures, including by using automated detection tools".[39] Despite widespread misgivings as to the current accuracy of such tools, even amongst EU Member State governments within the Commission's own Impact Assessment,[40] the third, most far-reaching option was chosen.

The Commission has done so whilst acknowledging that should proactive measures inadequately distinguish between unlawful and lawful conduct, this may risk undermining, *inter alia*, freedom of information in contravention of settled EU law.[41] The apposite rule is found in the much-discussed Article 15 of the e-commerce Directive,[42] providing *inter alia* that Member States "shall not impose a general obligation on providers (…), to monitor the information which they transmit or store, nor a general obligation actively to seek facts or circumstances indicating illegal activity". The tension between standing CJEU case law interpreting this provision as barring the imposition of systematic ISP (internet service provider) filtering for copyright breaches[43] and the draft TCO Regulation is reflected in the wording of recital 16 to the draft Regulation, which states – disconcertingly – that whilst on the one hand it is up to HSPs to determine what proactive measure should be put in place, on the other hand "(t)his requirement *should not* imply a general monitoring obligation".[44]

[39] Impact Assessment, op. cit., 105.

[40] Cf. Impact Assessment, op. cit., 120: "Regarding filtering, the (Dutch) government has indicated that this does not work adequately, since in case of terrorism unlawful content is not as evident as compared to e.g. child pornography, resulting in a disproportionate interference with the right to freedom of speech".

[41] Impact Assessment, op .cit., 105.

[42] Directive 2000/31/EC.

[43] CJEU, 24 November 2011, Case C-70/10, *Scarlet Extended SA v SABAM*, esp. para 40, paras 50 et seq; CJEU, 16 February 2012, Case C-360/10, *SABAM v Netlog,* esp. para 38.

[44] Emphasis added. Noting in the copyright context that a prohibition on the imposition of general monitoring obligations does not equal a prohibition on the imposition of *specific* monitoring obligations, and concluding that "(i)n the end, it is a question of the technical solution used by the provider", see Nordemann, J. B., *Liability of Online Service Providers for Copyrighted Content – Regulatory Action Needed?,* Study prepared on request of the European Parliament's Committee on the Internal Market and Consumer Protection, January 2018, 2.5. Arguing that "(i)t 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", see Kuczerawy (2018), op cit, 14. For an earlier comparative view of relevant case law on proactive monitoring orders across diverse

The Parliament's LIBE Committee responded sharply by deleting the core of these provisions (prevention of re-uploading of known terrorist content; detection, identification and removal of terrorist content), substituting "specific" measures for "proactive" measures, and further shifting the emphasis back to user flagging whilst expressly tacking to the stance of the CJEU in the *Scarlet Extended* case law that any measures "shall not include automated content filters or other measures that entail the systematic monitoring of user behaviour".[45] Thus the stage would appear to be set for an(other) interinstitutional face-off.

Although space precludes a detailed unpacking here of the numerous connected EU-level instruments regulating service providers' responses to illicit online content, it is worth querying whether the solidity of the Parliament's position on proactive measures in the TCO Regulation (along with the associated CJEU jurisprudence) may risk being undermined by its activities elsewhere. In particular, its adoption in March 2019 of a new Copyright Directive producing changes in the liability regime for online content-sharing service providers has been criticised as opening the door for "upload filters" (as widely termed in the policy debate, if not in the legal provisions) to be included in national implementations.[46] Whether the EU legislator's stance on copyright proves to be precedent or exception remains to be seen[47] – but surely few would doubt the extra political clout of the terrorism debate: if this is acceptable for copyrighted material, how can dangerous terrorist content be spared? In the meantime, national implementations of the Copyright Directive are already being sized up by digital rights groups for challenges before the Luxembourg court.[48]

Although less immediately apparent – and, conversely, largely untouched by the Parliament in its amendments – a further legal tangle may await the proposal in general and its drive toward proactive measures in particular given the vast amounts of data in question. Art. 7 of the draft Regulation obliges HSPs to preserve terrorist content which has been removed or disabled as a result of a removal order, a referral or proactive measures. With a preservation period set at six

sectors and jurisdictions, see also Frosio, G. F. (2017), 'The Death of 'No Monitoring Obligations': A Story of Untameable Monsters', *JIPITEC* 8: 199-215.

[45] Amendment 74.

[46] Article 17 of Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ L 130, 17.5.2019, p. 92-125. See e.g. 'Press Release: Censorship machine takes over EU's internet', *EDRi,* 26th March 2019.

[47] Compare for instance the language in Article 17(8) of the new Copyright Directive ("The application of this Article shall not lead to any general monitoring obligation") to that attached to the co-regulatory regime in the November 2018 AVMSD reform (*supra* n.9) ("Those measures shall not lead to any ex-ante control measures or upload-filtering of content which do not comply with Article 15 of Directive 2000/31/EC"). Surveying an "emerging tendency of the European lawmaker to encourage the use of proactive filtering mechanisms", see Riis, T. & Schwemer, S. F. (2019), 'Leaving the European Safe Harbor, Sailing Toward Algorithmic Content Regulation', *Journal of Internet Law* 22(7): 11-21.

[48] 'Filters Incorporated', *EDRi,* 9th April 2019.

months (extendable), the potential ramifications of this framework are bound to attract comparisons to the annulled Data Retention Directive – especially insofar as the obligation not only covers the targeted terrorist content per se but also extends to "related data removed as a consequence of the removal of the terrorist content" (Art. 7(2)).

The term "related data" is not defined in the main text of the draft Regulation, but subscriber data and access data are given as examples in recital 20. This is unlikely to quell fears of over-preservation by HSPs and will likely fuel calls for a more precise wording. Moreover, whatever data is preserved by HSPs as being in their estimation "likely to have a link with terrorist offences" (recital 21) may be used for the prevention, detection, investigation and prosecution of terrorist offences (Art. 7(1)(b)). Depending on definitions in place in national systems, these purposes may open broad channels of access to large amounts of data generated by the customers of HSPs. In turn, this approach triggers concerns regarding profiling, particularly in light of the CJEU's language on the use of non-content data to draw "very precise conclusions concerning the private lives of the persons whose data has been retained"[49] and lead to "the feeling that (users') private lives are the subject of constant surveillance".[50]

ii) Potential impacts of proactive measures

The mini-compliance framework for the automated detection and removal of terrorist content set out in the Commission and Council texts channels a fairly recent but major growth in political pressure[51] as well as industry activity, chiefly in the form of a pooled "hash database".[52] The extent to which such initiatives and the technologies they employ are truly effective at accurately

[49] CJEU, 8 April 2014, Joined Cases C-293/12 and C-594/12, *Digital Rights Ireland and Seitlinger*, para 27.
[50] *Digital Rights Ireland*, ibid, para 37 and CJEU, 21 December 2016, Joined Cases C-203/15 and C-698/15, *Tele2 Sverige AB and Secretary of State for Home Department v Tom Watson and Others*, para 100. Cf. CJEU, 2 October 2018, Case C-207/16, *Ministerio Fiscal*, wherein a targeted request for access to data concerning a stolen mobile telephone was deemed to "not therefore allow precise conclusions to be drawn concerning the private lives of the persons whose data is concerned"; para 60.
[51] See the European Council conclusions on security and defence of 22 June 2017, point 2: "Building on the work of the EU Internet Forum, the European Council expects industry to establish an Industry Forum and to develop new technology and tools to improve the automatic detection and removal of content that incites to terrorist acts. This should be complemented by the relevant legislative measures at EU level, if necessary". See also, subsequently to the Commission's proposal for a Regulation, European Parliament, Report on findings and recommendations of the Special Committee on Terrorism (2018/2044(INI)), 21 November 2018, "47. Underlines the need to achieve automatic detection and systematic, fast, permanent and full removal of terrorist content online on the basis of clear legal provisions including safeguards, and human review; […] welcomes the Commission' legislative proposal […]; calls on the co-legislators to urgently work on the proposal […]".
[52] Under the aegis of the Global Internet Forum to Counter Terrorism, founded in June 2017 by the Big Four of Facebook, Microsoft, Twitter and YouTube. See <https://gifct.org/about/> accessed 3 April 2019.

identifying terrorist content (as distinct from removing vast swathes of content)[53] raises a set of thorny issues which are likely to play out from late 2019 onward, particularly given the European Parliament's stated position on the importance of transparency in the use of algorithms. In this context, the Parliament has previously called on the Commission and the Member States to "examine the potential for error and bias in the use of algorithms in order to prevent any kind of discrimination, unfair practice or breach of privacy".[54]

Counterbalancing its provisions designed to foster the uptake of proactive measures, the draft Regulation had stipulated standard GDPR-era obligations on HSPs to provide a meaningful explanation of such tools in their terms and conditions, to publish transparency reports (including "information" and absolute figures on action taken and complaint procedures) and to provide safeguards to ensure that decisions taken concerning stored content are "accurate and well-founded", consisting in particular of "human oversight and verifications where appropriate and, in any event 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" (Art. 9).

Since then, the Parliament has (as noted above) moved to simultaneously emasculate the proactive approach preferred by the Commission and the Council and to boost the transparency provisions – such as they were in relation to proactive measures – which would only apply to the substitute scope of "specific measures". Whilst this strong statement from the Parliament must be acknowledged, it would be complacent to discount some proactive elements either creeping back into the text at the trilogue stage or growing in practice – not least given the established goal of executing takedowns within the hour. Whether at an official level or in the field, therefore, we can expect to see the putative safeguard of human oversight and/or review backing up otherwise-automated content removal. In this respect, it is important to heed the warning of David Kaye, UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and

[53] Cf. the comment of D. Keller, "Inception Impact Assessment: Measures to Further Improve the Effectiveness of the Fight Against Illegal Content Online", 29 March 2018, available at SSRN: <http://dx.doi.org/10.2139/ssrn.3262950>, accessed 3 April 2019: "Technical filters cannot assess context or tell whether potentially terrorist content is actually illegal. No existing machine – be it a simple filter or the most advanced artificial intelligence – can review new material, or look at old material in a new context, and say with certainty whether it violates the law". Also see more recently with regard to the Christchurch shootings: "But the technology used to create the fingerprints is fragile and can be defeated by simple methods such as filming the screen and uploading the resulting video. One such recording, tracked by the Guardian over the course of Friday, was on Facebook for more than five hours before being taken down , despite being headlined "New Zealand mass shooting"" ; 'Facebook and YouTube defend…', *supra.*

[54] European Parliament, Report on online platforms and the digital single market (2016/2276(INI)), 31.5.2017, at 17 (Legal Affairs Committee) and 12 (full report).

expression, who a few weeks before publication of the draft Regulation in reference to the application of artificial intelligence (AI) tools to online content stated that:[55]

"Even when algorithmic content moderation is complemented by human review – an arrangement that large social media platforms argue is increasingly infeasible on the scale at which they operate – a tendency to defer to machine-made decisions (…) impedes interrogation of content moderation outcomes, especially when the system's technical design occludes that kind of transparency".

Reports emerging post-Christchurch would seem to support these views: as reported in the UK press, "YouTube said it had tried to keep on top of the unprecedented number of videos uploaded, eventually going so far as to eject human reviewers from the loop in order to let automated systems take down more videos instantly".[56]

Conclusions

As the draft TCO Regulation enters the next stage of the EU legislative process, the very definition of "terrorist content" included therein is bound to excite much debate, and with good reason for it will underpin and shape every other aspect of the instrument. It is the definition that national competent authorities will be tasked with wielding in order to issue removal orders to HSPs, and it will provide the first baseline for the latter, along with content providers, to appeal against such orders. Consequently, it will inevitably operate as some form of filter for all manner of exchanges taking place online – but to what effect on open democratic values on the internet?

In the immediate, much will depend on the outcome of the multiple tensions between the EU co-legislators which have been set out in the foregoing. Key scoping issues from a freedom of expression perspective – particularly in light of the illiberal tendencies exhibited in recent times by some Member State governments – include the identity and independence of national authorities which may issue removal orders, and the geographical reach of the latter.

Although it has not been possible in this paper to delve into the nascent experiences from national jurisdictions which are rolling out comparable mechanisms,[57] or concerns over the relative weight which the removal or blocking of terrorist content ought to exert in antiterrorism or

[55] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 29 August 2018, A/73/348, para. 15.
[56] 'Facebook and YouTube defend response….', op cit.
[57] For a UK perspective, see McIntyre, T.J., 'Internet Censorship in the United Kingdom: National Schemes and European Norms' in Edwards, L. (Ed.), *Law, Policy and the Internet*, Hart, 2019, 291-330.

criminal justice policies,[58] the headline one-hour removal window in the draft TCO Regulation is bound to trigger scrutiny of whether it is fair on smaller operators, and on precisely what terms it may be deemed necessary – likely leading to criticism centred on the principle of evidence-led policy-making.

Yet in the medium to long term, if and when a compromise text passes, insofar as that text incorporates or encourages the taking of "proactive measures" on the part of private entities to pre-emptively remove or block terrorist content online, the legal territory inhabited by the TCO Regulation will shift, not least through legal challenges before the CJEU. Meanwhile, it is our democratic spaces online which will feel the effects – reason enough to tread carefully.

References

'FACEBOOK AND YOUTUBE defend response to Christchurch videos', *Guardian,* 19th March 2019.

'FILTERS INCORPORATED', *EDRi,* 9th April 2019.

'LEADERS AND TECH firms pledge to tackle extremist violence online', *Guardian,* 15th May 2019.

'PRESS RELEASE: Censorship machine takes over EU's internet', *EDRi,* 26th March 2019.

'TERRORIST CONTENT REGULATION: Successful "damage control" by LIBE Committee', *EDRi,* 6th April 2019.

COCHE, E. (2018), 'Privatised enforcement and the right to freedom of expression in a world confronted with terrorism propaganda online', *Internet Policy Review* 7(4).

D. KELLER, "Inception Impact Assessment: Measures to Further Improve the Effectiveness of the Fight Against Illegal Content Online", 29 March 2018, available at SSRN: <http://dx.doi.org/10.2139/ssrn.3262950>, accessed 3 April 2019.

EDRI, "EU's flawed arguments on terrorist content give big tech more power", 24 October 2018, < https://edri.org/eus-flawed-arguments-on-terrorist-content-give-big-tech-more-power/> accessed 3 April 2019.

EUROPEAN PARLIAMENT, Report on online platforms and the digital single market (2016/2276(INI)), 31.5.2017, at 17 (Legal Affairs Committee) and 12 (full report).

FROSIO, G. F. (2017), 'The Death of 'No Monitoring Obligations': A Story of Untameable Monsters', *JIPITEC* 8: 199-215.

[58] See Walker, C. (2017), 'The War of Words with Terrorism: An Assessment of Three Approaches to Pursue and Prevent', *Journal of Conflict and Security Law* 22(3): 523-551.

KUCZERAWY, A. (2018), 'The proposed Regulation on preventing the dissemination of terrorist content online: safeguards and risks for freedom of expression', available at <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3296864> accessed 3 April 2019, 7.

LIGETI, K. & Robinson, G., 'Transnational Enforcement of Production Orders for Electronic Evidence. Beyond Mutual Recognition?' in Kert, R. & Lehner. A. (eds.), *Vielfalt des Strafrechts im internationalen Kontext: Festschrift für Frank Höpfel zum 65. Geburtstag,* NWV, 2018, 625–644.

MCINTYRE, T.J., 'Internet Censorship in the United Kingdom: National Schemes and European Norms' in Edwards, L. (Ed.), *Law, Policy and the Internet*, Hart, 2019, 291-330.

MITSILEGAS, V. (2018), 'The privatisation of mutual trust in Europe's area of criminal justice The case of e-evidence', *Maastricht Journal of European and Comparative Law,* 263-265.

NORDEMANN, J. B., *Liability of Online Service Providers for Copyrighted Content – Regulatory Action Needed?,* Study prepared on request of the European Parliament's Committee on the Internal Market and Consumer Protection, January 2018, 2.5.

REPORT OF THE Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 29 August 2018, A/73/348.

RIIS, T. & Schwemer, S. F. (2019), 'Leaving the European Safe Harbor, Sailing Toward Algorithmic Content Regulation', *Journal of Internet Law* 22(7): 11-21.

VAN HOBOKEN, J. (2019), 'The Proposed EU Terrorism Content Regulation: Analysis and Recommendations with Respect to Freedom of Expression Implications', Working paper of the *Transatlantic Working Group on Content Moderation Online and Freedom of Expression*, 3rd May 2019.

WALKER, C. (2017), 'The War of Words with Terrorism: An Assessment of Three Approaches to Pursue and Prevent', Journal of Conflict and Security Law 22(3): 523-551.

Gavin Robinson
Gavin Robinson is a postdoctoral researcher in criminal law and IT law at the University of Luxembourg. His recent published work deals with electronic evidence, negotiated justice, cybercrime and jurisdiction, UK-EU judicial cooperation post-Brexit and cooperation duties on online platforms to identify and remove hate crime and terrorist content in Germany and the UK. Dr Robinson previously wrote a doctoral thesis on law enforcement access to commercial data.

Enviado em: 06 de agosto de 2019

Aprovado em: 15 de novembro de 2019