ISSN 2384-9169
Fascicolo n. 2 - 2024
rivista.eurojus.it

Digital Exceptionalism, Freedom of Expression and the Rule of Law: The Case of Targeting Terrorist Content Online

VALSAMIS MITSILEGAS* E CLEMENTINA SALVI**

SOMMARIO. 1. Introduction. – 2. Criminal Law and terrorist speech. – 3. Criminal Law for the purpose of removing content. – 4. The Role of private providers in preventing the dissemination of terrorist content online. – 5. Normalisation of exceptionalism? - 6. Conclusion

1. Introduction

The European Union's (EU) approach to counter-terrorism has evolved to adapt to the technological advancements in the digital landscape, especially since the 2015 and 2016 terrorist attacks in Paris and Brussels.[1] The dissemination of easily accessible extremist textual, audio, and video content in the online environment had contributed to and even facilitated, the phenomena of so called foreign fighters, home-grown terrorism and lone wolves increasingly operating in Europe and other western jurisdictions.[2] The EU policy has progressively expanded

* Professor of European and Global Law and Dean of the School of Law and Social Justice at the University of Liverpool
** Dottoranda di ricerca presso la Queen Mary University of London

[1] K. LUYTEN, *Addressing the dissemination of terrorist content online* in *European Parliament Research Service* (PE 649.326), 2021, p. 2.

[2] The term 'foreign fighters' descrives 'individuals who travel to a State other than their States of residence or nationality for the purpose of the perpetration, planning, or preparation of, or participation in, terrorist acts or the providing or receiving of terrorist training, including in connection with armed conflict' (United Nations (UN) Security Council, Resolution 2178 (2014) adopted at its 7272nd meeting on 24 September 2014); while 'home grown terrorism' conceptualises the phenomenon of terrorist activities pusured by 'individuals being born and

in light of concerns related to the changes in the terrorist threat, shifting from a purely repressive approach to prevention, addressing broadly the factors that stimulate radicalisation and create incentives for terrorist activities.[3] This has led to both the criminalisation of terrorist speech and the introduction of a wider range of measures to prevent the dissemination of dangerous content online more broadly. In parallel, the need to tackle these new challenges has resulted in the progressive privatisation of surveillance and enforcement by imposing a series of legal duties on the private sector, which controls a massive volume of personal data,[4] thus encouraging the use of new technology and tools to improve automatic detection and removal of content that incites or relates to terrorist acts.[5] In this context, this contribution will argue that the EU regulatory approach to online targeting content has departed from the criminalisation of 'illegal' speech in the offline environment, at the expense of fundamental rights and the rule of law in its multifaceted declinations.[6] The first part of the chapter explores the process of criminalisation of the dissemination and publication of content that might - directly or indirectly - incite or encourage individuals to commit an act of terrorism. The second part considers the role of online platforms in policing illegal content by looking at the adopted regulation on addressing the dissemination of terrorist content online (TERREG), emphasising the risks beyond the use of automated tools and algorithms on content moderation. The third part expands the reach across the EU policy on targeting online content, by examining the recently adopted Digital Service Act (DSA), which covers illegal or otherwise harmful content, and the proposal for a Directive on combating violence against women or domestic violence, with a

raised in the West—or at least having a strong attachment to the West', highlighting the idea of individuals or groups acting on their own behalf without taking orders from a militant groupabroad (M. CRONE and M. HARROW, *Homegrown Terrorism in the West* in *Terrorism and Political Violence*, XXIII, 2011, p. 521; B. SCHUURMAN, E. BAKKER and Q. EIJKMAN, *Structural Influences on Involvement in European Homegrown Jihadism: A Case Study* in *Terrorism and Political Violence*, vol. 30, 2018, p. 97). The term 'lone wolf' refers to terrorist violence committed by lone actor extremist. It is worth hilighiting that it is a very contested term especially because it is believed it implies 'a high level of cunning and lethality that it is often not present among these individuals'(B. SCHUURMAN and others, *End of the Lone Wolf: The Typology That Should Not Have Been* in *Studies in Conflict & Terrorism*, vol. 42, 2019, p. 771. See also, N. BOUHANA and others, *Background and Preparatory Behaviours of Right-Wing Extremist Lone Actors: A Comparative Study*, Perspectives on Terrorism, vol. 12, 2018, p. 150).

[3] J. BURCHETT, A. WEYEMBERGH, *Counterterrorism Policies, Measures and Tools in the EU: An Assessment of the Effectiveness of the EU Counterterrorism Policy*, Study Requested by the LIBE Committee, PE 730.581, 2022, p. 69 available at: <https://data.europa.eu/doi/10.2861/38694> accessed 19 December 2022; V. MITSILEGAS, *European Criminal Law and the Dangerous Citizen* in *Maastricht Journal of European and Comparative Law*, vol. 25, 2018, p. 733.

[4] Council of 22–23 June 2017; The European Council of 28 June 2018. See in the extensive literature, T. GILLESPIE, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*, Yale University Press, 2018.

[5] V. MITSILEGAS, *The Privatisation of Surveillance in the Digital Age* in V. MITSILEGAS, N. VAVOULA (eds), *Surveillance and Privacy in the Digital Age*. *European, Transatlantic and Global Perspectives*, Hart Publications, 2021.

[6] V. MITSILEGAS, *Rule of Law: Theorising EU Internal Security Cooperation From a Legal Perspective* in R. BOSSONG and M. RHINARD (eds), *Theorising Internal Security Cooperation in the European Union*, Oxford University Press, 2016, p. 109; B. TAMANAHA, *On the Rule of Law: History, Politics, Theory*, Cambridge University Press, 2004; P. CRAIG, *Formal and Substantive Conceptions of the Rule of Law: An Analytical Framework*, Routledge, in *Public Law*, 1997, p. 467.

focus on the removal of online content in relation to offences of cyber violence. The impact on fundamental rights under the rule of law lens is eventually highlighted in the concluding remarks.

2. Criminal Law and terrorist speech

A bulwark of modern liberal societies, the right to freedom of expression and information – traditionally protected under national and international instruments – generally encompasses even radical and controversial extremist speech, which can only be restricted under exceptional, necessary and proportionate circumstances.[7] Within the EU, Article 11 of the European Charter on Fundamental Rights (CFREU) specifies the legitimate conditions under which freedom of expression can be limited, which is when interference is clearly prescribed by law, pursues a legitimate aim and has been shown to be necessary.[8] Counter-terrorism policies are emblematic of the conceivable limits that can be posed to freedom of speech for security reasons. However, as is explained below, public provocation and glorification offences can go too far in creating undue constraints on fundamental rights. This section shows the evolution of the EU path of criminalising terrorist speech.

The progressive expansion of criminal liability has always been justified as matter of urgency and in the light of the changes in the terrorist threats in the digital landscape. The waves of terrorist attacks that have occurred in Europe since 9/11 can be seen as representing a significant shift in both the character and the *modus operandi* of those involved: often 'homegrown' terrorist networks as well as self-recruited or self-trained individuals (lone wolves), radicalised via the internet and based in European countries, have become perpetrators of multiple attacks.[9] The connection between terrorist attacks and terrorist-related material – disseminated by, and accessed on, dedicated websites, forums, social networks and file-sharing websites – has been an object of concern for governments and law enforcement authorities alike.[10] The necessity to provide an adequate response has led to the progressive criminalisation of activities such as

[7] J. BARATA, *Terrorist content online and threats to freedom of expression: From legal restrictions to choreographed content moderation* in *Verfassungsblog*, 2022, available at: <https://verfassungsblog.de/os4-content-threats/> accessed 19 December 2022.

[8] Art. 11 CFREU recalls article 10 Article 10 of the European Convention on Human Rights (ECHR): The article corresponds to Article 10 of the European Convention on Human Rights; the limitations which may be imposed on the rights defined by article 11 of the European Charter may not exceed those provided for in Article 10(2) of the Convention.

[9] Already Madrid and London attacks, respectively in 2004 and 2005, had led to a change in the perception of the terrorist threat as no longer something external, but as an internal threat connected to "homegrown terrorism" (J. BURCHETT, A. WEYEMBERGH, cit., p. 16). Amongst the literature of those years, M. DEN BOER, *9/11 and the Europeanisation of Anti-Terrorism Policy: A Critical Assessment*, Policy papers n. 6, 2003, p. 31; KL. THACHUCK and others, *Homegrown Terrorism: The Threath Within*, in *Center for Technology and National Security Policy, National Defense University*, 2008; P. BRUNST, *Terrorism and the internet: new threats posed by cyber-terrorism and terrorist use of the internet*, in M. WADE and A. MALJEVIC (eds), *A war on terror? The European Stance on a New Threat, Changing Laws and Human Rights*, Springer, 2009.

[10] K. LUYTEN, cit., p. 2; UNODC, *The Use of the Internet for Terrorist Purposes*, in *United Nations Publications*, 2012, p. 3.

«the encouragement» of, «glorification» of and/or «apology» for terrorism (albeit in an undefined future and at undefined places) as well as the dissemination and the publication of relevant material[11] at both the national and international level.[12]

At the EU level in particular, the first legislative step in that direction was made in 2008, when the pre-Lisbon Framework Decision on combating terrorism – the first tool aimed at harmonising the definition of terrorist offences through criminal law – was amended by the Framework Decision 2008/919/JHA (from now on simply Framework Decision).[13] At that time, the extent of the EU competence to define criminal offences and criminal sanctions before the Lisbon Treaty, which expressly provides the Union's power to criminalise under Article 83 TFEU, has been a matter of debate due to the silence of EC Treaty on the matter. However, the introduction of Framework Decisions as a form of third pillar law in Amsterdam in order to respond to security issues related to terrorism was not questioned *per se*.[14] Terrorism has been seen as one of the most serious violation of the values of human dignity, freedom, equality and solidarity, and enjoyment of human rights and fundamental freedoms, on which the Union is founded, and thus an EU policy against it as indispensable and unavoidable.[15]

This latter required Member States to criminalise intentional acts of public provocation to commit a terrorist offence, the recruitment for purpose of terrorism, and training for purposes of terrorist activities.[16] Public provocation to commit a terrorist offence includes «the distribution, or otherwise making available, of a message to the public, with the intent to incite the commission of one of the offences listed […], where such conduct, whether or not directly advocating terrorist offences, causes a danger to one or more such offence may be committed».[17] According to the Framework Decision, «these forms of behaviour should be

---

[11] F. GALLI, *Freedom of Thought or "Thought-Crimes"? Counter-Terrorism and Freedom of Expression* in A. MASFERRER and C. WALKER (eds), *Counter-Terrorism, Human Rights and the Rule of Law: Crossing Legal Boundaries in defence of the State*, Edward Elgar Publishing, 2013, p. 106.

[12] A panorama of member states legislative initiative of those years can be explored in Council of Europe (ed), "Apologie du Terrorism" *and Incitement to terrorism*, 2007, p. 12.

[13] The 2008 EU framework decision has followed the Council of Europe Convention for prevention of Terrorism 2005 (STE 196), but already since 2004-2005 radicalisation has become a core element of the EU holistic approach to counter terrorism.

[14] The adoption of 2002 Framework Decision was rather delayed because of three parliamentary scrutiny reserves. The concerns expressed by the European Parliament regareded the necessity to provide a clear enough distinction between terrorist acts and demonstrators (J. MONAR, *The European Union's response to 11 September 2001: Bases for action, performance and limits*, in *Albany Edu*, available at https://www.albany.edu/~rk289758/BCHS/col/JHA-TERRORISM-NEWARK.doc (accessed 10 March 2023). Differently, the Community compentence in the field of criminal offences and sanctions have been largely discussed in the form of inter-institutional battles in the proposal ald negotiation of Community legislation in other areas of law such environmental crime. For completenss on the debate before and after Lisbon, see '*Substantive Criminal Law: From Securitised to Functional Criminalisation*' in V. MITSILEGAS, in *EU Criminal Law*, Hart publications, II ed., 2022.

[15] Council Framework decision on combatting terrorism (2002/475/JHA) OJ L 164 of 22.06.2002. See for an overall, M. DEN BOER, *9/11 and the Europeanisation of Anti-Terrorism Policy: A Critical Assessment*, in Policy Paper Notre Europe – Institute Jacques Delors, Vol. 6,v2003, p. 6.

[16] Prior to that, basically terrorist offences were limited set a list of intentional acts aimed at seriously intimidating the population, compelling governments to carry out or abstaining from certain acts, or seriously destabilising the order of the targeted countries. While the intention was certainly a key element the focus was still on material acts.

[17] Council Framework Decision 2008/919/JHA of 28 November 2008 amending Framework Decision 2002/475/JHA on combating terrorism, Art. 3 as replaced.

equally punishable in all Member States irrespective of whether they are committed through the Internet or not».[18] The role of the internet was highlighted as a source of inspiration and mobilisation to local terrorist networks and individuals in Europe and also as a container of information on terrorist means and methods, functioning as a «virtual training camp».[19] The Framework Decision thus opted for the criminalisation of activities only linked to those commonly referred to as terrorist offences, in order to contribute to a wider general policy objective of preventing terrorism through the removal of those materials which might incite persons to commit terrorist attacks.[20] This legislative development was identified as part of a broader paradigm shift towards prevention in criminal law,[21] which is still ongoing, targeting "dangerous" behaviours and "potential future harms", with negative consequences in terms of freedoms and fundamental rights.[22] In fact, the new inchoate offences were widely criticised not only because they moved significantly from the actual commission of terrorist acts, criminalising the dissemination of ideas or opinions that *might* incite attacks, but also for the lack of clear and precise definitions, impacting on the legitimate limits to freedom of expression (art. 11 CFREU or Charter) as well as the principle of legality and foreseeability of the consequences of individuals' conducts (art. 49 Charter).[23]

The 2015-2016 attacks in Paris and Brussels as well as the significant flow of EU citizens travelling to Syria and Iraq in the previous years have given further impetus to the fight against terrorism.[24] In this respect, a catalyst was also a series of scandals relating to the misuse of online services by terrorist groups and their supporters to spread their message, radicalise others, and facilitate terrorist activities.[25] For example, the Daesh terrorist organisation had allegedly attracted several EU citizens by using an influential communication strategy together with the dissemination of online propaganda aimed at recruiting new 'followers'.[26] During this period, the objective of combating terrorism and preventing radicalisation became a priority in the agenda of European programmatic instruments dealing with broader security-related issues.[27]

---

[18] Ibidem, Art. 11.

[19] Ibidem.

[20] Ibidem, Art. 7.

[21] V. MITSILEGAS, *"Security Law" and Preventive Justice in the Legal Architecture of the European Union* in U. SIEBER, V. MITSILEGAS, C. MYLONOPOULOS, E. BILLIS and N. KNUST (eds), in *Alternative Systems of Crime Control – National, Transnational and International Dimension*s, Duncker and Humblot, 2018, p. 203; P. ASP, *Preventionism and Criminalization of Nonconsummate Offences* in A. ASHWORTH, L. ZEDNER AND P. TOMLIN (eds) in *Prevention and the Limits of Criminal Law*, Oxford University Press, 2014, p. 33.

[22] K. SUGMAN STUBBS, F. GALLI, *Inchoate offences. The sanctioning of an act prior to and irrespective of the commission of any harm* in A.WEYEMBERGH and F. GALLI (eds), *EU counter-terrorism offences. What impact on national legislation and case-law?*, Éditions de l'Université Libre de Bruxelles, 2012, p. 291 ff.

[23] The introduction of such inchoate offences were initially described exceptional and legitimised by their emergency of fight against terrorism, targeting only specific groups of people, Francesca Galli, cit., p. 124.

[24] J. BURCHETT, A. WEYEMBERGH, cit., p. 17.

[25] Ibidem; G. ROBINSON, *A Democratic Dénouement? The EU vs Terrorist Content Online*, in *Revista Publicum*, vol. 5, 2019, p. 189.

[26] K. LUYTEN, cit., p. 2.

[27] Commission, 'The European Agenda on Security' COM(2015) 185 final (28 April 2015).

A turning point was in 2017 with the introduction of Directive (EU) 2017/541 on combating terrorism (from now on simply 'Directive'), which replaced the amended Framework Decision 2002/475/JHA. The Directive, based on Article 83(1) TFEU, was proposed about three weeks after the terrorist attacks in Paris in November 2015 and rapidily adopted without the lack of an impact statement (see below), is currently still the main criminal law instrument at the EU level, imposing minimum rules in the area of terrorist offences, offences related to a terrorist group and offences related to terrorist activities. The Directive was proposed in response to similar standard-setting initiatives by the UN and the Council of Europe to address the developing terrorist threat associated primarily with the phenomenon of «an increasing number of individuals who travel abroad for the purposes of terrorism and the threat they pose upon their return», commonly referred to as foreign terrorist fighters.[28] It is worth noticing that this is a classic example of the concerning tendency by the Commission of what has been called the «re-emergence of emergency», since the proposal relied on *ex-post* evaluations of the Framework Decisions and on stakeholder consultations in the context of negotiations of the Council of Europe Additional Protocol.[29] The legislative proposal by the Commission was not even accompanied by an impact assessment and the legislative process was expedited, stating that this was necessary due to the «urgent need» to improve the EU security framework.[30]
The Directive lists a number of acts that Member States must qualify as terrorist offences as committed with a specific aim, such as travelling for the purpose of terrorism or receiving training for terrorism, providing Member States a certain margin of discretion to define these acts.[31] The Directive gained significant public attention and its impact on fundamental rights was among the key issues discussed during the legislative process and beyond, «considering the multiplication of "inchoate offences" defined in broad terms».[32] Overall, the Directive has three main points of friction with the rule of law and human rights.[33] The first is the presence in general of loose definitions that reduce «legal clarity», including a vague and broad definition of terrorism that clearly deviates from those provided at the UN level, such as the Security Council Resolution 1566 or the 1999 Terrorism Financing Convention, and from the Council of Europe Convention on the Prevention of Terrorism.[34] The second concerns the already mentioned (over)criminalisation of many preparatory acts that may be remote from intrinsically harmful conduct.[35] Finally, the existence of ancillary offences that are also potentially

[28] V. MITSILEGAS, *Counterterrorism and the Rule of Law in an Evolving European Union*: Plus Ça Change ? in 12 *New Journal of European Criminal Law*, vol. 12, 2021, p. 37.
[29] Ibidem p. 43.
[30] FRA, *Directive (EU) 2017/541*, cit., p. 13; V. MITSILEGAS, *Counterterrorism*, cit., p. 41.
[31] Ibidem p. 13.
[32] J. BURCHETT, A. WEYEMBERGH, cit., p. 14.
[33] T. GHERBAOUI, M. SCHEININ, *Time to Rewrite the EU Directive on Combating Terrorism*, in *Verfassungsblog*, 2022, <https://verfassungsblog.de/time-to-rewrite-the-eu-directive-on-combating-terrorism/> accessed 19 December 2022.
[34] This has to be further contextualised in the challenges in defining 'terrorism' in general and the lack of an internationally agreed-upon definition see M. LLOBET ANGLÌ, in A. MASFERRER, C. WALKER (eds), *Counter-Terrorism, Human Rights and the Rule of Law: Crossing Legal Boundaries in defence of the State*, Edward Elgar Publishing, 2013.
[35] V. MITSILEGAS, *Counterterrorism*, cit., p. 46.

accumulable. In the absence of any meaningful human rights impact assessment, the Directive included a clause providing for a five-year review by the Commission that would explicitly consider this. In November 2021, the Commission submitted a report to the European Parliament and the Council on the implementation of the Directive by Member States, assessing that Directive has achieved its objective and declaring that freedom and rights limitations largely meet the requirements of necessity and proportionality.[36] As persuasively argued, these considerations did not entirely consider the concerns expressed by some stakeholders, in particular, the EU Fundamental Rights Agency (FRA).[37] The 2021 FRA report includes a swift empirical assessment of the Directive's human rights implications based on extensive fieldwork, including interviews with experts and practitioners in several of the EU Member States (Belgium, Germany, Greece, Spain, France, Hungary and Sweden).[38]

In relation to terrorist speech specifically, Article 5 of the Directive recalls the offence of public provocation to commit a terrorist offence as provided in the Framework Decision, which also in this case covers indistinctly both online and offline speech. The offence necessitates an act *i)* of communicating a message advocating – directly or indirectly – the commission of terrorist offences and *ii)* causing an objective danger that an offence will be committed as a result of the act of communication (but there is no need for a terrorist crime to be actually prepared or attempted as a result of the provocation).[39] With regard to *mens rea*, the intent to incite the commission of such offences is required.

The scope of this is broader if compared to the Framework Decision, also encompassing glorification and justification of terrorism (direct as well as indirect provocation). As specified in Recital 10, «The offence of public provocation to commit a terrorist offence act comprises, inter alia, the glorification and justification of terrorism or the dissemination of messages or images online and offline, including those related to the victims of terrorism as a way to gather support for terrorist causes or to seriously intimidate the population». Even if recital 40 clarifies that the definition of the offence excludes expressing radical, polemic or controversial views in the public debate on sensitive political questions, it remains unclear how to adequately set the distinction between lawful and unlawful speech.[40]

Another problem of the Directive relates to the assessment of intent. Determining whether the identified speech or content has been disseminated with terrorist intent emerges as a major challenge, especially considering the different types of expression and the different channels of communication, both online and offline.[41] These difficulties have a clear impact «both to the principle of legality and proportionality of criminal offences and penalties (Article 49 Charter and Article 7 ECHR) and to freedom of expression and information (Article 11 Charter and

[36] Commission Report based on Article 29(2) of Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combatting terrorism and replacing the Council Framework Decision. 2002/457/JHA and amending Council Decision 2005/671/JHA.

[37] T. GHERBAOUI, M. SCHEININ, cit..

[38] European Union Agency for Fundamental Rights (FRA), *Directive (EU) 2017/541 on Combating Terrorism. Impact on Fundamental Rights and Freedoms Report*, Publications Office, 2021, pp. 6-7 (from now on FRA *Directive (EU) 2017/541*).

[39] FRA *Directive (EU) 2017/541*, cit., p. 51.

[40] V. MITSILEGAS, *Counterterrorism*, cit., p. 36.

[41] FRA, *Directive (EU) 2017/541*, cit., p. 56.

Article 10 ECHR)». Because of such nebulous definitions contained in the Directive, it is unsurprising that the Commission's report flags the fact that «several national authorities and judges reported difficulties in proving terrorist intent».[42] Most Member States in fact do not provide concrete guidelines on determining intent, as a means of guaranteeing protection from arbitrariness, and the concept of intent sometimes even appears to be ignored altogether.[43] Furthermore, the notion of "danger" that one or more terrorist offences can be committed, can be used to condemn a specific opinion rather than prevent a real danger, as respondents highlighted. This infringes on freedom of thought, conscience and religion (Article 10 Charter and Article 9 ECHR). Finally, respondents expressed concerns about the discriminatory impact of the provisions on specific groups (Article 21 Charter and Article 14 ECHR).[44]

On closer examination, these general issues are amplified when it comes to online speech, leaving room for arbitrariness and potential «cherry-picking».[45] French case law, especially on consulting the Internet, is particularly emblematic of these issues in the online space. Interviewees draw attention to jurisprudence concerning the French offence of *apologie du terrorism* which broadly covers condoning or inciting terrorism, including the favourable presentation of acts of terrorism and their perpetrators.[46] In fact, notwithstanding the criticism over its scope, the French courts have held that the offence is sufficiently precise to guarantee against the risk of arbitrariness and does not violate the principle of legality of criminal offences.[47] Particularly controversial is the case of the comedian Dieudonné, who was sentenced to 2 months imprisonment for *apologie* after having posted on social networks «*Je me sens Charlie Coulibaly*».[48] The Court of appeal rejected the arguments made by the defence against the constitutionality of Article 421-2-5 of the national Criminal Code, which would not clearly define the constituent elements of the offence. Also worrying are the declarations of some practitioners in Belgium, Germany and Greece that stated that «prosecuting provocation is not a problem, as the penalised behaviour is often "obvious" […] "I know it when I see it", says a law enforcement respondent about the criteria for identifying terrorist content online».[49] By contrast, some defence lawyers stress that «the criminal law response to public provocation is disproportionately severe in relation to the conduct, as it can entail long prison sentences, for example for posting content on social media. That limits freedom of speech more than is necessary».[50]

3. Criminal Law for the purpose of removing content

[42] Ibidem.

[43] Difficulties in proving terrorist intent, which mainly result from factual circumstances have also been highlighted by the same Commission, see above.

[44] FRA, *Directive (EU) 2017/541*, cit., p. 56.

[45] Ibidem p. 55.

[46] Ibidem p. 54.

[47] Ibidem.

[48] Ibidem.

[49] Ibidem p. 53.

[50] Ibidem p. 55.

A key aspect of digital exceptionalism relates to the development of content removal measures for preventive purposes: online terrorist content receives special legal treatment in comparison to offline content. The criminalisation of terrorist content is in fact coupled with additional administrative measures targeting online content.

The Directive, in addition to the introduction of minimum standards for the definition of terrorist offences and offences related to terrorism, provides a regulated obbligation on Member States to remove online public provocation content. In particular, Article 21 of the Directive obliges Member States to ensure the swift removal of, or blocking of access to, online content constituting the offence of public provocation according to Article 5, leaving to Member States the choice of the measures to adopt.[51] This combination of measures was considered necessary for combatting the radicalisation of individuals which can lead to terrorist acts, signing the first step towards legally binding provisions on the removal of online terrorist content, even if addressed not to private providers directly, but to EU Member States.[52]

The Commission went a step further in this direction in September 2018 with the Proposal for a Regulation on preventing the dissemination of terrorist content online,[53] eventually adopted after four years of negotiation in April 2021 as the EU Regulation on addressing the dissemination of terrorist content online (TERREG or The Regulation). This time, in order to ensure the effective detection of terrorist content, the Regulation directly imposes duties of care on hosting service providers to remove terrorist content. In particular, it introduces removal orders that oblige hosting service providers to remove or block access to flagged terrorist content within one hour of reception of a notification from a national competent authority (judicial or administrative). The Regulation also requires hosting service providers to take proactive measures, proportionate to the level of risk, and to remove terrorist material from their services, including by deploying automated detection tools. Online platforms thus have a legal duty to identify illegal or disallowed speech.[54] It is worth highlighting that the Regulation uses as a legal basis Article 114 TFEU, related to the harmonisation of rules in the internal market, which – however – in this case, serves the security objectives of the European Union, specifically the prevention of terrorism.[55] According to the EU Commission proposal, «Article

[51] *Directive (EU) 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 [2017] OJ L 88/6* (Terrorism Directive) Art 21. Recitals 22 and 23 specify that an effective means of combating terrorism on the internet is to remove online content constituting a public provocation to commit a terrorist offence at its source.

[52] V. MITSILEGAS, *The Privatisation*, cit..

[53] *Proposal for a Regulation on preventing the dissemination of terrorist content online* COM(2018) 640 final. See also the Commission Recommendation (EU) 2018/334 of 1 March 2018 on measures to effectively tackle illegal content online [2018] OJ L63/50. The Recommendation included a specific chapter laying down numerous measures to effectively stem the uploading and sharing of terrorist propaganda online, such as improvements to the referral process, a one-hour timeframe for responding to referrals, more proactive detection, effective removal, and sufficient safeguards to accurately assess terrorist content (recital 5 and pt 18).

[54] SERGIO CARRERA, VALSAMIS MITSILEGAS and others, 'Towards a Principled Level Playing Field for an Open and secure Online Environment', in *CEPS Task Force Report* (from now on simply: *CEPS task force report*), Brussels, 2022, p. 18.

[55] J. BURCHETT, A. WEYEMBERGH, cit., pp. 76-77.

114 is the appropriate legal basis to harmonise the conditions for hosting service providers to provide services across borders in the Digital Single Market and to address differences between Member State provisions which might otherwise obstruct the functioning of the internal market».[56]

In this sense, Germany expressed its doubt on the adequacy of the legal basis of the proposal, which would not encompass the clearly covered security issues.[57]

The proposal initially defined terrorist content in very broad terms, going significantly further than Article 5 of the Terrorism Directive. Firstly, by omitting the element of intent altogether, which may risk automatic deletion, irrespective of the context of its use (i.e. for confrontation, reporting, research or historical purposes).[58] Secondly, the proposal broadened the scope of expression that would be considered «terrorist» by including material encouraging the contribution, participation or support of terrorism or a terrorist group. In this way, the first draft presented a worrying situation where an administrative law measure eroded the already controversial legal certainty level reached in the criminal law definition of terrorism under the Directive.[59] In the course of the long legislative procedure, multiple stakeholders – especially human rights organisations – as well as the UN special rapporteurs,[60] widely criticised the proposal and its impact on fundamental rights such as freedom of expression, the right to information and the right to privacy.[61] Eventually, as noted, the final text is marginally better than the initial draft.[62] Firstly, recital 12 explicitly excludes material disseminated for educational, journalistic, artistic, or research purposes, or awareness-raising purposes against terrorist activity for the scope of TERREG. Secondly, the definition of terrorist content is now expressly aligned with the Directive, as a criminal action *aimed* at seriously intimidating the population, compelling governments to carry out or abstain from certain acts, or seriously

[56] *Proposal for a Regulation*, cit.

[57] K. LUYTEN, cit., p. 6. See *Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Verhinderung der Verbreitung terroristischer Online-Inhalte* COM(2018) 640 final; Ratsdok. 12129/18 (https://www.europarl.europa.eu/RegData/docs_autres_institutions/parlements_nationaux/com/2018/0640/DE_BUNDESRAT_CONT1-COM(2018)0640_DE.pdf)

[58] G. ROBINSON, cit., p. 189.

[59] V. MITSILEGAS, *The Privatisation*, cit.

[60] UN Special Rapporteurs, *Mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, the Special Rapporteur on the right to privacy and the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism*, 7 December 2018, p. 3; Open letter on behalf of civil society groups regarding the proposal for a regulation on Terrorist content online, available at: <https://edri.org/wp-content/uploads/2020/11/TERREG_Openletter_Liberties.pdf> accessed 30 December 2022.

[61] Ibidem. The definition – as it was – could encompass legitimate forms of expression, such as reporting conducted by journalists and human rights organizations on the activities of terrorist groups and on counterterrorism measures taken by authorities, in violation of the right to freedom of expression as protected under Article 19 of the International Covenant on Civil and Political Rights (ICCPR), Article 10 of the ECHR and Article 11 of the Charter. See also European Union Agency for Fundamental Rights (FRA), *Opinion 2/2019 on the Proposal for a Regulation on Preventing the Dissemination of Terrorist Content Online and Its Fundamental Rights Implications* (Publications Office 2019) available at: <https://data.europa.eu/doi/10.2811/818523> accessed 16 December 2022 (from now on: FRA *Opinion 2/2019*)

[62] *CEPS Task Force Report*, cit., pp. 43-44; 88.

destabilising the order of the targeted countries.[63] In particular, «Terrorist content» means material that «incites» or «advocate», directly or indirectly, the commission of one of the terrorist offences referred in the Directive, «such as by the glorification of terrorist acts»;[64] material that «solicits» someone to commit or to contribute to terrorist offences or to participate in activities of a terrorist group;[65] material that provides «instruction» on how to conduct attacks;[66] and material that constitue a threat to commit a terrorist offence.[67]

Competent authorities and hosting service providers should assess whether the material constitutes terrorist content within the meaning of this Regulation by taking into account certain factors such as «the nature and wording of statements, the context in which the statements were made and their potential to lead to harmful consequences in respect of the security and safety of persons».[68]

Nevertheless, despite the improvement, the definition adopted significantly broadens the scope of the obligation to online speech, expanding from public provocation to other material. Given its preventative nature, the Regulation covers not only material inciting terrorism but also material for recruitment or training purposes, reflecting other offences related to terrorist activities, which are also covered by the Directive. Since the definition of online terrorist content relies on the efinition of terrorist and terrorism-related offences in the directive, its implementation will once more assess the clarity and predictability of these provisions. As highlighted, «These challenges and concerns are likely to increase when the removal obligation is also applied to other content that is considered to be of a terrorist nature, not just content that clearly constitutes public provocation».[69] Furthermore, the addition in the text (compared to the proposal) that terrorist content can involve a 'threat' to commit a terrorist offence (see lett. (e)) has «the potential to become a residual category, as opposed to the more precise wording of the other kinds of material that constitute terrorism content».[70]

To conclude, the future implementation of this Regulation, which has applied from 7 June 2022, still gives rise to concerns. Especially considering the shortcomings of the Directive as shown by the assessment of its transposition,[71] many non-governmental organisations continue to see the new Regulation as a significant threat to freedom of expression, especially in relation to the concept of terrorist content, which has not been remedied by the compromise text between the European Parliament and the Council.[72]

---

[63] *CEPS Task Force Report*, cit., pp. 43-44.

[64] Art. 2 (7) lett. a) of the Regulation.

[65] Art. 2 (7) lett. b) and c) of the Regulation.

[66] Art. 2 (7) lett. d) of the Regulation.

[67] Art. 2 (7) lett.e) of the Regulation.

[68] Recital 11 of the Regulation.

[69] FRA *Directive (EU) 2017/541*, cit., p. 61.

[70] *CEPS Task Force Report*, cit., p. 88.

[71] J. BURCHETT, A. WEYEMBERGH, cit., p. 78.

[72] M. POLLET, *EU Adopts Law Giving Tech Giants One Hour to Remove Terrorist Content*, in *Euractiv*, 2021, available at: <https://www.euractiv.com/section/cybersecurity/news/eu-adopts-law-giving-tech-giants-one-hour-to-remove-terrorist-content/> accessed 30 December 2022.

4. The Role of private providers in preventing the dissemination of terrorist content online

The previous part illustrated the controversial EU path of regulation of terrorist speech that culminated in the TERREG, focusing on the scope, scheme, and definition of terrorist content introduced in the Regulation. This section looks at the duties imposed on private companies as well as fundamental rights concerns connected to these duties.[73] Since 2014, the EU has sought ways to control propaganda from terrorist organisations circulated via social media.[74] Private internet companies play a crucial role in hosting or facilitating the flow of online content, so there was a firm consensus on the need to involve them in efforts to prevent the propagation of extremist/terrorist content online.[75] The first efforts at EU level to counter terrorist content online commenced in 2015 through a framework of voluntary cooperation between Member States and hosting service providers in the EU Internet Forum by the Commission.[76] This was followed by analogous initiatives such as the 2016 Code of Conduct on Countering Illegal Hate Speech Online, to prevent and counter the rise of this phenomenon,[77] and the 2017 Global Internet Forum to Counter Terrorism which involved the major private-sector actors, with the aim of disrupting terrorist abuse of members' digital platforms through the creation of databases for terrorist material.[78]

---

[73] V. MITSILEGAS, *The Privatisation*, cit.

[74] As illustrated by R. GORWA, *The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content*, in *Internet Policy Review*, vol. 8, 2019, <https://policyreview.info/articles/analysis/platform-governance-triangle-conceptualising-informal-regulation-online-content> accessed 16 November 2022, back in 2010, The Netherlands, the UK, Germany, Belgium and Spain sponsored a European Commission project called 'Clean IT'. This latter would provide "general principles and best practices" on combating terrorist content and "other illegal uses of the internet [...] through a bottom up process where the private sector will be in the lead". GORWA highlighted 'The Clean IT coalition, which featured significant representation from European law enforcement agencies, initially appeared to be considering some very hawkish proposals (such as requiring all platforms to enact a real-name policy, and that "Social media companies must allow only real pictures of users"), leading to a push-back from civil society and the eventual end of the project. However, the project seemed to set the ideological foundations for the EU's approach to online terrorist content by advocating for more aggressive terms of service and industry takedowns without explicit legislation'. See also R. BELLANOVA and M. DE GOEDE, *Co-Producing Security: Platform Content Moderation and European Security Integration* in *Journal of Common Market Studies*, vol. 60, 2022, pp. 1316-1317. Emblematic of this is the Global Coalition's actions against Daesh's propaganda to which EU is a party, available at: <https://theglobalcoalition.org/en/mission/countering-daeshs-propaganda/> accessed 26 December 2022.

[75] J. BURCHETT, A. WEYEMBERGH, cit., p. 75.

[76] The Forum, composed by the Commission, Member States, the main private companies, Europol and other experts and civil society partners, provides a platform where these actors can exchange their views on trends, developments and the challenges posed by the presence of terrorist content online in order to reach *a joint, voluntary* approach based on a *public–private partnership* to detect and address harmful material online. Another major initiative is the setting up of the Internet Referral Unit (EU IRU) at Europol in 2015 following to the Charlie Hebdo attacks, to respond to the demand for closer collaboration with the Internet industry and to prevent radicalisation and related propaganda.

[77] Commission *Code of Conduct on Countering Illegal Hate Speech Online*, 30 June 2016.

[78] The original Forum was led by a rotating chair drawn from the founding four companies—Facebook, Microsoft, Twitter and YouTube (Google? No: https://gifct.org/membership/). In 2019 they were followed by Dropbox,

Some of the main platforms already had their internal content moderation policies and initiatives to ensure «online safety» by identifying, labelling, filtering, reporting, suspending and removing content that is defined as «harmful», «illegal» or «terrorist» according to national, international, or supranational regulations applicable in the various jurisdictions under which they operate.[79] These are set by online platforms under the Companies' Terms of Use, Terms of Services, Terms and Conditions (ToS), or Community guidelines.[80] As examined, «content moderation policies defined in ToS and/or community guidelines are often phrased vaguely and too broadly».[81] For example, the UN Special Rapporteur has expressed serious concerns regarding the Facebook (now Meta) community guidelines on dangerous individuals and organisations. These require the removal of content generated by a wide range of groups, not only terrorism or hate speech perpetrators, but also other violent non-state actors. These are mainly organisations that incite violence against a state, but not civilian or military organisations, even organisations which do not advocate violence at all, but are viewed as demonstrating an intent to do so.[82] For instance, the Special Rapporteur stressed that Facebook's designation of violent content was vague, imprecise, and could lead to the criminalisation of speech as protected under Article 19 of the ICCPR.[83]

Big tech companies have developed and used technological means, specifically semi or automated AI tools for content moderation, also known as «algorithmic moderation systems».[84] Online platforms often combine automated decision-making processes that are able to assess and detect the relevant content with human oversight (the decision is then checked by a team of human moderators).[85] The most used machine learning tools include flagging and filtering practices.[86] The latter, especially upload filters, are highly controversial.[87] One of the main issues is that – so far – the reliability and quality of the AI tools used in content moderation are

---

Amazon, WhatsApp. GIFCT membership is limited to companies operating Internet platforms and services. This group has also been criticised for its secretive processes and selective membership (see R. GORWA, R. BINNS and C. KATZENBACH, cit.); K. LUYTEN, cit., p. 3.

[79] *CEPS task force report*, cit., p. 16; S. T ROBERTS, *Behind the Screen. Content Moderation in the Shadows of Social Media*, Yale University Press, 2019; R. BELLANOVA and M. DE GOEDE, cit., p. 1316; K. KLONICK, *The New Governors: The People, Rules, and Processes Governing Online Speech*, in *Harvard Law Review*, vol. 131, pp. 1630 ff.

[80] *CEPS Task Force Report*, cit., pp. 21-22.

[81] Ibidem, p. 21.

[82] Ibidem, p. 23.

[83] Ibidem; see Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism (2021), *Input of the UN Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism to the Facebook Oversight Board Concerning its "Community Guidelines" and "Community Standard on Dangerous Individuals and Organizations*"'.

[84] R. BELLANOVA and M. DE GOEDE, cit., p. 1326; R. GORWA, R. BINNS and C. KATZENBACH, *Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance* in *Big Data & Society*, 7, 2020. p. 1317.

[85] *CEPS Task Force Report*, cit., p. 23.

[86] R. BELLANOVA and M. DE GOEDE, cit., p. 1326 ff.

[87] See the CJEU case law on filtering use of internet service provider for copyright breaches. In Case C-360/10 *Sabam v Netlog* [2012], the CJEU concluded that a filtering system, which targets a specific type of content while indiscriminately monitoring all information shared by platform users for unlimited period of time, amounts to such a prohibited general monitoring obligation.

still limited.[88] Additionally, these instruments have often proved to be biased by the training datasets that may incorporate the discriminatory assumptions involved in their development (for example targeted religion or categories).[89] Furthermore, it can be hard or even impossible to realise whether some discriminatory results occur as result of AI limits.[90] To avoid unlawful restrictions, more recently, social media platforms have also created their own internal mechanisms of adjudication against decisions made to remove or block users' content.[91] An important experiment undertaken by global online platforms in this domain is provided by the Facebook (now Meta) Oversight Board (MOB). Concerns have been raised on the legitimacy of these non-state mechanisms in relation to freedom of expression. For example, in Germany, the German Federal Supreme Court (BVerfG), ruled that Meta can develop its own internal rules prohibiting certain types of speech and can enforce those rules by removing posts and closing accounts breaching those rules.[92] However – because of the size of the company and its dominant position in the market, it should comply with clear due process requirements. For instance, this means that Meta should inform its users (at least *ex post*) of the removal of their content, and of its intention to block users' accounts.[93] A further issue concerns the accessibility of such remedies.[94] As noted, users have rarely the chance to demand a review of a case to these high-profile entities.[95]

In this scenario, characterised by the massive accretion of private power over-people's-speech,[96] the EU seeks to play an active role in controlling moderation policies, justified by the need to provide a clear legislative framework to further reduce accessibility to terrorist content and – at

[88] *CEPS Task Force Report*, cit., p. 24.

[89] Ibidem 25; *FRA Report, Bias in Algorithms: Artificial Intelligence and Discrimination*, 2022, pp. 49 ff. available at: <https://data.europa.eu/doi/10.2811/536044> accessed 30 December 2022; G. MAUGERI, *Automated Decision-making in the EU Member States: The Right to Explanation and Other 'Suitable Safeguards in the National Legislations*, in *Computer Law and Security Review*, vol. 5, 2019, p. 35; S. MACDONALD, S. GIRO CORREIA and A.L. WATKIN, *Regulating Terrorist Content on Social Media: Automation and the Rule of Law*, in *International Journal of Law in Context*, vol.15, 2019, p. 183.

[90] *CEPS Task Force Report*, cit., p. 2.

[91] The creation of the Oversight Board was anticipated by findings of critical gaps in Meta's internal 'grievance mechanisms': available at: <https://www.oversightboard.com/> accessed 30 December 2022. See chapter X in this volume.

[92] Case III ZR 192/20, III ZR 179/20 [2021] BVerfG.

[93] see also Case C-131/12 *Google Spain SL and Google Inc v. Agencia Española de Protecciín de Datos (AEPD)* and *Mario Costeja Gonzalez*, Judgment of 13 May 2014.

[94] These issues have been partially covered by the recently adopted Digital Service Act (DSA) see infra section 4. D. HOLZNAGEL, *A Self-Regulatory Race to the Bottom through Art. 18 Digital Services Act* in *Verfassungsblog*, 2021, <https://verfassungsblog.de/a-self-regulatory-race-to-the-bottom-through-art-18-digital-services-act/> accessed 16 November 2022.

[95] Ibidem.

[96] D. KAYE, *Speech Police – The Global Struggle to Govern the Internet*, in *Columbia Global Reports*, 2019, p. 126. See also J. M. BALKIN, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, in *University of California Davis Reviews*, vol. 51, 2018, p. 1151; E. COCHE, *Privatised Enforcement and the Right to Freedom of Expression in a World Confronted with Terrorism Propaganda Online*, in *Internet Policy Review*, vol. 7, 2018; D. KEATS CITRON, *Extremist Speech, Compelled Conformity and Censorship Creep*, in *Notre Dame Law Review*, vol. 93, 2018, p. 1035.

the same time – ensure the protection of the fundamental rights at stake.[97] The adoption of TERREG is key to the development of an EU normative framework on private-public partnerships in the field of online content moderation and the removal of content online.[98] As anticipated, the Regulation delegates private service providers' 'law enforcement duties' to remove or assess the nature of online content in ways that are both reactive, in response to national competent authorities orders,[99] and proactive, by including in its terms and conditions and applying provisions to address the misuse of its services for the dissemination to the public of terrorist content.[100] To enable the detection, identification and removal of content by online service providers, the Regulation legitimises the use of the above-mentioned automation practices in content moderation and the use of machine learning tools to perform online surveillance duties.[101] Given the number of concerns raised during the negotiations, the Regulation remarks the importance of preserving fundamental rights and the importance of providing human verification where automated tools are involved in the terrorist content detection, both in the recitals and in a number of provisions of the legal text.[102]

According to the final text, given the scale of the problem, hosting service providers should decide what are the most suitable, effective and proportionate to identify and remove terrorist content.[103] In this respect, «specific measures could include appropriate technical or operational measures or capacities such as staffing or technical means to identify and expeditiously remove or disable access to terrorist content, mechanisms for users to report or flag alleged terrorist content, or any other measures the hosting service provider considers appropriate and effective to address the availability of terrorist content on its services».[104]

Hence – differently from the proposal – there is expressly no obligation to use automated tools, but it is up to the private sector thus to find the best measures to deal with terrorist content. That said, hosting service providers are required to «ensure that users' rights to freedom of expression and information as well as the freedom and pluralism of the media as protected under the Charter are preserved».[105] Furthermore, they should also respect any requirement laid down in the law, including legislation on the protection of personal data, act with due diligence and implement safeguards, including human oversight and verifications, in order to «avoid any unintended or erroneous decision leading to the removal of or disabling of access to content that is not terrorist content».[106]

---

[97] The Commission in 2018 also adopted the 2018/334 Recommendation on measures to effectively tackle illegal content online [2018] OJ L 63/50.

[98] *CEPS Task Force Report*, cit., p. 11.

[99] Art. 3 and 4 of the Regulation.

[100] Art. 5 of the Regulation.

[101] *CEPS Task Force Report*, cit., p. 11.

[102] F. GIGLIO, *The New Regulation on Addressing the Dissemination of Terrorist Content Online: A Missed Opportunity to Balance Counter-Terrorism and Fundamental Rights?* in *CITIP blog*, 2021, available at: <https://www.law.kuleuven.be/citip/blog/the-new-regulation-on-addressing-the-dissemination-of-terrorist-content-online/> accessed 30 December 2022.

[103] Recital 22 of the Regulation.

[104] Recital 22 of the Regulation.

[105] Recital 23 of the Regulation.

[106] Ibidem.

Article 3 of the TERREG regulates removal orders and can be issued by a competent authority – not necessarily a judicial authority – in each of the Member States, requiring hosting services to remove terrorist content or to disable access to such content, in the whole European Union, as soon as possible and in any event within one hour of receipt of the removal order. The removal order should contain a statement of reasons qualifying the material to be removed or access to which is to be disabled as terrorist content and provide sufficient information for the location of that content.[107]

Article 5 concerns the specific measures that service providers shall include in its terms and conditions and apply provisions to address the misuse of its services for the dissemination to the public of terrorist content. The decision regarding the specific measures to adopt is always up to the hosting providers. These may incude, for example, staffing or technical means to identify and expeditiously remove or disable access to the targeted content; easily accessible and user-friendly mechanisms for users to report or flag to the hosting service provider alleged terrorist content; or mechanisms to increase the awareness of terrorist content on its services, such as mechanisms for user moderation.[108] It is worth highlighting that the Commission originally proposed the introduction of «proactive measures», according to which hosting service providers would have been obliged to apply measures to protect their services against the dissemination of terrorist content.[109] These measures were widely criticised since they appeared to enable the generalised monitoring of online content and incentivised the automaticity in the assessment of such content.[110] However, as persuasively argued, how the finally adopted «specific measures» effectively differ from the «proactive measures» is hard to tell, considering the decision of what type of measures should be adopted remains to the providers.[111] Furthermore, in case of non-compliance, Member States should adopt rules on penalties, which can be of an administrative or criminal nature, as well as, where appropriate, fining guidelines.[112]

In terms of the rights safeguards of content providers whose content has been removed, it is relevant to mention Article 7 which includes a set of transparency obligations on service providers that shall clearly provide in their terms and conditions their policy for addressing the dissemination of terrorist content, «including, where appropriate, a meaningful explanation of the functioning of specific measures, including, where applicable, the use of automated tools».[113] Article 9 establishes the right to an effective remedy, encompassing the right to challenge a removal order issued pursuant to Article 3 before the courts of the Member State of the competent authority that issued the removal order. Article 10 on complaint mechanisms is

[107] Recital 18 of the Regulation.

[108] Article 5 of the Regulation.

[109] Commission Proposal for a Regulation on preventing the dissemination of terrorist content online, COM(2018) 640 final, Art 6(1).

[110] V. MITSILEGAS, *The Privatisation*, cit..

[111] L. DUTKIEWICZ KRACK NOÉMIE, *How to Notice without Looking: The "algorithmization" of Terrorist Content Moderation in the Proposal for a Regulation on Preventing the Dissemination of Terrorist Content Online [Part II]* in *CITIP blog*, 2020 <https://www.law.kuleuven.be/citip/blog/how-to-notice-without-looking-the-algorithmization-of-terrorist-content-moderation-in-the-proposal-for-a-regulation-on-preventing-the-dissemination-of-terrorist-content-online-part-ii/> accessed 16 January 2023.

[112] Recital 42 and Art. 18 of the Regulation.

[113] Art. 7 of the Regulation.

also worth a mention since it requires each hosting service provider to «establish an effective and accessible mechanism allowing content providers where their content has been removed or access thereto has been disabled as a result of specific measures pursuant to Article 5 to submit a complaint concerning that removal or disabling, requesting the reinstatement of the content or of access thereto».[114]

After the entry into force of the Regulation, some of the concerns raised during the negotiations remain. First of all, as mentioned even if the Regulation does not oblige the use of automated tools, the short time limit from the receipt of the removal order and the threats of fines are likely to incentivise these types of measure.[115]

Furthermore, while the Regulation specifies that it does not *impose* a general obligation for the hosting service providers to conduct general monitoring activities and actively seek illegal activities, the choice of specific measures to prevent the spread of terrorist content online is problematically left to the discretion of the providers.[116]

The e-Commerce Directive (art. 15) already provided the prohibition of general monitoring obligations, such prohibition is now expressly preserved by the DSA under art. 8. In few words, the DSA is in continuation with the rules on monitoring obligations already developed under the e-Commerce Directive in order to guarantee a fair balance of fundamental rights in the online world.

Eventually, with regard to judicial remedies and effective oversight, it should be highlighted that the Member States are entitled to designate their national competent authorities, stating that such competent authorities must carry out their duties in an objective and non-discriminatory manner, however, the «issue of judicial review of the decisions on the removal and thus of effective oversight has been left unanswered; neither the proposal nor the final text includes anything in that regard».[117] This undermines due process and the rule of law and raises depoliticization concerns,[118] considering «a very large degree of discretion is still left to companies with regard to the level of independence, accessibility, transparency, and predictability of their in-house appeal mechanisms in the context of internal oversight mechanisms and review procedures».[119] The compatibility of the Regulation with the EU Charter of Fundamental Rights and national fundamental laws will be put to the test.

It is worthy to highlight that surprisingly French Constitutional Council rejected a challenge against some provisions of a law on terrorist content online, declaring the constitutionality of art. 3 of the Regulation on removal orders. [120] In fact the provisions of the so-called 'Avia law', concerning the obligation imposed upon online service providers to remove content flagged by users as «manifestly illegal» (terrorist content and hate speech), have already been declared unconstitutional by the French Constitutional Council. The latter *inter alia* held that several

[114] Article 10 of the Regulation.

[115] *CEPS Task Force Report*, cit., p. 94.

[116] Ibidem.

[117] Ibidem.

[118] R. GORWA, R. BINNS and C. KATZENBACH, cit, p. 12.

[119] Ibidem p. 7.

[120] Conseil Constitutionelle Decision n. 2022-841 Dc Du 13 Aout 2022 available at: <https://www.conseil-constitutionnel.fr/decision/2022/2022841DC.htm> accessed 29 December 2022.

provisions of the law restricted the exercise of the freedom of expression in a manner that is not necessary, appropriate and proportionate.[121]
The question of legitimacy specifically concerned the possibility for an administrative authority to determine the terrorist nature of the content on which basis it can order the removal of such content within one hour, under threat of even criminal penalties and without providing for a suspensive remedy, nor any other guarantee compensating for the absence of an *ex-ante* judicial assessment, without breaching the freedom of expression and information.[122] According to the *Conseil Constitutionnel*, the provisions of the Regulation, and in particular its articles 9, 12 and 18, only require the Member States of the European Union to designate a competent authority to issue a removal order under Article 3 of the same regulation, to provide an effective remedy allowing providers of hosting services to challenge such order before the courts of the Member State of the authority which issued it, and to determine the system of penalties applicable in the event of default. According to the *Counseil*, the determination of the terrorist nature of the content in question is not left to the sole discretion of the administrative authority designated by the contested provisions to issue removal orders. Firstly, on the one hand, the removal order likely to be issued by the competent administrative authority can only concern content of a terrorist nature precisely defined and exhaustively listed in Article 2 of the Regulation. On the other hand, Article 3 provides that the removal order issued by the competent administrative authority must include not only a reference to the type of content concerned, but also a sufficiently detailed statement explaining the reasons for which it is considered to be of a terrorist nature.[123] In addition, the competent authority mentioned in Article 6, designated within it by the Regulatory Authority for Audiovisual and Digital Communication, which is an independent administrative authority, must be informed of these requests for withdrawal.[124] Thus the *Counseil* concludes that the contested provisions do not violate freedom of expression and information and are in conformity with the French Constitution.

5. Normalisation of exceptionalism?

As anticipated terrorist speech is part of a broader challenge of how moderating illegal content and societal risks against other harmful online activities as well as what societal and legal responsibility should the private platforms have in that respect. The EU regulatory approach is progressively going towards the "normalisation" of the exceptional legal treatment reserved to targeted online content, provided originally as a matter of urgency in order to promptly respond to extremely worrying phenomena such as terrorism. In this respect, the recently adopted Digital Service Act (DSA), which amends the 2000 e-commerce Directive, regulates the obligations of digital services in general, aiming at creating a safe, predictable and trustworthy

[121] P. BREYER, *French Law on Illegal Content Online Ruled Unconstitutional: Lessons for the EU to Learn* , 2020 <https://www.patrick-breyer.de/en/french-law-on-illegal-content-online-ruled-unconstitutional-lessons-for-the-eu-to-learn/> accessed 2 January 2023.
[122] Conseil Constitutionelle, Decision n. 2022-841, cit.
[123] Ibidem par. 13 and par. 14 in relation to Art. 3 of the Regulation.
[124] Ibidem par. 15.

online environment.[125] The DSA focuses on issues of liability for illegal content and responsibility in content moderation processes, laying down due diligence obligations that will apply to all digital services that connect consumers to goods, services, or content depending on their roles and size, within the EU, even if established outside the EU. Specific due diligence are provided for hosting services; more extensive rules apply instead to very large online platforms (also very large online search machines) because of their significant societal and economic impact. The DSA signs an important development in the EU policy, finally departing from the self-regulatory paradigm for online service responsibilities, by defining clear and proportionate obligations for online services.[126] The purpose of the DSA is in fact to fill a regulatory gap on systemic issues which were not appropriately addressed on a horizontal level, accompanying sector-specific legislations. This means the DSA provisions complement the TERREG provisions, which still apply as *lex specialis*.[127] To sum up the main novelties, it should be mentioned a new tiered «system of due diligence obligations for (very large) intermediary services, the regulation of content moderation through terms of service enforcement, systemic risk assessment obligations for the most widely used platforms, and access to data for researchers».[128]

A central aspect of these obligations concerns new standardised procedures for the detection, flagging and rapid removal of illegal content.[129] However, the DSA does not impose a general obligation for service providers to monitor the information transmitted or stored or to actively seek facts or circumstances indicating illegal activity.

The concept of illegal content though is not defined at the EU level, it should broadly reflect the existing rules in the offline environment.[130] The DSA defines illegal content by *reference* to any information not in compliance with either EU law or the law of *any* Member State, irrespective of the precise subject matter or nature of that law.[131] It may refer to information as considered in itself (for example terrorist content, child sexual abuse material, hate speech and unlawful discriminatory content) or in relation to an illegal activity (unlawful non-consensual sharing of private images, online stalking, the sale of non-compliant or counterfeit products, the non- authorised use of copyright protected material). In this sense, «the rather broad definition of illegal content entails that the DSA does not impose any limits as to what content can be criminalised at the national level and the clarifications do not help in this regard».[132] The general rule remains that providers of hosting services are not liable for user-generated content.

---

[125] EU Regulation 2022/2065 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)) OJ L 277/1 of 19.10.2022 (DSA).

[126] I. BURI, *The DSA Proposal Impact on Digital Dominance*, in *Verfassungsblog*, 2022, <https://verfassungsblog.de/power-dsa-dma-01/ >; Id, *Can the EU Digital Services Act Contest the Power of Big Tech's Algorithms?* in *European Digital Rights (EDRi)* <https://edri.org/our-work/can-the-eu-digital-services-act-contest-the-power-of-big-techs-algorithms/> accessed 30 December 2022.

[127] Recital 9 and recital 10 of DSA; Proposal for a Regulation on a Single Market for Digital Service (DSA), Explanatory Memorandum, 4.

[128] I. BURI, cit.

[129] See Chapter 2 and Chapter 3 of the DSA; *CEPS Task Force Report*, cit., p. 52.

[130] Recital 12 of the DSA.

[131] *CEPS Task Force Report*, cit., p. 88.

[132] Ibidem.

However, in order to benefit from the exemption from liability for hosting services, the provider should, upon obtaining actual knowledge or awareness of illegal activities or illegal content, act expeditiously to remove or to disable access to that content (similarly to what yet provided in the e-commerce Directive). Recital 22 specifies the ways through which the provider can obtain the actual knowledge or awareness of the illegal nature of the content, including by its own-initiative investigations as well as via sufficiently precise notices submitted to it by individuals or entities. In other words, providers can be excluded from liability because they, in good faith and in a diligent manner, carry out voluntary own-initiative investigations or take other measures aimed at detecting, identifying, or disabling of access to illegal content, or take measures to comply with EU or national law.

Intermediary services shall enforce orders to act against one or more specific items of illegal content issued by national judicial or administrative authorities enabled by national laws.[133] The same obligation applies to orders to provide specific information about one or more specific individual recipients of the service.[134] The providers must inform the issuing authority of the removal order, which can be both a judicial or an administrative one, without undue delay that such order was applied. Furthermore, platforms must cooperate with the so called «trusted flaggers», meaning organisations that have demonstrated particular expertise and competence to identify and remove illegal content.[135] Very large online platforms are obliged to take mitigating measures at general level, to protect their users from illegal content, goods and services on the basis of risk assessment mechanisms.[136]

The new rules require the removal of online content regard illegal but not harmful material. As explained in the Proposal, «there is a general agreement among stakeholders that 'harmful' (yet not, or at least not necessarily, illegal) content should not be defined in the Digital Services Act and should not be subject to removal obligations, as this is a delicate area with severe implications for the protection of freedom of expression».[137] However, very large online platforms' and very large online search engines are obliged to deal with «systemic risks» such as disinformation, harms to vulnerable groups and other emerging societal harms.[138] In particular, they are obliged to perform annual risk assessments and take corresponding risk mitigation measures. Furthermore, the DSA sets out a co-regulatory framework where service providers can address online harms, manipulative and abusive activities through codes of conduct such as a revised Code of Practice on disinformation, and crisis protocols.[139] According to the DSA, these measures should be carefully balanced against restrictions of freedom of expression.[140]

Furthermore, in terms of users protection and effective remedies, recital 39 stresses the necessity to respect fundamental right to an effective judicial remedy and to a fair trial as

[133] Art. 9 of the DSA.
[134] Art. 10 of the DSA.
[135] Art. 22 of the DSA.
[136] Artt. 35 ff. of the DSA.
[137] *Proposal for a Regulation on a Single Market for Digital Service (DSA) and amending Directive 2000/31/EC, COM (2020) 825 final*, p. 9.
[138] Art. 34 of the DSA.
[139] Recital 104 of the DSA.
[140] Recital 86 of the DSA.

provided for in Article 47 of the Charter. In particular, it is required «to provide information on redress mechanisms available to the provider of the intermediary service and to the recipient of the service who provided the content include a requirement to provide information about administrative complaint-handling mechanisms and judicial redress including appeals against orders issued by judicial authorities»[141].

Article 14 on «terms and conditions» requires providers to include «information on any restrictions that they impose in relation to the use of their service in respect of information provided by the recipients of the service, in their terms and conditions».[142] This means that they should inform in an easily and accessible language users on «any policies, procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making and human review, as well as the rules of procedure of their internal complaint handling system»[143]. Furthermore, «Providers of intermediary services shall inform the recipients of the service of any significant change to the terms and conditions».[144] Article 15 includes a list of further transparency reporting obligations for providers, i.e. «any use made of automated means for the purpose of content moderation, including a qualitative description, a specification of the precise purposes, indicators of the accuracy and the possible rate of error of the automated means used in fulfilling those purposes, and any safeguards applied».[145]

The DSA provides new mechanisms that allow users to flag illegal content[146] and access to an effective internal complaint-handling system[147] and if the providers become aware or suspect that a criminal offence has been committed, they have to promptly inform the judicial authorities or law enforcement of the Member State concerned.[148]

Article 20 requires providers of online platforms to provide users, for a period of at least six months, the access to an effective internal complaint-handling system. In addition, Article 21 foresees the possibility for on out-of-court dispute settlement mechanism. In particular, users must be entitled to select any certified out-of-court dispute settlement body in order to resolve disputes relating to the decisions taken by the providers of online platforms. Users can still decide to contest those decisions before a court in accordance with the applicable law. In general, the providers are supervised by the Commission, which thus assumes a key role. Member States are instead required to set up Digital Service Coordinators, which have to investigate and in case of infringements, inform the Commission.[149] An independent group composed of Digital Services Coordinators and chaired by the Commission, named the European Board for Digital Services, is also established. In case of non-compliance and infringements of the DSA large fines are also provided.

It is worth highlighting here that during the negotiations the European Parliament attempted to introduce explicit rules for users on effective remedies against orders, including restoration of

[141] Recital 39 of the DSA.
[142] Art. 14 (1) of the DSA.
[143] Art.14 (1) of the DSA.
[144] Art. 14 (2) of the DSA.
[145] Art. 15 (1) lett.c) of the DSA.
[146] Art. 16 of the DSA.
[147] Art. 20 of the DSA.
[148] Art. 18 of the DSA.
[149] Art. 49 of the DSA.

the content that has been erroneously considered as illegal. However, this is as yet undeveloped, and «reference to the possibility of restoration has only remained in the Preamble».[150]
The DSA certainly represents an appreciated effort to regulate and redefine online platforms policies, signifying a departure from the previous discourse on content governance characterised by opaque actions by private actors to make decisions with significant implications for human rights and democratic discourse. However, it remains debatable whether the solution of enforcing human rights responsabilities on private intermediaries can exempt state state authority from its duty to prevent infringements of fundamental rights.[151]
The debate at EU level remains alive, even within specific sectors, targeting other kinds of content. An interesting initiative in this sense is the 2022 released proposal for a Directive on combating violence against women and domestic violence,[152] recently approved by the Parliament on the 24th of April 2024.[153] The Directive, which found its legal basis on Articles 82(2) and 83(1) TFEU, aims at preventing and combat violence against women and domestic violence to ensure a high level of security and the full enjoyment of fundamental rights within the Union, taking into account also more recent area of online abuse due to current digital transformation and the increase of cyber violence.[154] The Commission highlights the connection between the usage of internet and social media, tools of fast and broad sharing of hate speech reinforced by the online disinhibition effect, and the sharp rise in incitement to violence and hatred.[155] In this respect, «women are often the target of sexist and misogynous hate online, which can escalate into hate crime offline». [156] According to the Commission, this should be caught at an early stage. While the language used in this type of incitement does not always directly refer to the sex or gender of the targeted person(s), «the biased motivation can be inferred from the overall content or context of the incitement».[157] The draft provides minimum rules on the definitions of offences concerning certain forms of violence against women or domestic violence offline and online. These latter, that expressily amount to computer crime, include: non-consensual sharing of intimate or manipulated material (Article 5), offences concerning cyber stalking (Article 6), offences concerning cyber harassment (Article 7), and cyber incitement to hatred or violence (Article 8).[158] In addition, the legislation provides the removal of online content in relation to such offences of cyber violence, and a possibility of judicial redress for affected users. The draft requires Member State to take the necessary measures to ensure the prompt removal of the material referred to the computer crimes offences provided in the text, without prejudice to the DSA and within the limits set

[150] *CEPS Task Force Report*, cit., p. 55.
[151] M. SENFTLEBEN, *Human Rights Outsourcing and Reliance on User Activism in the DSA* in (21 February 2024)
[152] *Proposal for a Directive on combating violence against women and domestic violence COM (2022)* 105 final, 8.3.2022.
[153] See the amendemnts by the EU Parliament to the Commission proposal (A9-0234/2023), 16.4.2024. The adopted text is available on the following link: https://www.europarl.europa.eu/news/en/press-room/20240419IPR20588/parliament-approves-first-ever-eu-rules-on-combating-violence-against-women#:~:text=The%20directive%20calls%20for%20stronger,of%20private%20information%20and%20cyberflashing.
[154] Ibidem, Explanatory Memorandum, cit. pp. 2-3.
[155] Recital 22 of Proposal for a Directive on combating violence.
[156] Ibidem.
[157] Ibidem.
[158] See the amendemnts by the EU Parliament to the Commission proposal (A9-0234/2023), cit. *supra*.

therein regarding orders to remove illegal online content. In particular, amongst these measures, Member States shall include the possibility for their competent judicial authorities to issue binding legal orders to remove or disable access to such material addressed to relevant providers of intermediary services, upon application by the individual affected. Member States shall ensure that the end-users of the relevant services are informed, where appropriate by the intermediary service providers concerned, of the reasons for the removal of or disabling access to the material pursuant to the orders or other measures provided and that those end-users have access to judicial redress.[159]

According to the Commission, «despite cyber violence's wide prevalence, regulation has been highly fragmented to date, and significant legal gaps have been identified at both EU and Member State level».[160] The adopted text specifies the prohibition of imposing general obligations of monitoring or active fact-finding as provided in the DSA. The approach is similar to the one taken in the Directive regarding terrorist content: we have a criminal law legal basis (Articles 82(2) and 83(1) TFEU), under which hybrid-kind measures for the removal of the material related to cyber violence offences against women are provided for Member States. As explained, this differs from the Regulation, which uses as a legal basis Article 114 TFEU despite the security purposes, and it is directed to providers.

6. Conclusion

This contribution has identified the dimensions of digital exceptionalism justified by the need to adapt regulations and policies to the threat of terrorism posed by the digital domain. It has further analysed how the EU regulatory approach to online targeting content has departed from the criminalisation of 'illegal' speech in the offline environment, at the expense of freedom of expression and the rule of law in its substantive and formal components.[161]

The over-criminalisation of online speech, accompanied by vague and broad definitions of terrorism and terrorist-related offences, has clear consequences in terms of accessibility and foreseeability, undermining the principle of legality and proportionality.[162] While already in the offline environment the boundaries of criminal liability have become porous, based – as argued – on the predominantly mental element over the *actus reus*, the regulation of online terrorist speech has increased these issues. This is due to the difficulties in determining intent beyond an internet user, especially through automated algorithmic tools, with intrinsic risks of bias and discrimination. But even where human oversight is involved it is questionable whether private companies are equipped to assess the intentions of content-generators.[163] The lines between criminal law and prevention have also become blurred thanks to the introduction of additional measures to detect material that it is illegal for the (subjective) intent behind its publication,

[159] Art. 25 (4) (5) of the Proposal for a Directive on combating violence.
[160] Explanatory Memorandum, cit., p. 3.
[161] V. MITSILEGAS, *Rule of Law*, cit..
[162] See the copious ECtHR case law on art. 7 ECHR (Liivik v. Estonia, parr. 96-104 (2009); Cantoni v. France, par. 35 (1996); Contrada v. Italy n. 3 (2015).
[163] G. ROBINSON, cit., p. 189.

coupled with risk that some recipients will be incited to terrorist related offences.[164] In this sense, the nature of the online removal measures provided in the TERREG remains controversial. The fact that an administrative authority, during investigations or even outside criminal proceedings, may have the power to determine the terrorist nature of the content on which basis it can order the removal of such content within one hour, without any guarantee compensating for the absence of prior intervention by a judge, is particularly worrying. The decision can be made on the basis of mere suspicion, even where offline elements of a crime cannot be proven. As noted, «there is a concrete risk over censorship, and controversial thoughts, and infringing on freedom of expression if there is no court assessment of risk or intent».[165]

Secondly, the prominent role of internet providers in policing content online and the imposition by the state of duties to private actors to essentially undertake fundamental rights compliance assessments raise deep concerns. The TERREG has signified a further step in the process of the privatisation of surveillance by attributing to service providers far-reaching responsibilities for the prevention of terrorism.[166] As highlighted, «providers are not only required to comply with state requests swiftly in a reactive manner, but are also asked to exercise their judgment in order to proactively remove content that they consider to be related to terrorism».[167] Fundamental rights concerns are aggravated by the fact that action and the exercise of decision-making by private providers can be based on automaticity and speed and reliance on AI tools to take decisions which impact significantly on individuals.[168] Under internal and external pressures, private platforms have enforced their online moderation activities outside formal legal proceedings, without any judicial oversight and often an accessible and effective supervision, even when in-house appeal mechanisms are established.[169] These aspects though are of crucial importance in delivering access to justice and remedies sufficient to ensure effective legal protection, central for the respect of the rule of law.[170] It remains open whether the horizontal provisions of the DSA offer sufficient and effective remedies, access to complaint and redress mechanisms.[171] Otherwise, consequences may be serious: «the potential of general monitoring of the online environment by providers, coupled with the imposition of fines for non-compliance, may result in creating incentives for over-removal of content»,[172] without an effective judicial supervision and enforcement.

The rule of law is also undermined by the legal uncertainty on what content should be removed and what is a disallowed speech, exacerbated by the fact that there is no exact overlap between offline and online. As said, the EU law, TERREG specifically, provides definitions that tend to be over inclusive, jeopardising legal certainty and disproportionately affecting individual rights

[164] H. SARGEANT and others, *Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual*, Organization for security and cooperation in Europe (OCSE) publication, 2022.
[165] FRA *Directive (EU) 2017/541*, cit. p. 61.
[166] V. MITSILEGAS, *The Privatisation*, cit..
[167] Ibidem.
[168] Ibidem.
[169] *CEPS Task Force Report*, cit., p. 94.
[170] Ibidem.
[171] I. BURI, cit.
[172] *CEPS Task Force Report*, cit., p. 93.

and freedoms, primarily freedom of expression. Even now that the Regulation adopted aligns the definitions of terrorism and terrorist offences with those adopted in the Terrorism Directive - which are already problematic -, further guidance remains necessary for the involved actors. This is key considering that the definition adopted significantly broadens the scope of the obligation to online speech, expanding from public provocation to other material such as training for terrorist purpose. Whether the notion of illegal content should be fully aligned to criminal law remains unsolved in the horizontals provision of the DSA, which broadly refers to what is not in compliance with either EU law, or the law of *any* Member State, irrespective of the precise subject matter or nature of that law and thus whether that content has been expressly criminalised at the national or EU level.

It also remains debated what exactly constitutes harmful (but not illegal) online and who shall draw such boundaries.[173] So far, big companies have taken the lead in the definition and moderation of such content on their services, prohibiting and disallowing speech on the basis of their own guideline and terms of use. The DSA only partially clarify these issues by attributing regulatory oversight, accountability and transparency to online service providers, relying mainly on co-regulation and code of conducts, but it does not provide a definition of harmful content in the text. Questions are multiple and complex, but answers are necessary to enhance legal certainty and freedom of expression online. The lack of regulatory clarity in relation to what is illegal, unlawful or harmful content, especially ambiguous categories such as «online harm», «extremist content», «disinformation», increases the risk that platform services policies governing content moderation over-restrict fundamental rights, democracy and the rule of law.[174]

[173] *CEPS Task Force Report*, cit., pp. 14; 49; 97.
[174] Ibidem p. 97.