Regulating the moderation of illegal online content

De Streel, Alexandre; Ledger, Michele

Published in:
Unravelling the Digital Services Act package

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

3. Regulating the moderation of illegal online content

Alexandre de Streel, professor of EU law at Namur University and the Namur Digital Institute (NADI), academic co-director at the Centre on Regulation in Europe (CERRE) and chair of the expert group for the EU Observatory on the online platform economy.

Michèle Ledger, head of practice at Cullen International and senior researcher at NADI.[*]

3.1. Scope and structure of this chapter

This chapter studies the EU regulatory framework applicable to hosting intermediaries when they moderate online content which is illegal or in breach of their terms and conditions.[24] Each of those concepts are already – or are about to be – defined in EU law: (i) **Hosting intermediaries** comprise all organisations which store information provided by, and at the request of, a recipient of the services;[25] (ii) **Content moderation practices** cover all the measures that intermediaires take to manage content which is in violation of the law or of their terms and conditions as well as to manage their users (e.g. the suspension or termination of the user's account);[26] (iii) **Illegal content** comprises any information which does not comply with EU or national law, irrespective of the precise subject matter or nature of that law.[27]

On this last concept, it is important to distinguish between: content (i) which violates EU or member state law and hence is illegal according to the proposed DSA definition; (ii) which does not violate a law but violates the terms and conditions of a

[*] The authors wish to thank Maja Cappello and Francisco Cabrera for their very useful comments and discussions; as always, responsibility for the content of this article is the authors' alone.

[24] This chapter is partly based on de Streel A. et al., "Online platforms' moderation of illegal content online", study for the European Parliament, 2020. On the rules on content moderation, see Floridi L. and Taddeo M. (eds), *The responsibility of online service providers,* Springer, 2017 and Frosio G. (ed), *The Oxford handbook of online intermediary liability,* Oxford University Press, 2020.

[25] Directive 2000/31 of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the internal market [2000] OJ L 178/1, art.14; https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32000L0031, Proposal of the Commission of 15 December 2020 for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31, COM(2020) 825, Art. 5(1), https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN.

[26] DSA Proposal, Art. 2(p) defines content moderation as "the activities undertaken by providers of intermediary services aimed at detecting, identifying and addressing illegal content or information incompatible with their terms and conditions, provided by recipients of the service, including measures taken that affect the availability, visibility and accessibility of that illegal content or that information, such as demotion, disabling of access to, or removal thereof, or the recipients' ability to provide that information, such as the termination or suspension of a recipient's account', https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN.

[27] DSA Proposal, Art. 2(g), https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN.

platform where it is posted; and (iii) which violates neither a law nor the platform's terms and condition but creates harm to users, especially to the most vulnerable ones (such as minors). Our paper focuses mostly on the EU rules applicable to the moderation of the first category of online content, which is the most heavily regulated and only touches, when needed, on the second and third category of content.[28]

The chapter follows the evolution over the years of the EU regulatory framework as the Internet has increased in importance for the economy and society. At the turn of the century when digital intermediaries were in their infancy, the Internet remained relatively free from state intervention as famously suggested by John Perry Barlow in his 1996 Declaration of Independence of Cyberspace[29] (section 5.2). New rules started to be adopted to cater for particular types of illegal content or particular types of digital intermediaries, marking the beginning of the end of digital exceptionalism (section 5.3). Now, new horizontal rules applicable for all platforms and all content are in the making, indicating the end of cyberspace independence (section 5.4). Although these rules are certainly a step in the right direction, some clarifications and improvements are possible (section 5.5).

3.2. The independence of cyberspace: the e-Commerce Directive

In 2000, the e-Commerce Directive established a special liability regime for online intermediary services. As explained by the European Commission,[30] this regime pursued four main objectives: (i) to share responsibility for a safe Internet between all the private actors involved and to promote good cooperation with public authorities – thus, injured parties should notify online platforms about any illegality they observe and online platforms should remove or block access to any illegal material of which they are aware; (ii) to encourage the development of e-Commerce in Europe by ensuring that online platforms do not have an obligation to monitor the legality of all material they store; (iii) to strike a fair balance between the fundamental rights of the several stakeholders, in particular privacy and freedom of expression, freedom to conduct business (for platforms) and the right to property including intellectual property of injured parties;[31] and (iv) to strengthen the digital single market by adopting a common EU standard on liability exemptions, especially at a time when national rules and case law were increasingly divergent.

Thus, the e-Commerce Directive creates an exemption from the national liability regime to which the hosting platform is subject and determines the requirements to be met

[28] On online disinformation which is often of the third category, see Chapter 6 of this publication.
[29] https://www.eff.org/fr/cyberspace-independence.
[30] Explanatory Memorandum of the Commission of 18 November 1998 for the proposal for a directive on certain legal aspects of electronic commerce in the internal market, COM(1998)586, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:51999PC0427&rid=3.
[31] As protected by the Charter of Fundamental Rights of the European Union, Articles 7, 8, 11, 16 and 17, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT.

by the providers to benefit from such an exemption.[32] A hosting platform can escape liability for illegal material uploaded by users when it "does not have actual knowledge of illegal activity or information and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent". Should the platform have such knowledge or awareness, it can however benefit from the liability exemption if it "acts expeditiously to remove or to disable access to the information". Liability exemptions are horizontal: all types of illegal content or activities are covered (unfair market practices, violation of data protection rules, damage to honour and reputation, etc.), as well as various kinds of liabilities (criminal or civil).[33]

To benefit from the liability exemptions, the hosting platform should also be neutral in the sense that its conduct is, in the words of the Court of Justice, "merely technical, automatic and passive, pointing to a lack of knowledge or control of the data which it stores".[34] A related issue is whether the e-Commerce Directive disincentivises the online platforms to proactively monitor the legality of the material they host because, if they were to do so, they might lose the benefit of the liability exemption. This is sometimes referred to as the good Samaritan paradox. For instance, a platform carrying out *ex ante* moderation practices could be considered as playing an active role and, therefore, be excluded from the liability exemption. During the public consultations organised by the European Commission on the e-Commerce Directive, online platforms mentioned this legal risk of voluntarily introducing more proactive measures.[35] However, in its Communication of September 2017 on tackling illegal online content, the European Commission considered that voluntary proactive measures "do not in and of themselves lead to a loss of the liability exemption, in particular, the taking of such measures need not imply that the online

[32] e-Commerce Directive, Art. 14.On the liability exemption, see Chapter 2 of this publication. Also, Kuczerawy A., *Intermediary liability and freedom of expression in the EU: From concepts to safeguards*, Intersentia, 2018, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32000L0031.

[33] Note that, even when a digital intermediary cannot benefit from the liability exemption, it would not necessarily be considered liable under the applicable legal framework. In this case, the national jurisdiction should determine whether legal requirements applicable in the member state are fulfilled (e.g. negligence under civil law) and, if so, decide that the intermediary should be held liable.

[34] Cases C-236/08 to C-238/08 *Google France v Louis Vuitton,* EU:C:2010:159, https://curia.europa.eu/juris/liste.jsf?num=C-236/08; https://curia.europa.eu/juris/liste.jsf?language=en&num=C-238/08, and Case C-324/09 *L'Oreal and Others v. eBay and Others* EU:C:2011:474, https://curia.europa.eu/juris/liste.jsf?num=C-324/09. These cases are well explained in van Hoboken J., Quintais J.P., Poort J. and van Eijk N., "Hosting Intermediary Services and Illegal Content Online", study for the European Commission, 2018, https://op.europa.eu/en/publication-detail/-/publication/7779caca-2537-11e9-8d04-01aa75ed71a1/language-en

[35] For the 2011 public consultation: Commission Staff Working Document of 11 January 2012, Online services, including e-Commerce, in the Single Market, SEC(2011) 1641, p.35, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52011SC1641. For the 2015-2016 consultation, Communication from the Commission of 25 May 2016, Online Platforms and the Digital Single Market Opportunities and Challenges for Europe, COM(2016) 288, p. 9, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016DC0288 and Commission Staff Working Document of 10 May 2017 on the Mid-Term Review on the implementation of the Digital Single Market Strategy, SWD(2017) 155, p. 28, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017SC0155&rid=1.

platform concerned plays an active role which would no longer allow it to benefit from that exemption".[36]

Another pillar of the e-Commerce Directive consists in the prohibition, for EU member states, on imposing a general obligation on the hosting platforms to monitor the material hosted.[37] The Court of Justice has drawn a blurred line between general monitoring measures and specific monitoring measures, in particular in case of suspected violations of intellectual property rights. The first are prohibited;[38] the second are allowed when a fair balance between the fundamental rights of the different stakeholders is achieved.[39] Although imposing a general obligation to monitor is not allowed, online platforms could decide, on a voluntary basis, to carry out spot checks on the online content. This is not prohibited but by doing this, the online platform could be considered as playing an active role as explained above.

In addition, member states may impose on hosting providers the duty to cooperate with the competent authorities.[40] Two types of duties are possible: spontaneous communication to the authorities or communication at their request. Information related to identification of the user who posted illegal content anonymously could be communicated to the victim of the illegal content (so they may bring a claim against the author) or only to the competent authorities.

The last pillar of the e-Commerce Directive is the encouragement of co- and self-regulation in implementation of the rules and principles of the Directive.[41] In particular, the Directive mentions the importance of involving consumers in drafting codes of conduct to ensure that the rules remain balanced. To ensure the effectiveness of those rules, monitoring implementation of the codes is essential.[42] This provision has led, as explained in the next section, to increasing reliance on co- and self-regulation to tackle certain types of illegal materials which have a very negative impact on society, such as hate speech, child abuse content or terrorist content.

---

[36] Communication of the Commission of 28 September 2017, Tackling Illegal Content Online. Towards an enhanced responsibility for online platforms, COM (2017) 555, p.13, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52017DC0555.

[37] e-Commerce Directive, Art. 15(1). On this, see Husovec M., *Injunctions against intermediaries in the European Union: Accountable but not liable?*, Cambridge University Press, 2017

[38] Case C-360/10 *SABAM v. Netlog* EU:C:2012:85; https://curia.europa.eu/juris/documents.jsf?num=C-360/10, Case C-70/10 *Scarlet Extended v. SABAM* EU:C:2011:771 ,https://curia.europa.eu/juris/liste.jsf?language=en&num=C-70/10; Case C-18/18, *Glawischnig-Piesczek v. Facebook Ireland* EU:C:2019:821, https://curia.europa.eu/juris/liste.jsf?num=C-18/18.

[39] Case C-314/12 *UPC Telekabel Wien v Constantin Film Verleih GmbH* EU:C:2014:192, https://curia.europa.eu/juris/liste.jsf?num=C-314/12; Case C-484/14 *Mc Fadden*, para 96, https://curia.europa.eu/juris/liste.jsf?num=C-484/14.

[40] e-Commerce Directive, Art. 15(2), https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32000L0031.

[41] e-Commerce Directive, Art. 16, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32000L0031.

[42] In that regard, the Commission has developed some principles for better self- and co-regulation. These principles relate to the conception of the rules: they should be prepared openly and by as many relevant actors as possible; they should set clear targets and indicators and be designed in compliance with EU and national law. The principles also relate to the implementation of the rules: they should be monitored in a way that is sufficiently open and autonomous, improved in an iterative manner (learning by doing) and non-compliance should be subject to a graduated scale of sanctions.

3.3. The beginning of the end: The emerging EU regulatory framework for online content moderation

As the Internet became increasingly important in the economy and influential in society, the EU started to take back control of cyberspace and adopted new rules for content moderation, first focussing on the most harmful illegal content[43] and then on some specific types of digital intermediaries.

3.3.1. Regulation of the moderation of specific types of online content

3.3.1.1. Racist and xenophobic hate speech

Already back in 2008, the EU adopted a Counter-Racism Framework Decision which seeks to combat particularly serious forms of hate speech and provides that member states must ensure that racism and xenophobia are sanctioned by criminal law.[44] However, this Decision does not provide for detailed obligations related to online content moderation practices and more generally, the fragmentation of criminal procedural rules across member states makes it difficult to enforce the Decision effectively.[45]

Therefore, in 2016 at the initiative of the Commission, the main online platforms agreed on an EU Code of Conduct on countering all forms of illegal hate speech online[46] with a series of commitments: (i) drawing users' attention to the types of content not allowed by their community standards/guidelines and specifying that they prohibit the promotion of incitement to violence and hateful behaviour; (ii) putting in place a clear and effective process to review reports/notifications of illegal hate speech in order to remove them or make them inaccessible; reviewing notifications on the basis of the community standards / guidelines and national laws, and reviewing the majority of valid reports within 24 hours; (iii) regularly training online platform staff, particularly in relation to societal developments; (iv) encouraging the reporting of illegal hate speech by experts, including through partnerships with civil society organisations – so that they can potentially act as trusted reporters – and strengthening partnerships and collaboration with these

[43] This chapter does not deal with content and material violating IP law as this is covered in Chapter 4 of this publication.

[44] Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law, O.J. [2008] *L 328/55,* https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32008F0913.

[45] Report of the European Commission of 27 January 2014 on the implementation of Council Framework Decision 2008/913 on combating certain forms and expressions of racism and xenophobia by means of criminal law, COM(2014)27, https://op.europa.eu/en/publication-detail/-/publication/ea5a03d1-875e-11e3-9b7d-01aa75ed71a1.

[46] The Code is available at: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en.

organisations to support them; and (v) strengthening communication and cooperation between online platforms and national authorities, in particular with regard to procedures for submitting notifications; collaborating with other online platforms to improve and ensure the exchange of best practices between them.

While considered a step in the right direction, commentators have pointed towards the following weaknesses: risks of private censorship through the priority application of community standards / guidelines; lack of precision in determining the validity of a notification; absence of appeal mechanisms for users whose content has been withdrawn; absence of a requirement for illegal content to be reported to the competent national authorities when removed on the basis of the community standards / guidelines; and the observation that the 24-hour deadline could either make it impossible for online platforms to meet their commitments or could lead to over-blocking practices.[47]

3.3.1.2. Child sexual abuse material

In 2011, the EU adopted the **Child Sexual Abuse and Exploitation Directive** which requires member states to take content removal and blocking measures against websites containing or disseminating child sexual abuse material.[48] Such measures must be based on transparent procedures and provide adequate safeguards, in particular be necessary and proportionate, inform the users on the reasons for restriction and ensure the possibility of judicial redress.[49] In practice, member states have adopted two categories of measures: (i) notice-and-takedown measures with national hotlines to which Internet users can report child sexual abuse material that they find online[50]; and (ii) measures based on national criminal law such as general provisions that allow the seizure of material relevant to criminal proceedings (e.g. material used in the commission of an offence) or more specific provisions on the removal of child sexual abuse material.[51]

In parallel to the efforts made by member states, a series of self-regulatory initiatives were taken by digital intermediaries – often encouraged by the European

---

[47] Quintel T. and Ullrich C., "Self-regulation of fundamental rights? The EU Code of Conduct on Hate Speech, related initiatives and beyond" in Petkova B. and Ojanen T., *Fundamental rights protection online: The future regulation of intermediaries*, Edward Elgar, 2019.

[48] Directive 2011/92 of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography, OJ. [2011] L 335/1, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32011L0093**.** The Directive provides at Art. 2, for a broad definition of child sexual abuse material that includes real child pornography that visually depicts a child engaged in real or simulated sexually explicit conduct or virtual child pornography, i.e. computer-generated pornographic material involving children. In general on the EU strategy and rules to fight online child pornography, see Jenay P., "Combating child sexual abuse online", study for the European Parliament, 2015, https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2015)536481.

[49] Child Sexual Abuse and Exploitation Directive**,** Art. 25. Measures may consist in various types of public action, such as legislative, non-legislative, judicial or others.

[50] Moreover, INHOPE, a global umbrella organisation for the hotlines, encourages exchange of expertise, https://www.inhope.org/EN.

[51] Report from the Commission of 16 December 2016 assessing the implementation of the measures referred to in Article 25 of Directive 2011/93 on combating the sexual abuse and sexual exploitation of children and child pornography, COM(2016) 872, https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2016)872&lang=en.

Commission – to better protect minors and make the Internet a safer place for children.[52] In 2017, the Alliance to Better Protect Minors Online, a multi-stakeholder forum facilitated by the European Commission, was set up in order to address emerging risks that minors face online, such as illegal and harmful content (e.g. violent or sexually exploitative content), conduct (e.g. cyberbullying) and contact (e.g. sexual extortion).[53] It is composed of actors from the entire value chain (device manufacturers, telecom operators, media and online platforms used by children). Its action plan includes the provision of accessible and robust tools that are easy to use, the provision of feedback and notification, the promotion of content classification, and the strengthening of cooperation between the members of the alliance and other parties (such as child safety organisations, governments, education services and law enforcement) to enhance best practice sharing.[54] In an evaluation of this alliance, Ramboll indicates that many commitments are difficult to measure, hence their effectiveness is difficult to assess. It also notes that the effectiveness of the alliance is limited by low public awareness and limited internal knowledge sharing. It therefore recommends increasing public awareness in order to strengthen the external monitoring of the commitments and to incentivise the participants to meet them and to reinforce sharing of good practices between members.[55]

3.3.1.3. Terrorist content

Terrorist content was the last type of content to be regulated at the EU level but is now the most strictly regulated. In December 2015 after terrorist attacks in several member states, an EU Internet forum to counter terrorist content online was established among EU interior ministers, high-level representatives of major online platforms (such as Facebook, Google, Microsoft and Twitter), Europol, the EU Counter-Terrorism Coordinator and the European Parliament.[56] One of its goals was to address the misuse of the Internet by terrorist groups and to reduce accessibility to terrorist content online. The forum led to an efficient referral mechanism in particular with the EU Internet Referral Unit of Europol, a shared database with more than 200,000 hashes, which are unique digital fingerprints of terrorist videos and images removed from online platforms.

Then in 2017, the EU adopted the Counter-Terrorism Directive which requires member states to take removal and blocking measures against websites containing or

---

[52] A CEO Coalition to Make the Internet a Better place for Kids was set up in 2011, https://ec.europa.eu/digital-single-market/en/self-regulation-and-stakeholders-better-internet-kids, and the ICT Coalition for Children Online was set up in 2012, http://www.ictcoalition.eu.

[53] European Commission, Alliance to better protect minors online, https://ec.europa.eu/digital-single-market/en/alliance-better-protect-minors-online.

[54] The common action is complemented by individual company commitments with a specific timeline to better protect minors online, see: https://ec.europa.eu/digital-single-market/en/news/individual-company-statements-alliance-better-protect-minors-online.

[55] Ramboll, "Evaluation of the implementation of the Alliance to Better Protect Minors Online", study for the European Commission, 2018, https://op.europa.eu/en/publication-detail/-/publication/122e3bdd-237b-11e9-8d04-01aa75ed71a1/language-en.

[56] European Commission press release of 3 December 2015, https://ec.europa.eu/commission/presscorner/detail/en/IP_15_6243.

disseminating terrorist content.[57] These measures must follow transparent procedures and provide adequate safeguards, in particular to ensure that they are limited to what is necessary and proportionate and that users are informed of the reason for the measures. In practice, as with the Child Sexual Abuse and Exploitation Directive, member states have adopted two main types of measures:[58] (i) notice-and-takedown measures, which differ among the member states on several issues such as offences covered, time limits for removal and consequences of non-compliance; and (ii) criminal law measures allowing a prosecutor or a court to order companies to remove content or block content or a website, within a period of 24 or 48 hours.

Finally in 2021, the EU went one step further with the adoption of the Terrorism Content Regulation which imposes duties of care on hosting services providers.[59] In addition to transparency reporting obligations,[60] the main new obligations for these hosting service providers are to (i) remove terrorist content within one hour of receiving a valid removal order stemming from a national competent – not necessarily a judicial – authority;[61] to (ii) preserve for six months removed terrorist content and related data necessary for administrative or judicial review or complaint handling or the prevention, detection, investigation and prosecution of terrorist offences;[62] and to (iii) take specific measures (if they have been previously exposed to terrorist content) to address the dissemination of terrorist material on their services, including by deploying automated detection tools.[63] It is interesting to note that where automated tools are used, safeguards should be put in place in particular through human oversight and verification. Although the specific measures are not precisely defined, platforms must in any case ensure they are targeted and proportionate to the risks of exposure and their size, are applied by taking into account the rights and legitimate interests of users (in particular their fundamental rights) and are applied in a diligent and non-discriminatory manner.

[57] Directive 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism**,** OJ [2017] L 88/6**,** Article 21, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32017L0541.

[58] Commission Staff Working Document of 12 September 2018, Impact Assessment Terrorism Content Regulation Proposal, SWD(2018) 408, p. 22, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=SWD:2018:408:FIN.

[59] Regulation 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online, OJ [2021] L 172/79, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32021R0784. This new Regulation will apply from 7 June 2022. Article 2(1) defines a hosting service provider as "a provider of information society services consisting in the storage of information provided by and at the request of the content provider".

[60] Terrorism Content Regulation, Art. 7.

[61] *Ibid*, Art. 3. These can either be administrative, law enforcement or judicial authorities provided they fulfil their tasks in an objective and non-discriminatory manner and do not seek or take instructions from any other body in relation to the exercise of the tasks under the regulation (recital 35 and Art. 13).

[62] *Ibid*, Art.6.

[63] *Ibid*, Art. 5.

3.3.2. Regulation of moderation by specific types of digital intermediaries: Video-sharing platforms

In addition to the regulation of specific types of online illegal content, the EU also started to regulate moderation practices by a specific type of digital intermediaries. Indeed, the 2018 revision of the Audiovisual Media Services Directive (AVMSD) envisages that video-sharing platforms[64] should take appropriate measures to protect: (i) the general public from online content which violates EU law (i.e., racism and xenophobia, child sexual abuse material and terrorist content); (ii) the general public from other forms of hate speech which violates the principles mentioned in the EU Charter of Fundamental Rights (i.e., sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation); and (iii) minors from content which may violate the law or be harmful and impair their physical, mental or moral development.[65] Although the European Commission had initially foreseen[66] that the chapter on video-sharing platforms should lead to maximum harmonisation, this was changed during the course of adoption of the Directive, and member states are therefore free to introduce more far-reaching obligations for video-sharing platforms.

The AVMSD lists the possible measures to be taken such as. transparent and user-friendly mechanisms to report and flag content; systems through which video-sharing platforms explain to users what effect has been given to the reporting and flagging; easy-to-use systems allowing users to rate content; transparent, easy-to-use and effective procedures for the handling and resolution of users' complaints. The Directive specifies that the measures must be appropriate in the light of the nature of the content, the potential harm, the characteristics of the category of persons to be protected, the rights and legitimate interests at stake (in particular those of the video-sharing platforms and the users having created and/or uploaded the content, as well as the public interest). The measures should also be proportionate, taking into account the size of the video-sharing

[64] Directive 2010/13 of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive), OJ [2010] L 95/1, as amended by Directive 2018/1808, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02010L0013-20181218. Article 1(1aa) defines a video-sharing platform service as "a service as defined by Articles 56 and 57 TFEU, where the principal purpose of the service or of a dissociable section thereof or an essential functionality of the service is devoted to providing programmes, user-generated videos, or both, to the general public, for which the video-sharing platform provider does not have editorial responsibility, in order to inform, entertain or educate, by means of electronic communications networks (...) and the organisation of which is determined by the video-sharing platform provider, including by automatic means or algorithms in particular by displaying, tagging and sequencing."

[65] Audiovisual Media Services Directive, Art.28b(1). on the new obligations imposed on video-sharing platforms, see Valcke P., "The EU regulatory framework applicable to audiovisual media services', in Garzaniti L. et al. (eds.), *Telecommunications, broadcasting and the Internet. EU Competition law & regulation*, 4th ed., Sweet & Maxwell, 2019, pp. 232-235.

[66] See Explanatory Memorandum to Proposal for a Directive of the European Parliament and of the Council amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning provision of audiovisual media services in view of changing market realities, COM(2016) 287, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2016%3A287%3AFIN.

platform and the nature of the provided service. A national regulatory authority (often the media regulator) should assess the appropriateness of the measures.[67]

According to the European Commission, the requirements of the AVMSD are compatible with the liability exemption of the e-Commerce Directive, as the measures imposed on video-sharing platforms relate to the responsibilities of the provider in the organisational sphere and do not entail liability for any illegal information stored on the online platforms as such.[68] Moreover, the measures imposed on video-sharing platforms cannot lead to any *ex ante* control measures or upload-filtering of content.[69]

3.3.3. Regulation for all: A re-interpretation of the e-Commerce Directive

To improve the content moderation practices of all digital intermediaries, the Commission also adopted in 2017 a Communication[70] and then in 2018 a Recommendation[71] setting principles for the providers of hosting services as well as member states to take effective, appropriate and proportionate measures to tackle illegal content online. It sets out the general principles for all types of illegal content online and recommends stricter moderation for terrorist content.

Regarding the notice-and-takedown procedures which were not regulated by the e-Commerce Directive and were very divergent across member states,[72] the Recommendation calls for procedures that: (i) are effective, sufficiently precise and adequately substantiated; (ii) respect the rights of content providers with the possibility of counter-notices and out-of-court dispute settlements; and (iii) are transparent.[73]

Regarding proactive measures taken by the digital intermediaries to find and remove illegal content, the Recommendation encourages appropriate, proportionate and specific measures, which could involve the use of automated means, provided some safeguards are in place, in particular human oversight and verification.[74]

Regarding cooperation, the Recommendation encourages close cooperation with national, judicial and administrative authorities and trusted flaggers with the necessary expertise and determined on a clear and objective basis; it also encourages cooperation

[67] Audiovisual Media Services Directive, Art.28b(3)-(7).
[68] Proposal for a Directive of the European Parliament and of the Council amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning provision of audiovisual media services in view of changing market realities, COM(2016) 287.
[69] Audiovisual Media Services Directive, Art.28b(3).
[70] See fn 13.
[71] Recommendation 2018/334 of the European Commission of 1 March 2018 on measures to effectively tackle illegal content online**,** OJ [2018] L 63/50, https://eur-lex.europa.eu/legal-content/GA/TXT/?uri=CELEX:32018H0334.
[72] See ICF, Grimaldi Studio Legale, and 21c Consultancy, "Overview of the legal framework of notice-and-action procedures in Member States", study for the European Commission, 2018, https://op.europa.eu/en/publication-detail/-/publication/a56ceb47-2446-11e9-8d04-01aa75ed71a1.
[73] Recommendation 2018/334, Points 5-17.
[74] *Ibid*, Points 16-21.

among hosting services providers, in particular smaller ones which may have less capacity to tackle illegal content.[75]

3.3.4. Summary of the EU regulatory framework and current practices of online moderation

The table below outlines the EU rules against illegal content online according to the nature of the legal instrument (hard law, soft law, or self-regulation).

Table 1. EU regulatory framework on moderation of illegal content online

	Hard law	Soft law	Self-regulation
BASELINE ***All types of hosting platforms and all types of illegal content online***	- Directive 2000/31 on e-Commerce	- Commission Communication (2017) on Tackling *Illegal Content Online* - Commission *Re commendation 2018/334* on measures to effectively tackle illegal content online, Ch. II	
Additional rules for video-Sharing Platforms	- Directive 2010/13 Audiovisual Media Services as amended by Directive 2018/1808		
Additional rules for hate speech	- Council Framework Decision 2008/913 on combating certain forms and expressions of racism and xenophobia		- Code of conduct on illegal hate speech online (2016)
Additional rules for child sexual abuse material	- Directive 2011/93 on combating the sexual abuse and sexual exploitation of children and child pornography		- Alliance to Better Protect Minors Online (2017)

[75] *Ibid*, Points 22-28.

	Hard law	Soft law	Self-regulation
Additional rules for terrorist content	- Directive 2017/541 on combating terrorism - Regulation 2021/784 on addressing the dissemination of terrorist content online	- Commission *Recommendation 2018/334* on measures to effectively tackle illegal content online, Ch. III	- EU Internet Forum (2015)

Source: de Streel et al. (2020, p.33)

Current practices of online moderation vary according to the type and size of the platforms. They deploy a range of content moderation practices, which may be automated and/or which involve human review processes. Some also deploy prevention measures to make sure that harmful content is not seen by users, for instance by preventing certain users from uploading content or by making sure that minors do not see the content, through age verification or age assurance systems. According to recent research, a majority of platforms do not review content before it is uploaded, but they place the responsibility on the uploader to make sure that the content is compliant with the terms and conditions of the platform by ticking a box to that effect.[76] Most platforms have in place systems to detect content that may be in conflict with its terms and condition through flagging measures (including sometimes by trusted flaggers). Automated moderation is particularly widespread to detect child sexual abuse material and can lead to automatic removal if the illegal content is present (hashed) in a database. However, most other content detected by algorithms is reviewed by human moderators before it is removed. According to the same research, medium and large platforms invest 9% of their annual spend on in-house content moderation and they invest 16% to 29% of their annual spend on developing automated systems.

3.4. The end of the independence of cyberspace: The Digital Services Act

As a logical development of the regulatory move initiated 10 years ago, the Commission proposed in December 2020, with the Digital Services Act, new horizontal rules for the moderation of illegal content online applicable to all digital platforms and all content, clearly marking the end of the independence of cyberspace.[77]

[76] Report by Ernst and Young LLP, commissioned by the UK government "Understanding how platforms with videosharing capabilities protect users from harmful content online", 2021, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1008128/EYUK-000140696_EY_Report_-_Web_Accessible_Publication_2.pdf.

[77] As often explained by Commissioner Thierry Breton, the current progressive regulation of digital space repeats the previous progressive regulation of the terrestrial and then maritime spaces.

3.4.1. The proposed DSA and content moderation

The proposed DSA provides for four main categories of online intermediaries as a series of Russian dolls.[78] As we go from the biggest to the smallest doll, the rules imposed by the DSA become more numerous and stricter.

(i) The broadest category, the biggest doll, is the **provider of intermediary service**, which covers all providers of mere conduit, caching[79] and hosting services;

(ii) Then comes the **provider of hosting services** defined as storage of information provided by, and at the request of, a recipient of the service. This category includes the providers of cloud, file-sharing, and webhosting services;

(iii) Then comes the **online platform** defined as a provider of hosting services which, at the request of a recipient of the service, stores and disseminates to the public information. Such a category includes the providers of marketplaces, social media, app stores, and the collaborative economy;

(iv) Finally, the smallest doll is the **very large online platform** (VLOP) which is an online platform with at least 45m monthly active users in the EU (i.e., 10% of the EU population). This category includes most of the GAFAM.[80]

Rules on content moderation are scattered throughout the DSA but the approach fits with the logic of the proposal which is to introduce asymmetric rules depending on the type of intermediary (or Russian doll). In terms of the substantive rules, the proposed DSA introduces for the first time in EU law: transparency and due diligence obligations over content moderation practices; harmonised notice-and-action mechanisms with an obligation to motivate removal decisions; and rules on the suspension of accounts while granting rights to users to challenge content moderation decisions. VLOPs are subject to additional rules to ensure more comprehensive public oversight of their content moderation practices.

Before we turn to the rules on content moderation per se, it is important to note that the proposal introduces a so-called good Samaritan clause. It states that digital intermediaries "shall not be deemed ineligible for the exemptions from liability (…) solely because they carry out voluntary own-initiative investigations or other activities aimed at detecting, identifying and removing, or disabling access to, illegal content, or take the necessary measures to comply with the requirements of Union law, including those set out in this Regulation".[81] However, it has been said that this clause may lead to over-removals since, unlike under the US safe harbour and the US good Samaritan clause,[82] providers are not guaranteed protection if they fail to remove content once they have detected illegal content themselves. To be sure to be shielded from liability for third-party illegal content,

---

[78] Resp. DSA Proposal, Art. 2(f), 2(h) and 25, https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN.

79 For instance, Internet access providers, domain name registries and wi-fi hotspots.

80 GAFAM is the name given to the five largest and most dominant companies in the information technology industry of the United States, i.e. Google, Apple, Facebook, Amazon and Microsoft.

81 DSA Proposal, Art. 6, https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN.

82 Section 230(c) of the Communications Act of 1934, as amended by the Telecommunications Act of 1996 (47 U.S.C. § 230), https://www.law.cornell.edu/uscode/text/47/230.

providers could prefer to remove or disable access to the potentially illegal content, leading potentially to over-removals, which may impact the protection of fundamental rights, and in particular freedom of expression.[83]

3.4.2. Asymmetric obligations

3.4.2.1. All digital intermediaries: Transparency

All digital intermediaries in scope (technical intermediaries, hosting service providers, online platforms and VLOPs) would need to clearly inform users in their terms and conditions of any restrictions they impose on the use of their services, including their content moderation policies and in particular algorithmic decision-making and human review. Also, service providers would need to act in a diligent, objective, and proportionate manner in applying any restrictions, with due regard to the rights and legitimate interests of all parties involved, including applicable fundamental rights.[84]

On top of this, all intermediaries in scope (except for micro-enterprises) would need to produce annual reports on their content moderation activities, including the number of removal orders received from national authorities or notices received from users or flaggers, how fast they acted, and a detailed overview of their own-initiative content moderation activities (number and type of measures taken) and of the complaints-handling activities.[85]

The obligation becomes stricter for online platforms as they would have to report on any automatic content moderation procedures, by providing information on the purpose, indicators of accuracy and any safeguards applied.[86] VLOPs would need to publish transparency reports more frequently: every six months.[87]

3.4.2.2. Hosting intermediaries: Notice-and-action procedures

Hosting service providers would need to put in place notice-and-action systems to allow individuals and entities to notify them of allegedly illegal content.[88] The proposed DSA sets out the elements that need to be included in the notices. When all the elements are present, the provider is deemed to have actual knowledge, potentially triggering liability for third-party illegal content, if the provider fails to take down the illegal content. From receiving the notice, the provider would need to act quickly by sending a confirmation of receipt of the notice to the sender (and of whether automated means of processing or decision-

[83] For a detailed discussion on this point, see Chapter 2 of this publication.
[84] DSA Proposal, Art. 12, Ibid.
[85] DSA Proposal, Art. 13, Ibid.
[86] DSA Proposal, Art. 23, Ibid.
[87] DSA Proposal, Art. 33, Ibid.
[88] DSA Proposal, Arts. 14-15, Ibid.

making have been used) and by informing the sender of its decision which must be taken in a timely, diligent, and objective manner, and of the redress possibilities.

Hosting service providers that decide to remove or disable access to content would always need to inform the user at the latest at the time of removal of the decision by providing a statement of reasons, which would have to contain certain elements. These elements are for instance, the facts leading to the decision, if automated means were used, and a reference to legal grounds or to the provider's terms and conditions that were breached. The decisions and statement of reasons would need to be published in a publicly accessible database managed by the European Commission.

3.4.2.3. Online platforms: Trusted flaggers, user complaints and account suspension

Obligations become stricter for online platforms which also need to deal with notices submitted by trusted flaggers as a matter of priority and without delay.[89] The status (and revocation) of trusted flaggers would be decided by the Digital Service Coordinator[90] of the member state where the applicant/flagger is established if a number of set conditions are fulfilled. It is important to note that the status of trusted flagger is only foreseen to be awarded to entities and not to individuals.[91]

Online platforms would also need to provide their users with easy means to challenge content moderation decisions. As a first step, they would need to put in place internal complaint handling systems to allow users to complain about content moderation decisions.[92] Complaints would have to be receivable for at least six months following the contested decision. Systems would have to be available electronically, free of charge and be easy to access. Complaints would need to be handled in a timely, diligent, and objective manner and could lead to the reversal of the decision without undue delay. Online platforms would also need to inform complainants without undue delay of their decision and of the possibility of further redress mechanisms. Importantly, these decisions could not be taken by online platforms solely on the basis of automated means.

On top of this, users that have been the subject of a content moderation decision would be allowed to resort to a certified out-of-court dispute procedure to seek redress.[93] The proposed DSA sets out the conditions under which the Digital Service Coordinator would have to certify out-of-court dispute resolution bodies. These conditions are aimed at

[89] DSA Proposal, Art.19, Ibid..
[90] Digital Service Coordinators would be responsible for all matters relating to application and enforcement of the DSA in a member state, unless a member state has assigned certain specific tasks or sectors to other competent authorities. The requirements for Digital Service Coordinators are specified in Art. 39 of the DSA Proposal: they must perform their tasks in an impartial, transparent and timely manner; they must have adequate technical, financial and human resources to carry out their tasks; they should act with complete independence and remain free from any external influence; and they should neither seek nor take instructions from any other public authority or any private party.
[91] DSA Proposal, Rec.46, Ibid.
[92] DSA Proposal, Art. 17, Ibid. This covers decisions leading to the removal or disabling of access to information, and the suspension or termination of the service to the recipient, or of the user's account
[93] DSA Proposal, Art. 18, Ibid.

ensuring in particular that the bodies are impartial and independent of the online platforms and that they have the necessary expertise. Online platforms would have to engage in good faith with the selected body and would be bound by the decision taken by the body. If the dispute is settled in favour of the user, the platform would need to reimburse all fees and expenses incurred by the user to settle the decision. Of course, users could also seek redress in court, in accordance with their national law.

Of a different nature but worth noting, recipients of services would also have the right to lodge a complaint against providers (if they infringe the DSA) with the Digital Service Coordinator of the member state where the recipient resides.[94] This is not a dispute resolution mechanism though, since the Digital Service Coordinator only needs to assess the complaint and where appropriate, transmit it to the Digital Services Coordinator of establishment.

The proposed DSA also frames the conditions under which online platforms would be able to suspend the provision of services, in other words to suspend user's accounts.[95] This would only be possible for users that frequently provide manifestly illegal content, that is to say where it is evident to a layperson, without any substantive analysis, that the content is illegal.[96] Suspension could only be temporary and after issuance of a prior warning. Platforms would have to take the decision on a case-by-case basis by taking into account a number of listed circumstances including the gravity, the number of occurrences, and the intention. The terms and conditions of the online platforms would have to set out their policy in this respect, which could contain stricter measures in case of manifestly illegal content related to serious crimes. A similar procedure is also foreseen to suspend the processing of manifestly unfounded complaints and notices. Users would be able to challenge suspension decisions as explained above.

3.4.2.4. Very large online platforms: Systemic risk assessment

Very large online platforms would have to identify, analyse and assess at least once a year any significant systemic risks[97] stemming from their services, including the dissemination of content which violates the law but also content which does not violate the law but is harmful.[98] When doing so, they would have to take into account how their content moderation systems influenced any of the systemic risks. On the basis of the assessement, the VLOPs would need to put in place reasonable, proportionate and effective mitigation measures (such as adapting the content moderation practices) tailored to the specific systemic risks identified. Moreover, the European Commission would be able to issue

[94] DSA Proposal, Art. 43, Ibid.
[95] DSA Proposal, Art. 20, Ibid.
[96] DSA Proposal, Rec. 47, Ibid.
[97] Systemic risks are not defined in the DSA Proposal which only provides that three categories of systemic risk should be analysed, DSA Proposal, Art. 26: "(i) the dissemination of illegal content through their services; (ii) any negative effects for the exercise of the fundamental rights to respect for private and family life, freedom of expression and information, the prohibition of discrimination and the rights of the child (…); (iii) intentional manipulation of their service, including by means of inauthentic use or automated exploitation of the service, with an actual or foreseeable negative effect on the protection of public health, minors, civic discourse, or actual or foreseeable effects related to electoral processes and public security.", Ibid.
[98] DSA Proposal, Arts. 26-27, Ibid.

general guidelines in relation to specific risks, in particular to present best practices and recommend possible measures.

3.4.2.5. Overseeing content moderation

There are no special rules on the oversight of content moderation by public authorities in the proposed DSA although individual decisions can be challenged via internal complaints, by (certified) out-of-court dispute settlement and also in courts. The designated Digital Service Coordinator of the member state where the intermediary is established would be in charge of ensuring application and enforcement of the DSA, unless special tasks were assigned to other competent authorities.[99] On top of receiving the transparency reports, the Digital Service Coordinator would receive certain powers of investigation such as the power to require providers to deliver information, to carry out on-site inspections and to ask members of staff to provide explanations.

Strengthened rules are also foreseen in relation to the supervision of VLOPs, including the appointment of a compliance officer, the intervention of independent auditors and special rules to access data[100] as well as the possibility for the European Commission to directly regulate the VLOP instead of the Digital Service Coordinator of the member state where the platform is established.[101]

3.5. Concluding remarks

The evolution of the EU regulatory framework on the moderation of illegal online content, including its most recent step with the proposed Digital Services Act, is interesting. States are progressively regulating cyberspace, which has become increasingly important for the life of their citizens and businesses, and which has not delivered on the – admittedly naïve – promises of the libertarians.[102] In this endeavour, states could be mindful of preserving the greatest opportunities of the Internet, in particular to enhance the exercise of our fundamental freedoms. In that regard, the approach followed by the EU is also interesting. On the one hand, by introducing procedural accountability obligations, it regulates the process of content moderation and not its results. On the other hand, it tailors the obligations to the risks created by illegal content and by platforms.[103] However, some aspects of the proposed DSA could perhaps be clarified and improved.

[99] DSA Proposal, Art. 38, Ibid.
[100] DSA Proposal, Arts. 28, 31 and 32, Ibid.
[101] DSA Proposal, Arts. 50-66, Ibid.
[102] Like Barlow's hope of creating a "civilization of the Mind" more humane and fairer than what states had created before.
[103] In favour of a risk-based approach and asymmetric rules, see among others, Buiten M., de Streel A., and Peitz M., "Rethinking liability rules for online hosting platforms", *International Journal of Law and Information Technology* 28, 2020, pp. 139-166, https://academic.oup.com/ijlit/issue/28/2.

3.5.1.1. Scope

Content moderation can take place for all kinds of illegal content, with no distinction between manifestly illegal content and other forms of illegal content. However, a different take-down procedure – possibly with accelerated deadlines, enhanced communication channels to public authorities and retention obligations regarding evidence (similar to what is foreseen under the Terrorism Content Regulation) could be envisaged for manifestly illegal content where it is evident to a layperson, without any substantive analysis, that the content is illegal.[104]

Also, clear rules on the territorial scope of application of content moderation decisions are missing. Since illegal content is also defined by reference to national law, content may be illegal according to the legislation of one member state but not by reference to the legislation of another member state. It is therefore important to address the territorial scope of take-down decisions in the DSA since this could lead to over-removal which could jeopardise freedom of expression in certain countries.

3.5.1.2. Challenging content moderation decisions

The solution envisaged in the proposed DSA for online platforms is sound in our view because certified out-of-court dispute resolution bodies would be able to reassess and potentially reverse content moderation decisions. The proposal puts in place a number of guarantees, such as independence, but it will be important to correctly inform users of the redress mechanism and to specify deadlines to settle the dispute. As it stands, the proposal only allows "recipients of the service" (i.e. a user of a service) addressed by a content moderation decision to select an out-of-court dispute body to resolve a dispute. This means for instance that associations representing specific interests would not have the right to challenge content moderation decisions.

3.5.1.3. Oversight of the use of AI content moderation tools

Aside from the requirement to be transparent on the use of automated content moderation systems, the proposed DSA does not refer to criteria to be met by technology used for the detection of illegal content. Thus any automated content moderation would only be subject to the general EU law applicable to automated systems.[105] It would be helpful if any automated moderation system were bound to comply with the six key requirements proposed by the EU High-Level Expert Group on AI: human agency and oversight; technical

[104] Also Frosio G. and Geiger C., "Taking fundamental rights seriously in the Digital Services Act's platform liability regime", *European Law Journal*, 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3747756.

[105] In particular the need for human oversight when privacy is at stake, Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), OJ [2016] L 199/1, Art. 22) https://eur-lex.europa.eu/eli/reg/2016/679/oj, and the possible prohibition of manipulative AI systems, Proposal of the European Commission of 21 April 2021 for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), COM(2021) 206, Art. 5), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

robustness and safety; privacy and data governance; transparency, diversity, non-discrimination and fairness; societal and environmental wellbeing; and accountability.[106] Moreover, the VLOPs, which have the data, expertise and financial means to develop automated techniques, may usefully share these technologies with small and medium-sized or new platforms.[107] Finally, it is interesting to see that the UK's Online Safety draft bill specifies that the regulator (Ofcom) will be given the power to require that a service provider uses accredited technology, at least to identify and remove terrorist content and child sexual exploitation if Ofcom has reasonable grounds to believe that the service provider is not removing such content.

3.5.1.4. VLOPs and fundamental rights when moderating content

Given that VLOPs may be considered as organising a "public space",[108] it may now be time to ask them to respect the fundamental rights enshrined in the Charter of Fundamental Rights of the EU in their content moderation practices.[109] The reference to the fact that online platforms need to have due regard for the rights of all parties involved, including applicable fundamental rights could possibly become a positive duty to respect fundamental rights, which of course will need to be balanced out between each other. The terms and conditions of VLOPs could also be scrutinised *ex ante* by the Digital Services Coordinator and/or the Commission to make sure they respect all applicable legislation.

3.5.1.5. Journalistic content or content edited by audiovisual media service providers

As a contrast to the UK's Online Safety Bill,[110] the proposed DSA does not contain any special treatment in relation to professionally edited content, such as journalistic content or content that is under the editorial responsibility of audiovisual media service providers. The UK Bill specifies that so-called Category 1 services[111] have special duties (to be specified by Ofcom in dedicated codes of conduct) to protect content of democratic importance and

---

[106] European Communication White Paper of 19 February 2020 on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:65:FIN; High-Level Expert Group on Artificial Intelligence, Ethics Guidelines of 8 April 2019 for Trustworthy AI. See also Terrorism Content Regulation, Art.5(3), https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[107] European Commission Recommendation 2018/334, Point 28.

[108] Elkin-Koren N. and Perel M., "Guarding the guardians: Content moderation by online intermediaries and the rule of law' in Frosio G. (ed), *The Oxford Handbook of Online Intermediary Liability*, Oxford University Press, 2020, pp. 669-678, https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780198837138.001.0001/oxfordhb-9780198837138-e-34.

[109] Pollicino O., *Judicial protection of fundamental rights on the Internet: A road towards digital constitutionalism?,* Hart, 2021.

[110] Draft published on 12 May 2021 (Bill CP 405), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf.

[111] Category 1 services are subject to additional rules, and the thresholds to be met will be determined by the minister in charge (the Secretary of State). At least one of the threshold conditions would have to be the number of users.

journalistic content. In particular, the UK draft foresees that a special complaints procedure would need to be put in place in relation to content moderation decisions affecting access to journalistic content,[112] with the terms and conditions of platforms having to specify the importance of freedom of expression when taking content moderation decisions in relation to such content. The recently adopted Terrorism Content Regulation also contains a special carve-out for material disseminated to the public for "educational, journalistic, artistic or research purposes or for the purposes of preventing or countering terrorism, including material which represents an expression of polemic or controversial views in the course of public debate".[113]

3.5.1.6. Protection of minors and harmful content

With regard to legal but harmful content which could be damaging to minors, the only rules would apply to VLOPs and relate to systemic risk assessments and risk mitigation measures which would need to be taken. These measures are not defined at this stage. Nothing is foreseen in relation to other digital intermediaries, which will mean that this matter will be addressed in the platforms' terms and conditions, without public intervention. Age verification measures and content rating systems are difficult areas to address at the EU level but leaving this whole area to member state legislation would lead to continued tensions between the member states and could weaken the digital single market. In this regard, it is interesting to note that the revised AVMSD foresees that video-sharing platforms should protect minors from content which may impair their physical, mental or moral development. Also, the UK's Online Safety Bill which echoes many of the provisions of the proposed DSA foresees that all providers in scope would need to conduct a risk assessment of whether children are likely to access their services and providers will only be able to conclude that it is not possible for children to access a service if robust systems and processes such as age verification are in place.

[112] Interestingly, the text defines journalistic content as content generated for the purpose of journalism, and which is "UK-linked".
[113] Terrorism Content Regulation, Art. 1.3.