To what extent can online service providers adopt voluntary content moderation measures without losing their immunity shields? A comparative analysis of online service providers' liabilities in the European Union and the United States

by **Adriana Berbec** *

Abstract: TImmunity for engaging in voluntary content moderation measures is a new addition to the European Union legal framework for intermediaries' liabilities. Article 7 of the Digital Services Act guarantees eligibility for immunity to online service providers undertaking good faith voluntary own-initiative investigations similar to the Good Samaritan provision originating in Section 230 of the US Communication Decency Act. The latter has been in place for more than two decades and the breadth of US case law sheds some insights into the strengths and weakness of this provision. This research paper aims to identify similarities and differences between the rules that protect online Good Samaritans in both jurisdictions and determine whether the rules effectively fight illegal content online without undermining the immunity of online service providers. It does so by looking at the relevant jurisprudence and the existing legal provisions on liability exemptions for voluntary content moderation in both jurisdictions. It further examines the proposals to amend Section 230 in the US which are a symptom of the dissatisfaction surrounding the broad immunity granted to online service providers and the perceived, occasionally misconstrued, shortcomings of the provision. Additionally, they provide indications as to whether limiting the immunity shields to online service providers engaging in voluntary content moderation measures aligns with the standards of good faith and diligence set forth in the Digital Services Act.

Keywords: Content Moderation, Digital Services Act, Section 230, Liability, Online Platforms

Recommended citation: Adriana Berbec, To what extent can online service providers adopt voluntary content moderation measures without losing their immunity shields? A comparative analysis of online service providers' liabilities in the European Union and the United States, 15 (2024) JIPITEC 13 para 1.

A. Introduction

1 In both the European Union ("EU") and the United States ("US"), the rules on liability exemptions are meant to protect online service providers undertaking voluntary content moderation measures to remove or disable access to illegal or objectionable content. In the EU, Article 7 of the Digital Services Act (DSA)[1] guarantees that online service providers do not lose their eligibility for liability exemptions when they carry out, in good faith and in a diligent manner, voluntary own-initiative investigations to

* Adriana Berbec holds an LLM in IP and ICT Law from KU Leuven, Belgium. The information and views set out in this article are those of the author and do not necessarily reflect the official opinion of the author's employer.

1 Regulation of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (the "Digital Services Act") (DSA).

remove or disable access to illegal content. Their exemption from liability is however conditioned by the requirement that the service provider expeditiously removes illegal content once they acquire knowledge or awareness of it. In the US, Section 230(c)(2) of the US Communications Decency Act (CDA)[2], also known as the Good Samaritan provision, is meant to promote good faith voluntary content moderation measures by protecting online service providers for any action undertaken to remove content that the provider or user of an interactive computer service deems objectionable. Their exemption from liability is not conditioned by promptly removing objectionable content once they acquire knowledge about it. The provisions in both jurisdictions nevertheless share the common element of acting in good faith when undertaking voluntary content moderation measures. The DSA is more detailed and explicit about what constitutes good faith. Section 230 does not provide a statutory definition of what constitutes good faith, but US Courts have generally interpreted the term based on what it does not constitute acting in good faith. The ambiguity of what constitutes objectionable content has also led to interpreting the term either as being an objective standard or a subjective one. In addition, the removal decisions of online service providers have been afforded immunity also under Section 230(c)(1) of the CDA which is meant to protect online service providers from publishing third-party content. This broad interpretation has raised criticism since Section 230(c)(1) does not provide for a good faith covenant, nor does it apply to a certain type of content, such as content similar to pornography, violence, obscenity or harassment outlined in Section 230(c)(2). Several proposals and legislative developments to amend Section 230(c)(2) in the US have therefore emerged that would make the immunity of service providers contingent upon specific safeguards or conditions when they engage in content moderation. One proposal aims that immunity for removal decisions should be available only under Section 230(c)(2), while others suggest a more precise definition of what constitutes good faith and objectionable content.

2 The aim of this paper is to undertake a comparative analysis between the voluntary content moderation measures outlined in Article 7 of the EU DSA and Section 230(c)(2) of the US CDA. There are several reasons for choosing to compare these two jurisdictions. First, they have both contributed to the existing legal framework on service provider liabilities either through statutory laws or jurisprudence. Second, the EU service provider liability exemptions have been influenced by the US system of knowledge-based liability doctrine of the US Digital Millennium Copyright Act (DMCA).[3] Third, the voluntary content moderation measures, known as the Good Samaritan principle, originate from Section 230(c)(2). Finally, the DSA addresses all major online service providers, the majority of which are US-based companies that offer their services to EU users. The ultimate goal is to assess whether the legal provisions in both jurisdictions manage to achieve their desired objective of fighting illegal content online while preserving the immunity status of online service providers. It does so by examining the legislative framework governing the online service provider liabilities in the EU and the US, in particular the rules on voluntary content moderation measures in the DSA and Section 230.

3 This research paper is structured as follows. Chapter A serves as an introduction. Chapter B provides the regulatory framework for online service providers liability in both the EU and the US. Chapter C examines the jurisprudence from the EU Court of Justice ("CJEU") and of the US Courts with regards to liability exemptions and voluntary content moderation measures. Chapter D analyses the recent proposals to modify the Good Samaritan provision in the US, the objective being to determine the perceived weaknesses of Section 230(c) and the solutions to tackle them. Chapter E concludes the findings of the research and provides some general reflections in relation to the interplay between voluntary content moderation measures and liability shields.

B. Regulatory framework for online service providers liability

4 This chapter describes the framework directive governing electronic commerce in the EU which, *inter alia*, regulates intermediary liability. It will touch upon the transition of intermediary liability regime from a directive to a regulation, by analysing the similarities and differences between the relevant legal provisions in the directive and the regulation. Similarly, the chapter explores the main legislation that governs intermediary liability in the US, along with the sequential steps that have led to the creation of Section 230. Subsequently, the chapter will examine the voluntary content moderation measures outlined in the DSA and in Section 230(c)(2), known as the Good Samaritan provision.

2 Communications Decency Act (CDA), also called Title V of the Telecommunications Act of 1996, enacted by the US Congress primarily in response to concerns about minors' access to pornography via the Internet.

3 Digital Millennium Copyright Act of 1998 (DMCA), < https://www.copyright.gov/dmca/ >, accessed 16 March 2024.

I. European Union

5 At the EU level, Directive 2000/31/EC on electronic commerce (hereby 'the Directive'), is the legislation that regulates central legal aspects of electronic commerce, including online communications, online contracts, and intermediary liability.[4] Its objective is to contribute to the proper functioning of the internal market by ensuring the free movement of information society services between Member States.

6 Section 4 of the Directive lays down intermediaries' liabilities exemptions for all the unlawful activities carried out by third parties, subject to conditions as laid down in Article 12 (*mere conduit*)[5], Article 13 (*caching*)[6] and Article 14 (*hosting*).[7,8] Among these three provisions, Article 14 is the most important one as it basically reflects the knowledge-based liability principle, and it applies to providers that host third-party content on their servers. The Directive clarifies that the liability exemptions in the Directive apply to intermediary service providers when their activity *"is of a mere technical, automatic and passive nature, which implies that the information society service provider has neither knowledge of nor control over the information which is transmitted or stored"*.[9] According to the Directive, the passive nature of the provider is commensurate with a lack of knowledge or control over the content. This passive nature of the service provider guarantees its liability exemptions.

7 More than two decades after the adoption of the Directive, the European Commission (the 'Commission'), in light of the *"new and innovative business models ...[that] have allowed business users and consumers to impart and access information and engage in transactions in novel ways"*[10], proposed a new Regulation on a single market for digital services and amending Directive 2000/31/EC (Digital Services Act), shortly known as the DSA. The DSA entered into force on 16 November 2022. The DSA maintains the liability regime in the Directive, but it introduces transparency requirements and due diligence obligations proportionate to the size of the intermediary service provider. The DSA also includes online search engines and online platforms which were left out in the Directive. Online platforms are defined as a sub-category of internet intermediaries that provide a digital hosting service at the request of a recipient of the service.[11] The hosting service includes the storing, but also the dissemination of information to the public, unless that activity is a minor and purely ancillary feature of another service or a minor functionality of the principal service.

8 The liability exemptions contained in Articles 12 to 14 of the Directive are now construed as references to Articles 4, 5, and 6 of the DSA. Likewise, the DSA also maintains the provision on the prohibition for general monitoring in Article 15 of the Directive, now construed as reference to Article 8 DSA.

9 Pursuant to Article 6 of the DSA (former Article 14 of the Directive), a (hosting) service provider is exempt from liability of third-party illegal content if it *"does not have actual knowledge of illegal activity or illegal content and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or illegal content is apparent"* (Article 6(1)a), or *"upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the illegal content"* (Article 6(1)b).[12] Illegal content is defined as information relating to illegal content, products,

4 Directive 2000/31/EC of the European Parliament and of the Council on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), Recital 7.

5 *Ibid.*, Article 12, according to which a service provider is not liable for the information transmitted or accessed if it *"does not initiate the transmission; does not select the receiver of the transmission; and does not select or modify the information contained in the transmission"*.

6 *Ibid.*, Article 13, according to which a service provider is not liable for the automatic, intermediate and temporary storage of information if it *"does not modify the information; complies with conditions on access to the information; complies with rules regarding the updating of the information [...]; does not interfere with the lawful use of technology [...] to obtain data on the use of the information; and acts expeditiously to remove or to disable access to the information upon obtaining actual knowledge of the fact that the information at the initial source of the transmission' has been removed or disabled"*.

7 *Ibid.*, Article 14, according to which a service provider is not liable for the information transmitted or accessed on the condition that *"(a) the provider 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; or (b) the provider, upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information"*.

8 For each of Articles 12, 13, and 14, even if no liability is established, national courts and administrative authorities can require the service provider to terminate or prevent an infringement or to remove or disable access to illegal information respectively, in accordance with the law of the Member State.

9 Directive 2000/31/EC, Recital 42.

10 DSA, Recital 1.

11 *Ibid.*, Article 3(i).

12 *Ibid.*, Article 6(1).

services, and activities[13] that are not in compliance with the law of the Union or of any Member State.[14]

10 According to Article 6(2) of the DSA, Article 6(1) does not *"apply when the recipient of the service is acting under the authority or control of the provider"*.[15] The DSA goes a bit further than the Directive and introduces Article 6(3) to indicate that Article 6(1) does not apply with respect to liability under consumer protection law of online platforms that allow customers to conclude distance contracts with traders, if the online platform leads an average consumer to believe that the product or service is provided either by the online platform itself or by a recipient of the service who is acting under its authority or control.

1. Voluntary content moderation measures under the DSA

11 Article 7 of the DSA, entitled "Voluntary own-initiative investigations and legal compliance" fills in a gap in the Directive[16] by introducing a provision relating to liability exemptions for intermediary service providers engaging in voluntary own-initiative investigations. An 'intermediary service' is defined as an information society service that provides either a 'mere conduit', 'cashing', or 'hosting' service.[17] For ease of comparison with the immunity of interactive computer service providers in Section 230, 'intermediary service providers' will be referred to as 'online service providers'.

12 The concept of extending protections to online service providers engaging in voluntary pro-active measures dates back to 2017, when the Commission considered the option of introducing a Good Samaritan provision aimed at encouraging service providers to tackle illegal content (*"proactive steps to detect, remove or disable access to illegal content (the so-called "Good Samaritan" actions"*).[18] This is now reflected in Article 7 of the DSA which holds that online service providers shall not lose the liability shields referred to in Articles 4, 5 and 6 of the DSA *solely* because they *"in good faith and in a diligent manner, carry out voluntary own-initiative investigations into, or take other measures aimed at detecting, identifying and removing, or disabling access to, illegal content"*.[19] Article 7 further aims at reassuring online service providers that any measures undertaken to comply with the requirements in the DSA (such as the due diligence obligations) do not lead to unavailability of the exemptions[20] due to categorizing such actions as 'active' involvement.[21] Unlike own-initiative investigations which are voluntarily initiated by online service providers and are meant to promote self-regulation, the measures undertaken for legal compliance pertain to mandatory (due diligence) obligations that online service providers must fulfill. To maintain the focus of the comparative analysis with the liability regime applicable under US Section 230(c)(2) concerning Good Samaritan voluntary content moderation measures, this research paper will center on the liability exemptions available for voluntary own-initiative investigations to remove or disable access to illegal content (from now on 'voluntary content moderation measures').

2. Knowledge and take-down

13 According to the DSA, liability exemptions for voluntary content moderation measures are subject to several conditions. First, pursuant to Article 6 of the DSA, liability exemptions are conditional upon online service providers (i) lacking actual knowledge of the illegal content or awareness of facts or circumstances from which the illegal activity or illegal content is apparent (Article 6(1)((a)), or (ii) acting expeditiously to remove illegal content once they obtain actual knowledge or awareness of the illegal content (Article 6(1)(b)). Since knowledge and awareness can be acquired not only through notices submitted by third parties, but also through own-initiative investigations,[22] online service providers can avoid liability if they act expeditiously to remove

13 *Ibid.*, Recital 12.

14 *Ibid.*, Article 3(h).

15 *Ibid.*, Article 6(2).

16 Directive 2000/31/EC only specifies in Recital 48 that the Directive does not prevent Member States to request hosting providers to apply a duty of care to detect illegal activities.

17 *Ibid.*, Article 3(g).

18 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Tackling Illegal Content Online, Towards an enhanced responsibility of online service providers COM(2017) 555.

19 DSA, Article 7.

20 *Ibid.*, Article 7 ("to comply with the requirements of Union law and national law in compliance with Union law, including the requirements set out in this Regulation").

21 Communication from the Commission COM(2017) 555, *supra* note 18.

22 *Ibid.*, Recital 22.

or disable access to the illegal content[23] in line with Article 6(1)(b) also when engaging in voluntary content moderation. This has been acknowledged by the Commission in its Communication and later on, its Recommendation on tackling illegal content online.[24] More specifically, the Commission explained that taking such voluntary proactive measures *"does not automatically lead to the online service provider losing the benefit of the liability exemption provided for in Article 14 of Directive"*.[25] This is due to the fact that acting expeditiously to remove or disable access to illegal content the online service providers continue to *"benefit from the liability exemption pursuant to point (b) of Article 14(1) [of the Directive]"*.[26] Accordingly, when online service providers undertake voluntary content moderation measures to remove content whose illegality is apparent[27] their acquired knowledge and awareness of that specific illegal content does not automatically render them liable. Online service providers lose their liability exemptions only if they fail to expeditiously remove the specific content whose illegality is obvious.[28] Similarly, the mere fact that online service providers undertake such voluntary measures does not make them active providers in respect of the illegal content posted on their servers[29] as interpreted in Recital 18 of the DSA. This is also the view expressed by the Advocate General in *YouTube and Cyando*.[30]

14 To better understand the rules on liability exemptions, a reading of the Recitals of the DSA is necessary. Although the Recitals are not legally binding, they play an important role in the decisions of the CJEU as they help with the interpretation of the operative provisions of the Regulation.

15 Thus, Recital 18 clarifies that liability exemptions are available as long as the service providers confine themselves to *"providing the services neutrally by a merely technical and automatic processing of the information"* and do not play *"an active role of such a kind as to give it knowledge of, or control over, that information"*.[31] This Recital rephrases the liability exemptions in the Directive, by focusing on the *active* nature of the service provider that *removes* the liability exemptions as opposed to the *passive* nature of the service provider that *guarantees* the liability exemptions in the Directive. Neutrality, as the Recital reads, is correlated with knowledge and an online service provider which acquires knowledge of illegal content can still benefit from liability exemptions provided it expeditiously removes that specific content in line with Article 6 of the DSA.

16 Recital 22 provides further insights into the interplay between liability exemptions and knowledge of illegal content. Therefore, the fact that an online service provider automatically indexes information, has a search function, or recommends information based on profiles or preferences is not sufficient to conclude it has a 'specific' knowledge of the illegal content. Nor would an online service provider become knowledgeable solely by being aware, in a general sense, that its service is also used to store illegal content.[32] In other words, being aware that online service providers, although designed to be used for legal purposes, are inevitably used by third parties also for illegal purposes, does not lead to

23 Joan Barata, 'Positive Intent Protections: Incorporating a Good Samaritan principle in the EU Digital Services Act' (2020), Centre for Technology and Democracy, <https://cdt.org/wp-content/uploads/2020/07/2020-07-29-Positive-Intent-Protections-Good-Samaritan-principle-EU-Digital-Services-Act-FINAL.pdf>, accessed 16 March 2024.

24 Communication from the Commission COM(2017) 555, *supra* note 18, section 3.3. See also Commission Recommendation of 1 March 2018 on measures to effectively tackle illegal content online, C(2018)1177.

25 *Ibid*.

26 *Ibid*.

27 DSA, Article 6 (*"does not have knowledge [...] or is not aware of facts or circumstances from which the illegal activity or the illegal content is apparent"*).

28 See also Domingos Fahrino, 'The Digital Services Act: A European Digital Republic, If We Can Keep It Or The Long And Winding Road' (*The Digital Constitutionalist*, 7 December 2022) (*"by having content moderation, service providers, especially hosting ones, do not waiver the exemptions of liability they are given, but, if in the course of moderating content, illegal one is found or made apparent (see article 6(1)(a) and (b)) than the service provider is liable for such content if it does not act to counter it."*), <https://digi-con.org/the-digital-services-act-a-european-digital-republic-if-we-can-keep-it-or-the-long-and-winding-road/>, accessed 16 March 2024.

29 Communication from the Commission, COM(2017) 555, *supra* note 18.

30 Opinion of Advocate General Saugmandsgaard Øe delivered on 16 July 2020 in *Frank Peterson v Google LLC and Elsevier Inc. v Cyando AG*, Joined Cases C682/18 and C683/18 (*'YouTube and Cyando'*), para. 166 (*"it is necessary to avoid an interpretation of the concept of 'active role' that could produce the paradoxical result whereby a service provider conducting research on its own initiative into the information which it stores[...], would lose the benefit of the exemption from liability laid down in Article 14(1) of that directive and would, therefore be treated more severely than a provider which does not"*).

31 DSA, Recital 18.

32 *Ibid*., Recital 22.

knowledge-based liability.[33]

17 Recital 26 further confirms that such measures *"should not be taken into account when determining whether the provider can rely on an exemption from liability, in particular as regards whether the provider offers its service neutrally [...] without this rule implying that the provider can necessarily rely thereon"*.[34] In other words, undertaking voluntary content moderation measures should not be taken into account to determine whether the online service provider can claim or invoke an exemption from liability,[35] in particular for determining that the online service provider offers its services neutrally. The rule does not mean that the online service providers can necessarily invoke an exemption from liability.[36] The only thing that these voluntary measures guarantee is that the online service providers can invoke eligibility for liability exemptions (*"shall not be deemed ineligible"*[37]). Whether the online service provider is exempted from liability depends on whether the service provider satisfies the conditions for liability exemptions in Article 6 of the DSA. Kuczerawy explained that *"taking voluntary actions in good faith neither guarantees nor precludes neutrality"* and that the online platforms may still lose immunity.[38]

3. Good faith and diligence

18 Second, pursuant to Article 7 of the DSA, online service providers do not lose their eligibility for liability exemptions referred to in Articles 4, 5 and 6 of the Regulation, *"solely because they, in good faith and in a diligent manner, carry out voluntary own-initiative investigations into, or take other measures aimed at detecting, identifying and removing, or disabling access to, illegal content"*.[39] Recital 26 provides more clarity by indicating that the mere undertaking of voluntary measures does not render the liability exemptions unavailable on the condition that these measures are taken in good faith and in a diligent manner. These conditions include *"acting in an objective, non-discriminatory and proportionate manner, with due regards to the rights and legitimate interest of all parties involved, and providing the necessary safeguards against unjustified removal of legal content"*.[40] To that aim, where automated tools are used, the technology used must be sufficiently reliable *"to limit to the maximum extent possible the rate of errors"*.[41]

19 It can be inferred from this Recital that acting in good faith means acting in an objective, non-discriminatory and proportionate manner, while duly considering the rights and legitimate interests of all the parties involved.[42] This interpretation is reinforced by the Directive on Unfair Terms in Consumer Contracts ('UCTD') which provides that the seller or the supplier can meet the good faith requirement by dealing *"fairly and equitably with the other party whose legitimate interests he has to take into account"*.[43] It is worth noting that 'fairly' and 'equitably' are synonyms of 'objective', 'non-discriminatory' and 'proportionate' and that the UCTD, just like the Recital 26 of the DSA, specifically requires that the legitimate interests (of the other party) must be taken into account when acting in good faith. The Commission Notice on the interpretation of UCTD also confirms that *"good faith is an objective concept linked to the question of whether [...] the contract term in question is compatible with fair and equitable market practices that take the consumer's*

33 By analogy with the safe harbours in DMCA, see for instance Emerald Smith, 'Lord of the Files: International Secondary Liability for Internet Service Providers'(2011) in *68*(3) *Wash.& L.L. Rev.* (*"The court interpreted the DMCA placing the burden of policing content on copyright owners as logical given that the service platforms in question contain both infringing and non-infringing works and submission methods can make it difficult to determine which is which"*), < https://scholarlycommons.law.wlu.edu/wlulr/vol68/iss3/24 >, accessed 16 March 2024.

34 DSA, Recital 26.

35 In French: *« si ledit fournisseur peut se prévaloir d'une exemption de responsabilité »*, DSA, Recital 26.

36 In French: *« [...]cette règle n'impliquant cependant pas que ledit fournisseur peut nécessairement se prévaloir d'une exemption de responsabilité »*, DSA, Recital 26.

37 DSA, Article 7.

38 Aleksandra Kuczerawy, 'The Good Samaritan that wasn't: voluntary monitoring under the (draft) Digital Services Act' (*Verfassungblog*, 12 January 2021), <https://verfassungsblog.de/good-samaritan-dsa/>, accessed 16 March 2024.

39 DSA, Article 7.

40 *Ibid.*, Recital 26.

41 *Ibid*.

42 See also Jacob van de Kerkhof, 'Good Faith in Article 6 Digital Services Act (Good Samaritan Exemption)' (*The Digital Constitutionalist*, 15 February 2023) (*"the components of good faith in Recital 26 are objectivity, non-discrimination, proportionality, due regard of rights and interests of users and necessary safeguards in place to ensure automated technologies are sufficiently reliable"*), <https://digi-con.org/good-faith-in-article-6-digital-services-act-good-samaritan-exemption/>, accessed 16 March 2024.

43 Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts, No L 95/30, Recital 16.

legitimate interests sufficiently into account".[44]

20 Recital 26 does not distinctly separate acting in good faith from acting diligently, and while the two standards may overlap,[45] a diligent operator must also remove content when its illegal character is manifestly evident.[46] The statement of the Advocate General Saugmandsgaard Øe in the *YouTube and Cyando* judgment accurately conveys this concept by explaining that *"a provider must remove such information only where its illegal nature is 'apparent', that is to say manifest"*.[47] This judgment aligns with the principles oultlined in Recital 53 of the DSA pursuant to which the illegal character of the content can be considered apparent, when a third-party notice contains sufficient information to enable a diligent operator to establish that character *"without a detailed legal examination"*.[48] The service provider is required to diligently assess the facts brought to its attention concerning specific illegal information[49] and to address it.[50] The requirement for a diligent online service provider to remove the content which, based on a notice from third-party, appears sufficiently illegal is likely to be applicable in relation to own-initiative investigations also.[51]

21 Exercising diligence in the context of voluntary content moderation measures does not come without challenges. In particular, questions have been raised about how to measure whether an online service provider acted diligently when failing to remove some but not all illegal content.[52] For instance, consider a scenario where the same illegal content is shared on two different platforms. If one platform identifies it and removes it, while the other overlooks it and fails to remove it,[53] such instances of unsuccessful content moderation could be considered as not undertaken in a diligent manner.[54] There can be two consequences. Either the number of online service providers that will remove illegal content will diminish[55] fearing potential liability for incomplete removal, or the online service providers will exercise excessive removal (including unjustified removal of legal content)[56] to avoid liability. The DSA strives to achieve a balance between encouraging the removal of content whose illegality is apparent[57] (without losing the liability exemptions)[58] and applying safeguards to prevent the arbitrary removal of legal content.[59]

22 On the other hand, raising the bar too high for

44 Commission notice-Guidance on the interpretation and application of Council Directive 93/13/EEC on unfair terms in consumer contracts, (2019/C 323/04), sub-paragraph 3.4.1.

45 See Guidance on the Implementation/Application of the Directive 2005/29/EC on Unfair Commercial Practices, COM(2016) 320 final (*"The notion of 'professional diligence' encompasses principles which were already well-established in the laws of the Member States before the adoption of the UCPD, such as 'honest market practice', 'good faith' and 'good market practice'. These principles emphasise normative values that apply in the specific field of business activity"*), pages 50-51.

46 DSA, Recitals 22 and Recital 53.

47 AG Opinion in *YouTube and Cyando*, *supra* note 30, para. 187.

48 DSA, Recital 53. See also Judgment of 22 June 2021, *YouTube and Cyando*, para.116.

49 AG Opinion *in YouTube and Cyando*, *supra* note 30, para.185.

50 *Ibid.,* para.194 (*"Removing information requires the service provider to react (diligently) to a notification"*).

51 Folkert Wilman, 'Between preservation and clarification, the evolution of the DSA's liability rules in light of the CJEU's case law' (citing YouTube judgment), (*Verfassungblog*, 2 November 2022) <https://verfassungsblog.de/dsa-preservation-clarification/>, accessed 16 March 2023.

52 See also Kuczerawy, *supra* note 38 *("Could unsuccessful voluntary actions be considered as not undertaken in a "diligent manner"? Could it actually discourage hosts from taking one-time voluntary decisions in particular cases if no coherent framework for 'diligence' is in place?"*).

53 This example is based on a similar example provided by Aleksandra Kuczerawy, *supra* note 38 (*"if a moderator trained to review for one type illegality (e.g. incitement to violence) looked at a video, but failed to recognize that it contained another type (e.g. defamation)"*).

54 *Ibid*.

55 Jan M. Smits, discussing good Samaritan's liability for non-rescue, 'The good Samaritan in European private law: on the perils of principles without a programme and a programme for the future' (2000), <https://doi.org/10.26481/spe.20000519js>, accessed 16 March 2024.

56 See also Wilman, *supra* note 51. ("*Even when sincerely meant to tackle illegal content, they* [the measures] *can cause considerable damage if not enacted diligently. For instance, the large-scale removal of content that is wrongly considered illegal comes to mind"*).

57 See also *infra* note 64 and the accompanying text.

58 See DSA, Recital 22. A diligent operator who becomes aware (through own-initiative investigations or third-party notices) of content whose character is clearly ilegal can continue to benefit from the exemptions from liability if it takes immediate action to address it.

59 See supra *notes* 40 and 41 and *infra* notes 66 and 67 and the accompanying text.

Article 7 of the DSA would (i) defeat its own purpose[60] and (ii) would be contrary to Article 8 of the DSA which prohibits a general monitoring obligation. First, the objective of introducing Article 7 is to encourage online service providers to moderate content without fear of losing the liability exemptions. Requiring absolute accuracy in moderating content would discourage them from doing so and would be contrary to the objective of introducing such provision. Second, as explained in Recital 30 of the DSA, *"nothing in this Regulation should be construed [...] as a general obligation for providers to take proactive measures in relation to illegal content"*.[61] More importantly, it confirms that online service providers should not be, neither *de jure* or *de facto*, subject to a monitoring obligation except in a specific case or when faced with an injunction, as interpreted by the CJEU.[62] In that respect, it addresses the concerns raised on how to reconcile the prohibition on general monitoring with proactive measures and how to distinguish between general and specific monitoring obligations.[63]

23 Notwithstanding these opposing approaches, it is worth noting that failing to remove content equates to leaving content up or continuing to host third-party illegal content. Article 6 of the DSA exempts online service providers from liability for hosting or leaving up illegal third-party content provided that they do not have knowledge of the content whose illegality is apparent[64] or upon obtaining knowledge, they expeditiously remove or disable access to that content. The natural train of thought would lead us to conclude that the apparent illegality makes the online service provider knowledgeable which in return allows it to remove that content, thereby acting diligently. Knowledge, which stems from apparent illegality, is necessary to allow good faith and diligent content moderation. In fact, Article 7 applies to online service providers that remove in good faith third-party content they *know* it is illegal, while Article 6(a) applies to online service providers that *unknowingly* host third-party illegal content.

4. Fundamental rights

24 Third, pursuant to Recital 22 of the DSA, liability exemptions for voluntary content moderation measures are conditional upon online service providers acting *"in the observance of the fundamental rights of the recipients of the service, including the right to freedom of expression and of information"*[65] when they expeditiously remove illegal content upon obtaining actual knowledge. The intention here is that online service providers *"avoid that removal and disabling measures affect legal and protected speech"*[66] and a balance is achieved between fighting illegal content and users' rights to freedom of expression and information.[67] The protection of freedom of expression and of information is further reinforced by the transparency requirements in Article 15 of the DSA according to which online service providers should make publicly available reports which include meaningful and comprehensible information about the content moderation engaged in at their own initiative.[68]

II. United States

25 In the US, Section 230 of the CDA provides limited federal immunity[69] to providers and users of interactive computer service,[70] protecting them from

60 Wilman, *supra* note 51.

61 See also Cases C70/10 *Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM)* (2011) and C-360/10 *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV* (2012), where an injunction would require the service provider to carry out a general monitoring obligation contrary to Article 15 of the Directive.

62 DSA, Recital 30.

63 See for instance Thomas Riis and Sebastian Felix Schwemer, 'Leaving the European Safe Harbor, Sailing towards Algorithmic Content Regulation', University of Copenhagen Faculty of Law Legal Studies Research Paper Series, paper no. 2019-64.

64 DSA, Article 6 (*"does not have knowledge [...] or is not aware of facts or circumstances from which the illegal activity or the illegal content is apparent"*).

65 *Ibid*., Recital 22.

66 Joan Barata, 'Digital Services Act and the Protection of Fundamental Freedoms-Recommendations for the trialogue process', (*Digital Services Act Observatory*, 11 April 2022), <https://dsa-observatory.eu/2022/04/11/digital-services-act-and-the-protection-of-fundamental-freedoms-recommendations-for-the-trilogue-process/>, accessed 16 March 2024.

67 See also *You Tube and Cyando*, para. 116.

68 DSA, Article 15(c) and Recital 66.

69 It does not apply to federal criminal law, intellectual property law, any state law "consistent" with Section 230, certain privacy laws applicable to electronic communications, or certain federal and state laws relating to sex trafficking.

70 Interactive computer service means *"any information service, system, or access software provider that provides or enables*

liability for content provided by third parties. The CDA, part of the Telecommunications Act of 1996, originally endeavoured to protect children from indecency and obscene material online.[71] However, the Supreme Court of the US (SCOTUS)[72] struck down the CDA almost entirely for being unconstitutional and violating the First Amendment[73] on freedom of speech.[74] Section 230, which was introduced as a free-standing bill[75] that promoted speech online while encouraging moderation and removal of obscene content,[76] was allowed to stand. Section 230 contains two different immunities listed under Section 230(c) under the title 'Protection of "Good Samaritan" blocking and screening of offensive material'. The first one is Section 230(c)(1) on 'treatment of publisher or speaker' and the second one is Section 230(c)(2) on 'civil liability'.

26 To understand how Section 230 emerged and was eventually enacted, it is useful to look at the influence of the two court cases, *Cubby, Inc. v. CompuServe, Inc.*[77] *('Cubby')* and *Stratton Oakmont, Inc. v. Prodigy Services Co.*[78] *('Stratton Oakmont').* These cases dealt with the issue of whether interactive computer service providers (from now on 'online service providers') could be held liable for defamatory third-party content hosted on their servers.[79]

27 In *Cubby,* the online service provider was considered merely a distributor, rather than a publisher of that information and therefore was not held liable for content hosted on its server. The defendant, CompuServe was the owner of an electronic library consisting of different interest forums, one of which was Rumorville, a daily newsletter that was reporting on broadcast journalism.[80] Rumorville was sued for defamation by Cubby who asserted that CompuServe should be held liable as a publisher of the content posted by Rumorville. The Southern District Court of New York disagreed and held that CompuServe would only be liable if it had knowledge of such defamatory content, therefore creating a notice and take down standard for defamation cases.[81]

28 In *Stratton Oakmont,* the online service provider was considered a publisher because it exercised editorial control, including by removing offensive content from its bulletin boards. It was therefore held liable for content hosted on its server. In reaching that conclusion, the SCOTUS held that Prodigy maintained control over the content by means of an automatic screening program in accordance with company guidelines that 'board leaders' were required to enforce.[82] Prodigy explained that it did not screen material on the bulletin boards, but rather screened and blocked postings containing 'the seven dirty words' and their equivalents in major languages (the so-called 'George Carlin screener').[83] Consequently, some postings, such as calling *"someone a piece of a*

computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions" (47 U.S.C. Section 230 (f)(2)).

71 Emine Ozge Yildirim, 'CDA §230: The Section Behind the Internet Boom', Georgetown University Law Center, (2017).

72 SCOTUS is the highest court in the federal judiciary of the United States. It has ultimate appellate jurisdiction over all federal court cases, and over state court cases that involve a point of U.S. Constitutional or federal law. Source: Wikipedia.

73 The First Amendment protects freedom of speech, the press, assembly, and the right to petition the Government for a redress of grievances, < https://www.whitehouse.gov/about-the-white-house/our-government/the-constitution/>, accessed 16 March 2024.

74 *Reno v. American Civil Liberties Union*, 521 U.S. 844 (1997).

75 Christopher Cox (Former United States Representative and co-author of Section 230), 'Section 230: A Retrospective' (*The Center for Growth and Opportunity at Utah State University*, 10 November 2022), <https://www.thecgo.org/research/section-230-a-retrospective/>, accessed 16 March 2024.

76 Jason Kelley, 'Section 230 is Good, Actually', (*Electronic Frontier Foundation*, 3 December 2020) <https://www.eff.org/deeplinks/2020/12/section-230-good-actually>, accessed 16 March 2024.

77 *Cubby, Inc. v. CompuServe Inc.*, 776 F. Supp. 135 (S.D.N.Y. 1991).

78 *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 1995 WL 323710 (N.Y. Sup. Ct. 1995).

79 Electronic Frontier Foundation, Section 230 Legislative History <https://www.eff.org/issues/cda230/legislative-history#:~:text=Cubby%20and%20Stratton%20Oakmont&text=v.,be%20held%20responsible%20for%20it.>, accessed 16 March 2024.

80 Josh Slovin, 'Section 230 of the Communications Decency Act: The "Good Samaritan" Law which Grants Immunity to "Bad Samaritans"' (2022) in *73*(2) Mercer Law Review.

81 Eric Goldman, 'An Overview of the United States' Section 230 Internet Immunity', in Giancarlo Frosio (ed) *Oxford Handbook of Online Intermediary Liability* (OUP 2020).

82 Marc Jacobson, (Vice President, General Counsel for Prodigy), 'Prodigy: It May Be Many Things To Many People, But It Is Not A Publisher For Purposes Of Libel, And Other Opinions' (1996), *3*(11) Journal of Civil Rights and Economic Development.

83 *Ibid.*

Shitake mushroom" were not screened due to being separate words.[84] The Court eventually held that by actively deleting notes from its bulletin boards on the basis of 'offensiveness' and 'bad taste', Prodigy exercised editorial control over the content and was therefore considered a publisher.[85] The Court drew a distinction between *Stratton Oakmont* and *Cubby* emphasising that CompuServe, unlike Prodigy, lacked the opportunity to monitor information on its website.

29 The *Stratton Oakmont* decision led to what is now known as 'the moderator's dilemma', pushing online service providers to choose between removing content (and potentially being treated as publishers and held liable for third-party content) and not removing content and thereby avoiding liability.[86] To address this issue, Section 230(c) was introduced as an amendment to the CDA and overruled *Stratton Oakmont* by establishing two key rules.[87]

30 The first one, Section 230(c)(1), specifies that a provider or user of an interactive computer service may not be treated as publisher or speaker of any content provided by another information content provider.[88] Section 230 defines the information content provider as *"any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service".*[89] When interpreting Section 230(c)(1), Courts employ either (i) a presumption standard, or (ii) a three-part test.[90] Under the presumption standard, a service provider incurrs no liability for third-party content unless it actively contributes to the development of the content.[91] The test for determining whether an online service provider benefits from liability protection under Section 230(c)(1) *"is whether the service provider developed the content that is the basis for liability".*[92] If it is found not to be a publisher, it can lead to online service providers being offered protection in situations where they negligently fail[93,94] or chose not to remove content from their websites even upon notification.[95] Under the three-step test (i) the defendant must be a provider or user of an interactive computer service, (ii) the defendant must not be an information content provider, and (iii) the plaintiff' claims must seek to treat the defendant as a publisher or speaker of the content.[96]

31 The second one, Section 230(c)(2) concerns civil liability, and it consists of two sub-paragraphs. The first sub-paragraph (A) deals with voluntarily removing or restricting access in good faith to objectionable material,[97] and the second sub-paragraph (B) deals with action taken by online service providers to provide users (or content

84 *Ibid.*

85 Yildirim, *o.c.*

86 Goldman, *o.c. supra* note 81.

87 Statement of Justice Thomas, Supreme Court of the United States, *Malwarebytes, Inc. v. Enigma Software Group USA, LLC*, On Petition For Writ of Certiorari to the United States Court of Appeals for the Ninth Circuit, 13 October 2020.

88 47 U.S.C., Section 230(c)(1): *"No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider".*

89 47 U.S.C, Section 230(f)(3).

90 Yaffa A. Meeran, 'As Justice So Requires: Making the Case for a Limited Reading of § 230 of the Communications Decency Act' (2018), 86 Geo. Wash. L. Rev. 257, 267.

91 *Ibid.*

92 Congressional Research Service ('CRS'), 'Section 230: An Overview,' (2021).

93 For instance, in *Zeran v. America Online, Inc.*, 129 F.3d 327 (4th Cir. 1997), Zeran brought negligence claims against America Online (AOL) as AOL *"had a duty to remove the defamatory posting promptly, to notify its subscribers of the message's false nature, and to effectively screen future defamatory material".* In *Doe v. MySpace, Inc.*, 528 F.3d 413 (5th Cir. 2008), Doe brought negligence claims against MySpace from failing to implement *"basic safety measures to prevent sexual predators from communicating with minors on its Web site".*

94 Negligence has been interpreted by US Courts as failure to investigate and remove a defamatory statement. See *Zeran v. AOL o.c.* (*"Publication does not only describe the choice by an author to include certain information. In addition, both the negligent communication of a defamatory statement and the failure to remove such a statement when first communicated by another party— each alleged by Zeran here under a negligence label—constitute publication."*). See also Amicus Brief in *Reynaldo Gonzalez, et al., Petitioners v. Google LLC*, No. 21-1333 (*"publication is an element of the tort of defamation that encompasses all "communication intentionally or by a negligent act to one other than the person defamed."*).

95 *Zeran v. AOL., o.c.* (*"Liability upon notice would defeat the dual purposes advanced by § 230 of the CDA"*).

96 CRS, *supra* note 92, Johnson and Castro, *o.c.*, Meeran, *o.c.*

97 47 U.S.C., Section 230(c)(2)(A) (*"no provider or user of an interactive computer service shall be held liable on account of any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected"*).

providers) with technical tools to restrict access to content, described in first sub-paragraph (A).[98] These tools refer to *"blocking and filtering technologies that empower parents to restrict their children's access to objectionable or inappropriate online material"*.[99]

32 Compared to Section 230(c)(1), Section 230(c)(2)(A) immunity is narrower as it applies only to good-faith content moderation of *certain type* of third-party objectionable content.[100] In contrast, Section 230(c)(1) is broader as it applies to liability for publishing *any* third-party content and the test employed by most Courts[101] is whether it has materially contributed to the development of the content. The general rule is that when it comes to litigation, if the service provider shows that it did not act as a speaker or publisher of the content, the Courts will not investigate whether it is immune under Section 230(c)(2)(A). If it is established that the service provider acted as publisher or speaker of the content, it can still enjoy immunity under Section 230(c)(2)(A) if it shows that it took down in good faith third-party content that the provider or the user considered objectionable.[102]

33 Even though both sections come under the heading of Good Samaritan principle, only Section 230(c)(2) sub-paragraph (A) (hereafter "Section 230(c)(2)(A)") would qualify for a Good Samaritan provision.[103] This is because Section 230(c)(2)(A) requires a voluntary *action* to restrict access to objectionable material in *good faith*[104] and therefore a duty of care, whereas Section 230(c)(1) requires *no action*. Additionally, some scholars and Court of Appeals claim that only Section 230(c)(2) confers immunity, being the only section under the heading "civil liability".[105]

34 The primary objective of this research paper is to conduct a comparative analysis between the liability shields afforded to online service providers that engage in voluntary good faith content moderation measures under Section 230(c)(2)(A) and Article 7 of DSA. Nonetheless, since US Courts have read Section 230(c)(1) to apply to removal and content moderation decisions,[106] an analysis of Section 230(c)(1) is necessary.

1. Voluntary content moderation measures under US Section 230

35 The Good Samaritan principle reflected in Section 230(c)(2)(A) immunizes interactive computer service providers and users in situations where they voluntarily take any action to remove illegal or objectionable content subject to the good faith safeguard. The provision reads as follows: *"no provider or user of an interactive computer service shall be held liable on account of (a) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected;"*[107]

36 A reading of the provision suggest that Section 230(c)(2)(A) has two limitations. Particularly, it requires the finding that the voluntary actions were taken in good faith and that the content removed is objectionable. Neither the term 'good faith' nor the term 'objectionable' have a statutory definition.

37 The US Courts have generally defined good faith in terms of what it is not considered good faith.[108] Other definitions refer to acting in good faith belief that the content requires moderation or making a good faith effort to moderate objectionable content.[109]

38 While Section 230(c)(2)(A) enumerates a specific type of content that is obscene, lewd, lascivious,

98 47 U.S.C., Section 230(c)(2)(B) (*"no provider or user of an interactive computer service shall be held liable on account of any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1)"*).

99 47 U.S.C., Section 230(b)(4).

100 CRS, *supra* note 92.

101 According to Meeran, *o.c.*, these concern the Fourth, Sixth, and Ninth Circuits. The three-steps test has been employed by the Tenth Circuit in *FTC v. Accusearch.*

102 Ian C. Ballon, 'The Good Samaritan Exemption-Section 230 of the CDA', Excerpted from Chapter 37 (Defamation, Torts and the Good Samaritan Exemption (47 U.S.C.A. § 230)) from *E-Commerce and Internet Law: Legal Treatise with Forms,* 2d Edition (2021).

103 Mary Anne Franks, 'Reforming Section 230 and Platform Liability', *Stanford Cyber Policy Center* (2021).

104 Ballon, *o.c.*

105 See Shlomo Klapper, 'Reading Section 230', *Buffalo Law review*, Volume 70, No.4, (2022), page 1304; See also Meeran, *o.c.*

106 Adam Candeub, 'Reading Section 230 as Written', *Journal of Free Speech Law*, <https://www.journaloffreespeechlaw.org/candeub.pdf >, accessed 16 March 2024.

107 47 U.S.C., Section 230(c)(2)(A).

108 See *infra* notes 200-202.

109 Klapper, *o.c.*, page 1304.

filthy, excessively violent, and/or harassing, it also mentions content that the provider or user finds 'otherwise objectionable'. The term 'otherwise objectionable' has been interpreted by some Courts either as content similar to the content enumerated before under the principle of *ejusdem generis*[110] or as a broad concept[111] encompassing any content the user or provider finds objectionable.[112] Under the *ejusdem generis* principle, otherwise objectionable material should relate to content similar to pornography, violence, or harassment.[113] If the *ejusdem generis* principle is applied, *objectionable* content becomes an objective standard and therefore excludes political viewpoints.[114]

III. Analysis

39 Both Article 7 of the DSA and Section 230(c)(2)(A) provide (eligibility) for liability exemptions for online service providers engaging in voluntarily good faith measures to remove or restrict access to third-party illegal or objectionable content.

40 Under Article 7 of the DSA, online service providers that engage in voluntary content moderation measures are eligible for liability exemptions. To be eligible for liability exemptions, voluntary content moderations measures must be undertaken: i) in good faith, meaning in an objective, non-discriminatory and proportionate manner, with due regards to the rights and legitimate interests of all parties involved[115], ii) in a diligent manner, ensuring the removal of content whose illegal character is apparent,[116] and iii) in the observance of fundamental rights, such as the freedom of expression and of information.[117] Provided that these conditions are met, online service providers become eligible for liability exemptions. To be exempted from liability, online service providers, upon obtaining actual knowledge or awareness of illegal content, must act expeditiously to remove or disable access to it.[118] Under Section 230(c)(2)(A), online service providers that engage in voluntary content moderation measures are exempted from liability under the safeguard of good faith and provided that the content removed is objectionable.

41 A couple of similarities on liability exemptions in the two jurisdictions can be observed from the text of the provisions. First, both Article 7 of the DSA and Section 230(c)(2)(A) require good faith voluntary *action*. Section 230(c)(2)(A) guarantees the liability exemptions when proactively taking *action* to remove objectionable content. Article 7 of the DSA guarantees that the liability exemptions are not lost and that online service providers can still benefit from immunity for their *actions* to remove content subject to the conditions in Article 6 of the DSA. This creates a nexus between the action of moderating content and the liability exemptions.[119]

42 Second, liability exemptions in both jurisdictions constitute rules,[120] whereas the good faith requirements for content moderation constitute standards.[121] The distinction between rules and standards is that the rules constrain the discretion of judges, whereas the standards leave a lot of discretion to judges when interpreting those provisions.[122] The standard of good faith, being an open term, is or can

110 *National Numismatic Certification, LLC. v. eBay, Inc.*, No. 6:08-CV-42-ORL-19GJK, 2008 WL 2704404, (M.D. Fla. July 8, 2008), *Song Fi, Inc. v. Google, Inc.*, 2015 WL 3624335 (N.D. Cal. June 10, 2015). The principle was also acknolwdged in *Holomaxx Techs. v. Microsoft Corp.*, 783 F. Supp. 2d 1097, 1104 (N.D. Cal. 2011), *Zango, Inc. v. Kaspersky Lab, Inc.*, 568 F.3d 1169 (9th Cir. 2009) (Judge Fisher concurring opinion).

111 *Burns v. United States*, 501 U.S. 129, 138 (1991), *Gollust v. Mendell*, 501 U.S. 115, 126 (1991).

112 Candeub, *o.c.*

113 Eric Goldman, 'Online User Account Termination and 47 U.S.C. § 230(c)(2)', 2 U.C. Irvine L. Rev. 659 (2012).

114 See also Klapper, *o.c.*, page 1296 (quoting Rep. Christopher Cox: *"Nor is Section 230 immunity automatically provided on account of moderation or curation policies that restrict access to or availability of content on the basis of political viewpoint"*).

115 DSA, Article 7 and Recital 26.

116 *Ibid.* See also *supra* notes 46 and 47.

117 *Ibid.*, Recitals 22 and 26.

118 *Ibid.*, Article 6.

119 See Klapper, *o.c.* (*"Section 230(c)(2) immunity applies only to cases where the entity would have otherwise been held liable because of the moderation decisions. The moderation must be essential to the alleged liability; it cannot be incidental*").

120 See also Eric Goldman, 'Why Section 230 Is Better Than the First Amendment' (2019) in 95(1) *Notre Dame Law Review Reflection* (*"Section 230 is like a rule; First Amendment defenses are like standards"*).

121 Martijn W. Hesseling, 'The Concept of Good Faith', in A.S. Hartkamp, E.H. Hondius, M.W. Hesselink, C.E. du Perron & M. Veldman (eds.) *Towards a European Civil Code*, 4th rev. and exp. ed. (pp. 619-649) (KLI, 2011).

122 William Fisher, CopyrightX Lecture Transcripts, <https://copyx.org/lectures/>, accessed 16 March 2024.

be subject to the interpretation of judges in both jurisdictions. As the rules on liability exemptions and content moderation measures will have time to mature, the interpretation by the judges of the good faith requirement will convert into guidelines.[123]

43 As to differences, it can be noted that first, both rules on liability exemptions have a test, but the test differs quite significantly. In the EU, the test is whether the online service provider has actual knowledge or awareness of illegal content. If the test is positive, the online service provider can continue to benefit from liability exemptions provided that it acts expeditiously to remove the specific illegal content. In the US, the test is whether the online service provider has materially contributed to the development of content (i.e. whether it is a publisher of the content) ((Section 230(c)(1)). If the test is positive, the online service provider can continue to benefit from the liability exemptions provided that it voluntary removes in good faith objectionable material (Section 230(c)(2)(A)).

44 Second, liability exemptions for voluntary good faith content moderation in Article 7 of the DSA are specifically related to illegal content,[124] whereas immunity for voluntary good faith content moderation in Section 230(c)(2)(A) pertains to a specific type of content that the provider or user finds objectionable. Based on the interpretation of objectionable content, it may refer to either content similar to pornography, violence, obscenity or harassment under the *ejusdem generis* canon, or to anything that service or provider finds objectionable under a broad interpretation. The latter means that objectionable content may include both illegal and legal but harmful content.

45 Third, the rules for liability exemptions under the DSA are subject to other safeguards which are not explicitly mentioned in Section 230. These relate to the principles of proportionality and non-discrimination attached to the good faith standard, the protection of fundamental rights and the diligence requirement. Nonetheless, as discussed in Chapter C II of this research paper, the US Courts interpretation of Section 230(c) suggests that some of these principles, although not explicitly mentioned in the provision, are embedded in the objectives of enacting Section 230, as well as in the definition pertaining to good faith and objectionable content.

123 *Ibid.* On the standard of fair use in US copyright law: *"over time the courts have tacitly subdivided the universe of cases implicating colorable fair use defenses into subfields and have converged on guidelines concerning how the four factors will be interpreted in each subfield".*

124 DSA, Recital 17 (*"the exemptions from liability established in this Regulation should apply in respect of any type of liability as regards any type of illegal content, irrespective of the precise subject matter or nature of those laws."*). See also 'Questions and answers on the Digital Services Act' (*"The new rules only impose measures to remove or encourage removal of illegal content, in full respect of the freedom of expression"*), <https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348>, accessed 15 March 2024.

C. Relevant jurisprudence

46 This chapter provides an overview of the applicable legal precedents concerning the establishment of liability exemptions for online service providers in relation to voluntary content moderation actions in both jurisdictions. Since the Directive on electronic commerce does not foresee voluntary content moderation measures and the DSA has only been recently adopted, there is a scarcity of legal cases addressing specifically liability exemptions and voluntary content moderation measures. In contrast, Section 230 has been in existence for over twenty years, leading to a substantial body of case law that aids in interpreting the relevant provisions. Nevertheless, the Good Samaritan principle under Section 230(c)(2) has been litigated less than Section 230(c)(1) following the Court's decision in *Zeran v. AOL* to treat removal decisions under Section 230(c)(1) instead of under Section 230(c)(2).[125]

I. European Union

47 The DSA confirms the case law of the CJEU on liability exemptions under the Directive and brings clarity to certain elements regarding liability exemptions for online service providers. The cases mostly deal with knowledge and awareness of illegal content, as well as the nature of the service provider (active or passive) that would determine whether the online service provider is exempted from liability.

1. Knowledge and take-down

48 The situations in which the online service providers become knowledgeable or aware of the illegal content as a result of both own-initiative investigations and notices by third parties have been examined by the CJEU in *L'Oréal v eBay*[126] and are reflected in

125 Klapper, *o.c.,* fn 142 and citing *Domen v. Vimeo, Inc.*, 433 F. Supp. 3d 592, 601 (S.D.N.Y. 2020), *Riggs v. MySpace*, Inc., 444 F. App'x 986, 987 (9th Cir. 2011), *Ebeid v. Facebook, Inc.*, No. 18-cv-07030, 2019 WL 2059662, at *5 (N.D. Cal. May 9, 2019).

126 C324/09 *L'Oréal SA v eBay International AG* [2011], para. 122.

Recital 22 of the DSA. In the same case, the CJEU defined the notion of knowledge as knowledge that results from information which is sufficiently and adequately substantiated.[127] Recital 53 parallels this notion by stating that a notice should contain sufficient information to enable a diligent provider of hosting services to identify, without a detailed legal examination, that it is clear that the content is illegal. *L'Oréal v eBay* judgment also brings clarity of what it means to play an active role of such a kind to give it knowledge, such as providing assistance which entails, in particular, optimizing the presentation of the offers for sale in question or promoting them.[128] Furthermore, the role played by the active nature of online service providers in determining liability exemptions, which were confirmed by the CJEU in *Google France and Google*[129] and *L'Oréal v eBay*[130], are reproduced in Recital 18 of the DSA.

49 The most recent case, *YouTube and Cyando*[131] which was adopted after the DSA proposal and related to liability of online service providers for hosting copyright infringing material, provides elements in relation to voluntary content moderation measures and liability exemptions which are also found in the DSA. Thus, the CJEU conclusion that automatic indexing, search function and recommending information does not lead to liability is reproduced in Recital 22 of the DSA. Furthermore, the fact that the online service provider *"is aware, in a general sense of the fact that its service provider is also used to share content which may infringe intellectual property rights"*[132] does not constitute actual knowledge or awareness, as well as the fact that actual knowledge refers to specific content,[133] is included in Recital 22 of the DSA. The CJEU further explained that implementing technological measures aimed at detecting and ending copyright infringing material does not mean that the online service provider plays an active role.[134] In other words, the CJEU conclusions in this respect, on one hand incentivised service providers to undertake voluntary measures, and on the other hand, confirms that such measures do not render the service provider active and therefore aware of the illegal content.

50 Finally, the DSA aligns with the CJEU ruling in *Eva Glawischnig v Facebook,*[135] by establishing that the absence of a general monitoring obligation does not mean that online service providers do not have an obligation to monitor in a specific case or when faced with an injunction from national authorities.[136]

2. Good faith and diligence

51 With regards to the *good faith* safeguard, more specifically acting in an objective, non-discriminatory and proportionate manner, with due regards to the rights and legitimate interests of all the parties involved, there is limited case law on the nexus between this safeguard and (voluntary) content moderation measures. Thus, these elements will have to be assessed on a case-by-case basis and interpreted accordingly by the CJEU once the rules on liability exemptions for voluntary content moderation measures in the DSA have matured. The *objective* requirement, if given a literal interpretation of the definition, can be understood as acting based on facts and can be closely related to the principle of non-discrimination.[137] The *non-discriminatory* requirement can be understood as a condition that the online service providers, when engaging in voluntary content moderation measures to remove or disable access to illegal content, do not discriminate based on speaker, the content of his/her message or other characteristics.[138] Such an interpretation reflects the non-discrimination principle in the Charter of Fundamental Rights of

127 *Ibid.*

128 *Ibid.*, para.123.

129 Joined Cases C-236/08 to C-238/08 *Google France SARL and Google Inc. v Louis Vuitton Malletier SA and Others* [2010], para. 114 (except that the DSA no longer refers to the passive nature, but rather active nature of the online service provider).

130 *L'Oréal v eBay*, para.113.

131 Joined Cases C682/18 and C683/18 *Frank Peterson v Google LLC and Elsevier Inc. v Cyando AG* [2021], para.114.

132 *Ibid.*, para. 111.

133 *Ibid.*, para.113.

134 *Ibid.*, paras. 94 and 109.

135 Case C18/18 *Eva Glawischnig Piesczek v. Facebook Ireland Limited* [2019].

136 DSA, Recital 30.

137 van de Kerkhof, *o.c.*

138 See also Christoph Busch, 'Platform Responsibility in the European Union' ("*While the DSA does not formulate an explicit requirement of platform neutrality, the reference to the principle of non-discrimination makes it clear that an arbitrary unequal treatment of content within the framework of content moderation would be a violation of the due diligence requirements*") (2022) <https://sites.tufts.edu/digitalplanet/files/2022/12/DD-Report_2-Christoph-Busch-11.30.22.pdf>, accessed 15 March 2024.

the European Union[139] and mentioned in Recital 3 of the DSA.

52 Regarding the *proportionality* requirement, the CJEU, in *Sabam v Netlog,* held that *"an injunction* [requiring the installation of a filtering system] *would result in a serious infringement of the freedom of the hosting service provider to conduct its business since it would require that hosting service provider to install a complicated, costly, permanent computer system at its own expense"*.[140] Therefore, when undertaking voluntary measures, online service providers must ensure that such measures are not excessively costly or burdensome on any of the parties involved and serve the purpose for which they are employed.[141] This reflects the principle of proportionality as defined in the Treaty of the EU as being suitable and necessary to achieve the desired end and not impose a burden on the individual that is excessive in relation to the objective sought to be achieved.[142]

53 The proportionality principle in Recital 26 of the DSA is intrinsically linked to the legitimate interests of both the recipients of the service and the service providers themselves. It entails, as indicated in Recital 22 of the DSA, the rights of all parties involved, not only the rights of the recipients of the service. Thus, in relation to the rights of the service providers, the *Sabam v Netlog* case tackles the principle of proportionality by looking at the service provider's freedom to conduct business, which is also the meaning given to proportionality in Article 17(5) of the Directive 2019/790[143] and Article 3 of Directive 2004/48 (such measures "*shall not be unnecessarily complicated or costly, or entail unreasonable time-limits or unwarranted delays"*).[144]

54 It is noteworthy that the *Sabam v Netlog* judgment prompted the CJEU to assert that requiring an online service provider to implement a filtering system would force them to actively monitor all user data, a practice prohibited by Article 15 of the Directive (now Article 8 of the DSA). Additionally, such a broad monitoring obligation would be inconsistent with Article 3 of Directive 2004/48,[145] which states that the measures referred to by the directive must be fair, proportionate, and not excessively costly.[146] Similar conclusions were reached in *Scarlet Extended*.[147] Recital 26 of the DSA confirms that online service providers can employ automated filtering tools when engaging in voluntary content moderation, provided that they do so diligently and minimize the rate of errors. Moreover, Article 7 of the DSA which guarantees eligibility for liability exemptions for voluntary content moderation, and the Commission's confirmation[148] that online service providers can maintain their liability exemptions if they promptly remove the illegal content, seem to suggest that online service providers can still benefit from liability exemptions when undertaking voluntary measures. Nonetheless, reconciling general monitoring obligation with voluntary content moderation remains challenging.

55 With regards to the *diligence* safeguard, although not dealt by the CJEU, but by the ECtHR, the *Delfi v Estonia*[149] case offers insights into undertaking (voluntary) content moderation measures in a diligent manner. The case concerned the liability of Delfi, an Internet news portal, for defamation

139 EU Charter of Fundamental Rights, Article 21.

140 *SABAM v Netlog NV*, para. 46.

141 The Directive 2017/541 on combating terrorism confirms this view: "*measures or removal and blocking* [of online content constituting a public provocation to commit a terrorist offence] *are limited to what is necessary and proportionate and that users are informed of the reason for those measures"*, Article 21 (3).

142 Article 5(4) of the Treaty on European Union.

143 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, Article 17(5) and Recital 66. See also Advocate General Saugmandsgaard Øe in C-401/19 *Republic of Poland v European Parliament, Council of the European Union* [2021], para. 156 (*"Article 17(5) of Directive 2019/790, [...] states that the measures to be taken by each supplier must be assessed, in the light of the principle of proportionality, with regard to factors such as the 'size of the service' or the 'cost' of available tools, seems to me to be more relevant to the question of compliance with the freedom to conduct a business, which is not the subject of the present case, than to freedom of expression"*).

144 Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights (OJ L 157, 30.4.2004). This was confirmed by Advocate General Cruz Villalon in C-314/12 *UPC Telekabel Wien GmbH v Constantin Film Verleih GmbH and Wega Filmproduktionsgesellschaft GmbH* [2014] para.79 (*"the measure is neither 'fair and equitable' nor 'proportionate' within the meaning of Article 3 of Directive 2004/48"*).

145 Directive 2004/48/EC, Article 3.

146 *Sabam v Netlog*, paras. 34 and 38.

147 *Scarlet Extended*, paras. 36 and 40.

148 Communication from the Commission COM(2017) 555, *supra* note 18.

149 *Delfi AS. v. Estonia*, App. nr. 64569/09 (European Court of Human Rights (Grand Chamber), 16 June 2015).

in its capacity as publisher[150] of the (anonymous) comments posted on its portal, despite the fact that it promptly removed the illegal comments upon receiving notification.[151] Since Delfi was not an intermediary service it could not rely on the liability exemptions specified in Articles 12 to 15 of the Directive.[152]

56 The Court ruled that the automatic word-based filter employed by Delfi, while it may have removed some defamatory comments, it failed to filter out and detect hate speech and incitement to violence. This failure limited Delfi's ability to expeditiously remove the defamatory comments with the consequence that the clearly illegal comments remained online for six weeks.[153] The Court found that the comments which were related to hate speech and incitement to violence *"did not require any linguistic or legal analysis since the remarks were on their face manifestly unlawful"*.[154] According to the Court, the majority of the comments lacked sophisticated metaphors, hidden meanings, or subtle threats. Instead, they were overt expressions *"of hatred and blatant threats"*.[155] Additionally, the Court noted that the comments did not contain any information that would necessitate excessive verification by the portal operator.[156] Recital 53 of the DSA reflects the same reasoning according to which a notice that contains sufficient information to enable a diligent operator to identify, without a legal detailed examination, the illegality of content gives rise to knowledge or awareness of illegality.[157] This aligns with the concept of apparent illegality as defined by the CJEU in the *L'Oréal v eBay* case, as being *"aware of facts or circumstances on the basis of which a diligent economic operator should have identified the illegality in question"*.[158]

57 The Court's conclusions indicate that if a news portal, such as Delfi, voluntarily implements content moderation measures through automated tools or by establishing a team of moderators to remove illegal content but fails to eliminate all such content, it may be held liable for third-party content, particularly when the content displays apparent illegality, such as hate speech or incitement to violence. Should Delfi have been found to be an intermediary service, the outcome of this ruling on voluntary moderation to remove third-party illegal content remain uncertain. Nevertheless, the CJEU has not yet dealt with a case similar to Delfi[159] to allow us to draw parallels with it.

3. Fundamental rights

58 With regards to the fundamental rights of the recipients of service, which include freedom of expression and of information (Recital 22 of the DSA), the CJEU held in *Scarlet Extended* that the *"filtering system may also infringe the fundamental rights of that ISP's customers, namely their right to protection of their personal data and their freedom to receive or impart information"*.[160] Thus, the unjustified removal of content by automated tools can pose a potential threat to the protection of fundamental rights. By analogy, online service providers, when undertaking voluntary content moderation measures, especially by employing algorithmic filtering tools, must ensure the implementation of necessary safeguards to protect fundamental rights of users, such as freedom to receive and impart information.[161]

59 The CJEU held in *UPC Telekabel* that when complying with an injection, the addressee of that injunction *"must ensure compliance with the fundamental right of internet users to freedom of information"* so that the measures implemented do not affect *"internet users who are using the provider's services in order to lawfully access information"*.[162] The case is illustrative of how

150 Delfi was found to be in control of the comments and thus acting as a media publisher since it i) invited and encouraged comments on its website, ii) economically profited from the number of visits which in turn depended on a number of comments, and iii) set out the rules for the comments section and made changes to it (removed comments) if those rules were breached.

151 *Ibid.*, para. 65. The Grand Chamber ruled that there was no violation of Article 10 of the Convention of Human Rights.

152 *Ibid.*, para. 13.

153 *Ibid.*, para. 156.

154 *Ibid.*, para. 117.

155 *Ibid.*, para. 156.

156 *Ibid.*, para. 16.

157 DSA, Recital 53.

158 *L'Oréal SA v eBay*, para. 120.

159 Liudmila Sivetc, 'Future of Internet Portals After the Case of Delfi', Master thesis (2016), University of Turku.

160 *Scarlet Extended*, para. 50.

161 See for instance Riis and Schwemer, *o.c.* (*"The finding of the CJEU in the Scarlet Extended and Netlog judgments that an order to implement filtering technologies violates Article 15 of the E-Commerce Directive and fundamental rights, in principle, must also be considered applicable to other rules that create an obligation to implement proactive measures"*).

162 *UPC Telekabel*, paras. 55-56.

CJEU interprets the knowledge and take down approach for liability exemption by giving due weight to the rights of users[163] *"whose content may be blocked or removed"*.[164]

II. United States

60 In the US, liability exemptions have only been updated once when Congress enacted in 2018 a law creating a sex trafficking exception[165] to the immunity provided by Section 230. Nonetheless, the breadth of case law available provides details about the judicial interpretation of Section 230(c), in particular subsection (c)(1). With regards to the Good Samaritan rule of Section 230(c)(2)(A), the case law dealt with either determining whether the removal of content was done in good faith or whether the content was indeed objectionable.

1. Knowledge

61 Pre-Section 230 enactment, the *Cubby* case upheld the common law distributor liability according to which distributors are liable for third-party content only if they have actual knowledge of the illegal character of the content.[166] Due to a contradictory ruling in *Stratton Oakmont*, Congress introduced Section 230 to address the issue of the moderator's dilemma.[167]

62 The Fourth Circuit was the first to interpret Section 230 after its enactment in *Zeran v. AOL*. The plaintiff brought negligence claims against AOL as AOL *"had a duty to remove the defamatory posting promptly, [...], and to effectively screen future defamatory material"*.[168] The Fourth Circuit Court asserted that Section 230(c)(1) creates a federal immunity for any cause of action that would make a service provider liable for information originating with a third-party.[169] It thus led to preserving immunity of online service providers in situations where they negligently fail to[170] or chose to not remove content from their websites.[171] Publishing, as well as removal and editing of material are considered basic editorial functions covered by Section 230 and confirmed consistently by the US Courts.[172] Such basic editorial functions do not deem a service provider as publisher or speaker of the content provided by a third-party.

63 Following *Zeran v. AOL*, subsequent decisions have followed the broad interpretation of Section 230(c)(1),[173] providing sweeping immunity to online service providers for any tort action.[174] Under this wide interpretation courts have dismissed lawsuits on a large set of causes of action, including sex trafficking of minors (*Doe v. Backpage*[175]), illegal sale of guns (*Gibson v. Craiglist*[176]), defective sale of products (*Lemmon v. Snap*[177]), the encouragement of terrorist acts (*Force v. Facebook*[178]),[179] and racially discriminatory removal of content (*Sikhs for Justice*

163 Written comments in the case of Delfi AS v. Estonia, No 64569/09, 6 June 2014, <https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2015/06/20140606-Delfi-intervention-FINAL.pdf.>, accessed 16 March 2024.

164 *UPC Telekabel*, para. 57.

165 The Allow States and Victims to Fight Online Sex Trafficking Act of 2017 ("FOSTA").

166 See *supra* note 81.

167 See *supra* note 86.

168 *Zeran v. AOL*.

169 *Ibid.*

170 See also *Doe v. MySpace, Inc.*, 528 F.3d 413 (5th Cir. 2008), Doe brought negligence claims against MySpace from failing to implement *"basic safety measures to prevent sexual predators from communicating with minors on its Web site"*.

171 See *supra* notes 93-95.

172 *Zeran v. AOL, o.c.*, where the Fourth Circuit Court held that Section 230(c)(1) precludes *"lawsuits seeking to hold a service provider liable for its exercise of a publisher's traditional editorial functions-such as deciding whether to publish, withdraw, postpone, or alter content"*. Such basic editorial functions were also noted in *Fair Housing Council of San Fernando Valley v. Roommates.com*, LLC, 521 F.3d 1157 (9th Cir. 2008).

173 Notably First, Third, and Tenth Circuit as indicated in *Force v Facebook*, On a Petition For Writ of Certiorari, enquiring about the meaning of Section 230(c)(1).

174 Michael L. Rustad and Thomas H. Koenig, 'The Case for a CDA Section 230 Notice-and-Take-Down Duty' (2023) 23 NEV. L.J. 533.

175 *Jane Doe No. 1 v. Backpage.com*, LLC, 817 F. 3d 12 (1st Circuit 2016).

176 *Gibson v. Craigslist, Inc.*, 08-CV-7735, 2009 WL 1704355 (S.D.N.Y. June 15, 2009).

177 *Lemmon v. Snap, Inc.*, 440 F. Supp. 3d 1103 (CD Cal. 2020).

178 *Force v. Facebook, Inc.*, 934 F.3d 53 (2nd Cir. 2019).

179 Franks, *o.c.*

v. Facebook[180]).[181] The SCOTUS has recently issued an opinion in favour of Google holding that Section 230 protects YouTube's recommender systems from liability under the anti-terrorism act.[182]

64 The wide consensus on the broad immunity of Section 230(c)(1) is not unanimously shared. The underlying reason of this conflicting approach is that there is a difference between immunizing only traditional functions and immunizing any activity of publishing.[183] Justice Thomas, writing on petition for writ of certiorari[184] in *Malwarebytes v. Enigma* provided a textual analysis of the provision which criticizes the consensus. He explained that Section 230(c)(1) applies when online service providers *unknowingly* leave up illegal third-party content, while 230(c)(2)(A) applies when they take down in good faith certain third-party content.[185] In supporting his argument, he stated that Section 502[186] of the CDA *"makes it a crime to knowingly display obscene material to children, even if a third party created that content".*[187]

What has instead prevailed in the Courts is that Section 230(c)(1) confers immunity even when a company distributes content that it *knows* is illegal as in *Zeran v. AOL.*[188] Courts have narrowly interpreted Section 230(f)(3) which defines a content creator as anyone *"responsible, in whole or in part, for the creation or development"* of the content to cover only extensive edits.[189] Referring to *Barnes v. Yahoo*, which held that *"Subsection (c)(1), by itself, shields from liability all publication decisions, whether to edit, to remove, or to post, with respect to content generated entirely by third parties",*[190] Justice Thomas stated that Courts have restricted the limits Congress placed on removal decisions.[191] His opinion was based on the dissenting opinion of Judge Katzman in *Force v. Facebook* who rejected the notion that Section 230(c)(1) should be construed broadly.[192]

65 The conflicts among the circuits[193] regarding the meaning of section 230(c)(1) have led the Courts of appeals to disagree not only about when section 230(c)(1) exempts service providers from liability, but also about what type of defence it is.[194] A majority of the Courts of appeals follow the *Zeran* reasoning and hold that Section 230(c)(1) immunity applies to any service provider that acts as a publisher of third-party content.[195] Thus, the immunity provided by Section 230(c)(1) depends on the nature of the defendant's *conduct* and whether the *defendant can show* it was acting as a publisher, and if available would apply to *all types of claims.*[196] The Seventh Circuit, on the other hand, holds that section 230(c)(1) does not create a form of immunity at all, but it is

180 *Sikhs for Justice Inc. v. Facebook, Inc.*, 697 Fed. Appx. 526 (9th Circuit 2017).

181 Statement of Justice Thomas in *Malwarebytes v. Enigma, supra* note 87.

182 *Reynaldo Gonzalez, et al., Petitioners v. Google LLC*, On a Petition For Writ of Certiorari to the United States Court of Appeals for the Ninth Circuit, 18 May 2023.

183 *Force v. Facebook*, On a Petition For Writ of Certiorari, enquiring about the meaning of Section 230(c)(1).

184 A Petition for Writ of Certiorari is an appellee's formal request to a state Supreme Court or to the Supreme Court of the United States to review a case for error or violation that occurred in a lower court.

185 Statement of Justice Thomas in *Malwarebytes v. Enigma*, *supra* note 87.

186 See Vincent Dumas, 'Enigma Machines: Deep Learning Algorithms As Information Content Providers Under Section 230 of The Communications Decency Act' (*"The Senate and House introduced two amendments, one from each chamber, as part of a unified CDA: Sections 223 and 230. Section 223 criminalized the transmission of obscene material or harassing communications over the internet"*.) <https://wlr.law.wisc.edu/wp-content/uploads/sites/1263/2023/01/14-F_Dumas-Camera-Ready-1581%E2%80%93-1616-PDF-.pdf>, accessed 16 March 2024, and Danielle K. Citron and Benjamin Wittes, 'The Problem Isn't Just Backpage: Revision Section 230 Immunity' 2 Georgetown Law Technology Review 453 (2018) (*"Section 502 of the final legislation contained the Senate's additions to 47 U.S.C. § 223. Section 509 contained the House's new Section 230*").

187 Statement of Justice Thomas in *Malwarebytes v. Enigma*, *supra* note 87.

188 *Ibid.*

189 *Ibid.*

190 *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1105 (9th Cir. 2009).

191 Statement of Justice Thomas in *Malwarebytes v. Enigma*, *supra* note 87.

192 *Force v. Facebook*, On a Petition For Writ of Certiorari, enquiring about the meaning of Section 230(c)(1).

193 The US Court of Appeals are the intermediate appellate courts of the United States federal judiciary and are divided into 13 Circuits. The US district courts are the trial courts of the US federal judiciary. District courts' decisions are appealed to the US court of appeals for the Circuit in which they reside, except for certain specialized cases that are appealed to the US Court of Appeals for the Federal Circuit or directly to the U.S. Supreme Court. (Source: Wikipedia).

194 *Force v. Facebook*, *supra* note 192.

195 *Ibid.*

196 *Ibid.*

rather a definition.[197] Thus, the defence provided by Section 230(c)(1) is limited to claims which require a *plaintiff to show* that the defendant was a publisher.[198]

2. Good faith

66 Since there is no statutory definition of the term *good faith*, the Courts have given interpretations of what it means when an online service provider does not act in good faith. For instance, plaintiff's claims that the defendant acted under an anticompetitive motive and therefore not in good faith were allowed to proceed[199] in several cases such as *e-ventures Worldwide v Google,*[200] *Spy Phone v Google,*[201] or *Darnaa v Google*.[202]

67 In *e-ventures Worldwide v. Google*, the Middle District Court of Florida denied Google's motion to dismiss under Section 230(c)(2) due to e-ventures presenting sufficient evidence about Google's anticompetitive motivations.[203] The Court asserted that moderation based on anticompetitive motives does not constitute good faith. Google however won the case on the basis that its decision to de-index all of e-ventures' websites so they would no longer appear in Google search results constituted speech protected under the First Amendment.[204]

68 In *Spy Phone v. Google*, the Northern District Court of California examined the good faith covenant by looking at whether Google's decision to remove Spy Phone app for violating its anti-spyware policy was "pretextual" since no such policy existed. The plaintiff's claim was that Google, *"by selling the keywords "Spy Phone" to developers of parental monitoring apps"* and granting priority placement to the purchaser of those keywords for its competitive app in Google Play, placed the plaintiff at a competitive disadvantage.[205] The same Court examined YouTube' decision to remove a video for an inflated view count which allegedly violated its terms of use in *Darnaa v. Google*. The Court found that *"the allegations in the complaint are sufficient to support a claim for contractual breach of the covenant of good faith and fair dealing"*.[206] In reaching that conclusion, the Court explained that YouTube's terms of service regarding its rights to remove and relocate videos were ambiguously drafted.

69 The good faith requirement has been discussed in other cases such as *Jurin v. Google*[207] where the Eastern District Court of California dismissed a breach of the duty of good faith and fair dealing by Google for not adhering to the terms of its Adwords policy. The Court noted that *"good faith and fair dealing is satisfied where the conduct at issue is either expressly permitted or at least not prohibited"*.[208] The Court ruled that Google followed the terms of its policy *"and because this conduct was expressly permitted, good faith is satisfied"*.[209] It further held that *"the implied covenant [of good faith] cannot override express provisions"*.[210] The claims were barred by the immunity provided by Section 230 and the case was quoted by the Northern District Court in *King v Facebook* where King alleged that Facebook removed multiple posts that Facebook considered to be in violation of its terms of use and *"that Facebook treats black activists and their posts differently than it does other groups"*.[211] Because *"each of King's claims against Facebook seeks to hold it liable as a publisher for either removing his posts, blocking his content, or suspending his accounts"*,[212] the Court applied Section 230(c)(1) to dismiss the case.

197 See for instance *Doe v. GTE Corp.*, 347 F.3d 655, 660 (7th Cir. 2003).

198 *Force v. Facebook*, *supra* note 192.

199 CRS, *supra* note 92.

200 *e-ventures Worldwide v. Google*, 2:14-CV-646-FTM-29CM, 2016 U.S. Dist. LEXIS 62855 (M.D. Fla. May 12, 2016).

201 *Spy Phone Labs LLC v. Google Inc.*, 15-CV-03756-KAW, 2016 WL 6025469, at *8 (N.D. Cal. Oct. 14, 2016).

202 *Darnaa LLC v. Google, Inc.*, 2015 WL 7753406 (N.D. Cal. Dec. 2, 2015), Order by Judge Ronald M. Whyte granting 15 Motion to Dismiss with leave to amend.

203 Ashley Johnson and Daniel Castro, 'The Exceptions to Section 230: How Have the Courts Interpreted Section 230?' (*ITIF*, 22 February 2021) <https://itif.org/publications/2021/02/22/exceptions-section-230-how-have-courts-interpreted-section-230/.>, accessed 16 March 2024.

204 *e-ventures Worldwide v. Google, Inc.*, 2:14-cv-00646-PAM-CM (M.D. Fla. Feb. 8, 2017) where the Court held that *"the First Amendment protects these decisions, whether they are fair or unfair, or motivated by profit or altruism"*.

205 *Spy Phone v. Google*.

206 *Darnaa v. Google*.

207 *Daniel Jurin v. Google Inc.*, No. 2:09-cv-03065-MCE-KJM, Memorandum of Order (E.D. Cal. Feb.15, 2011).

208 *Ibid*.

209 *Ibid*.

210 *Ibid*.

211 *King v. Facebook, Inc.*, No 19-cv-01987-WHO (N.D. Cal. Sept. 5, 2019).

212 *Ibid*.

3. Objectionable content

70 Some Courts have interpreted 'otherwise objectionable' broadly because Section 230(c)(2)(A) states that the provider or user is the one who determines whether the content is objectionable.[213] The subjective nature of objectionable content was considered in *e360Insight v. Comcast,*[214] where the Northern District Court of Illinois ruled that commercial unsolicited and bulk email could be deemed objectionable under Section 230(c)(2)(A) and that online service providers are immune from liability when they block content that they subjectively consider to be objectionable.[215] In *Holomaxx v. Yahoo*, the judge for the Northern District Court of California noted that *"no court has articulated specific, objective criteria to be used in assessing whether a provider's subjective determination of what is "objectionable" is protected by [Section] 230(c)(2.)".*[216] The Court eventually acknowledged that the harassing nature of the emails were sufficient to reasonably conclude that the content was objectionable.[217]

71 The Western District Court of Washington in *Zango v. Kaspersky* also considered the subjective nature of 'otherwise objectionable' since it is the provider or the user who determines what content in objectionable.[218] In its concurring opinion for the Ninth Circuit in *Zango v Kaspersky,*[219] Judge Fisher warned that 'otherwise objectionable' may be invoked by a blocking software provider to block content for anticompetitive reasons. The interpretation of 'otherwise objectionable' was ultimately not examined since the plaintiff did not raise it and thus waived it.[220]

72 Blocking for anticompetitive reasons was later addressed by the Ninth Circuit in *Enigma Software v. Malwarebytes*[221] which ruled that Section 230(c)(2)(B) did not apply[222] because objectionable content in Section 230(c)(2)(A) does not include blocking access to content for anticompetitive reasons.[223] By looking at the statute's policy goals to determine whether the competitors' content was objectionable and therefore its removal justified, the Court on appeal held that Section 230 objective is to promote the advancement of tools that maximise user control by granting immunity to *"providers of such tools, such as Malwarebytes, regardless of motive [...] But, to prevent misuse of those tools, [they must restrict content] by acting in good faith".*[224] This perspective implies that what the user or provider considers to be objectionable is not unlimited,[225] but must fall within the specific categories of content which are either enumerated in Section 230(c)(2)(A) or align with the policy goals of Section 230.

73 In *Song Fi v Google*, the Northern District Court denied immunity under Section 230(c)(2) to YouTube for removing a video because its view count was considered by YouTube to have been artificially inflated and thus *"its content violated YouTube's Terms of Service".*[226] The Court did not consider that the inflated view count qualifies as objectionable content as it was not in line with the policy goals

213 CRS, *supra* note 92.

214 *e360Insight, LLC v. Comcast Corp.*, 546 F. Supp. 2d 605 (N.D. Ill. 2008).

215 Ballon, *o.c.*

216 *Holomaxx Technologies v. Yahoo, Inc.*, CV-10-4926-JF (N.D. Cal. March 11, 2011).

217 *Ibid.*

218 *Zango, Inc. v Kaspersky Lab, Inc.*, No. 07-0807-JCC, 2007 WL 5189857, at *4 (W.D. Wash. Aug. 28, 2007).

219 *Zango v. Kaspersky*, *supra* note 110.

220 Eric Goldman, 'Anti-Spyware Company Protected by 47 USC 230(c)(2) - Zango v. Kaspersky (*Technology & Marketing Law Blog*, 26 June 2009) <https://blog.ericgoldman.org/archives/2009/06/antispyware_com.htm.>, accessed 16 March 2024.

221 *Enigma Software v. Malwarebytes,* where, according to the plaintiff, Malwarebytes' decision to block unwanted programs of Enigma served as a 'guise' for anticompetitive conduct. The Court concluded that *"if a provider's basis for objecting to and seeking to block materials is because those materials benefit a competitor, the objection would not fall within any category listed in the statute and the immunity would not apply".*

222 According to Eric Goldman, this decision overruled *Zango v. Kaspersky* which provided immunity to providers of malware software. The difference was that in *Zango v. Kaspersky* the interpretation of 'otherwise objectionable' had not been fully examined because there was no dispute over the objectionable nature of the blocked content, *supra* note 221.

223 Johnson and Castro, *o.c.*

224 *Enigma Software v. Malwarebytes,* No. 5:17-cv-02915-EJD (2019), On Appeal from the United States District Court for the Northern District of California, Petition For Panel Rehearing and Rehearing En Banc.

225 Annemarie Bridy, 'Remediating Social Media: A Layer-Conscious Approach' (2018) Boston University Journal of Science and Technology Law, *24*(193-228).

226 *Song Fi v. Google.*

of the CDA.[227] It therefore construed the term 'otherwise objectionable' more narrowly to avoid allowing online service providers to remove any content they wish.[228]

74 The cases dealing with the interpretation of 'good faith' and 'otherwise objectionable' content suggest that there is an overlap between the two terms. Some Courts interpreted the decision to remove or restrict access to material provided to have been done in good faith by looking at whether the material removed fell under the categories listed in Section 230(c)(2)(A) and was indeed objectionable. As Judge Fisher concurring opinion in *Zango v. Kaspersky* suggests, blocking software for anticompetitive reasons by invoking 'otherwise objectionable' can be considered to be acting in bad faith.[229]

III. Analysis

75 Although *Delfi v. Estonia* was not reviewed by the CJEU, but by ECtHR which does not have a jurisdiction to apply EU law,[230] and *Stratton Oakmont* was reviewed before Section 230 was enacted, these two notable cases are useful for setting the scene regarding the liabilities of news portals[231] for unsuccessful content moderation.

76 Similar to Delfi, Prodigy faced liability for defamatory comments posted by anonymous users on its bulletin board, whether it knew about the content or not. *Stratton Oakmont* and *Delfi v. Estonia* share resemblances in that both Prodigy and Delfi were found to be publishers on the basis of exercising editorial control over illegal content, including efforts to remove such content but failing to remove some of it. Prodigy's content moderation policies, technological measures, and employment of moderators (board leaders) to act as editors for the bulletin boards, contributed to the finding that it is a publisher. This parallel is evident in Delfi's engagement in automatic content filtering, the establishment of rules for the comments section, and the removal of comments, all of which granted it editorial control over the content and resulted in its classification as a publisher. Although the CJEU might have led to opposite conclusions had *Delfi* been an information society service, the case offers some perspectives and anticipates questions about how to determine the circumstances under which an online service provider can lose immunity for unsuccessful content moderation measures (i.e. failing to remove all illegal content).

77 The Courts' interpretation of the immunity available under Section 230(c)(1), as well as what constitutes 'good faith' and 'objectionable' content, provides valuable insights into the similarities and differences regarding the provisions offering immunity to online service providers for hosting and removing content in both jurisdictions.

78 Under a narrow interpretation of Section 230(c), the following similarities regarding the immunity provisions in Section 230(c) and Article 6 and 7 of the DSA can be noted.

79 First, both Section 230(c)(1) and Article 6(1)(a) of the DSA offers immunity to online service providers that *unknowingly* host illegal content on their websites. Under Section 230(c)(1), if the service provider acted as a publisher, it can still benefit from immunity under Section 230(c)(2)(A) by voluntary removing the content in good faith. Under Article 6 of the DSA, if online service providers become aware of the illegal content, either through third-party notices or own-initiative investigations, it can still benefit from immunity under Article 6(1)(b) by promptly removing the content.

80 Second, the Courts' narrow interpretation of good faith and objectionable content suggest that similar to the EU requirements for voluntary content moderation measures, the good faith standard under Section 230(c)(2)(A) is expected to be assessed from an objective perspective. Similarly, the term 'otherwise objectionable' has been given an objective reading and confirm the principle of *ejusdem generis* in interpreting objectively the term 'otherwise objectionable'. This narrow interpretation would render the removal decisions of online service providers to be objective and therefore non-discriminatory as it is required under the DSA.

227 Eric Goldman, 'Section 230(c)(2) Gets No Luv From the Courts–Song Fi v. Google'(*Technology & Marketing Law Blog*, 12 June 2015) <https://blog.ericgoldman.org/archives/2015/06/section-230c2-gets-no-luv-from-the-courts-song-fi-v-google.htm>, accessed 16 March 2024.

228 CRS, *supra* note 92.

229 *Zango v. Kaspersky* ("*Unless § 230(c)(2)(B) imposes some good faith limitation on what a blocking software provider can consider "otherwise objectionable," or some requirement that blocking be consistent with user choice, immunity might stretch to cover conduct Congress very likely did not intend to immunize*").

230 ECtHR, Guide on the case-law of the European Convention on Human Rights, European Union law in the Court's case-law (2022) <https://www.echr.coe.int/documents/d/echr/Guide_EU_law_in_ECHR_case-law_ENG>, accessed 16 March 2024.

231 Delfi was deemed a publisher and not an information society service provider which would have made it eligible for liability exemptions under Article 14 of the Directive.

81 Finally, the DSA explicitly mandates that online service providers diligently remove illegal content, especially when it can be established without a detailed legal examination that such content is illegal.[232] Although Section 230(c)(2)(A) does not explicitly require online service providers to moderate content diligently, the US Courts have, to some extent, reflected this requirement by asserting that good faith moderation efforts involve the removal of content which is indeed objectionable.

82 Under a broad interpretation of Section 230(c), the following differences between the immunity provisions in the two jurisdictions can be noted.

83 First, Section 230(c)(1) immunizes online service providers for any cause of action, including any decision to *knowingly* host objectionable content or even facilitate and encourage illegal activity. This is in contrast with the provisions in the DSA which immunizes service providers if they *unknowingly* host illegal content or promptly remove content upon obtaining knowledge.

84 Second, if 'otherwise objectionable' is anything that the user or provider finds objectionable, then the term constitutes a subjective standard.[233] Such a subjective approach would allow voluntary content moderation policies to be discriminatory as they would favour certain types of views or messages.[234]

85 Third, embracing this broad interpretation stemmed from policy considerations and purpose arguments to justify the promotion of unrestricted speech on the internet.[235] The Courts' overemphasis on free speech is, however, made to the detriment of public safety and welfare.[236] Additionally, the Courts' frequent reliance on Section 230(c)(1) instead of (c)(2) implies a primary consideration of fostering free speech and a secondary focus on addressing illegal content. In contrast, the CJEU has given significant consideration to the freedom of expression and right to information of service recipients when examining injunctions to remove or disable access to illegal content. This suggests an effort at striking a balance between the rights and interest of all parties involved and the objective of fighting illegal content online.

86 Finally, neither the statute nor the US case law suggests that content moderation measures should take into account the principle of proportionality, which under the DSA mean that voluntary content moderation measures should be implemented without unduly infringing upon the rights of other parties, including the freedom of online service providers to conduct their business. Nonetheless, the broad interpretation of Section 230(c)(1) has been based on the purposive argument of protecting free speech, such that service providers have the right, under the First Amendment to host or not host any content they want.[237]

D. Proposals to reform Section 230(c)(2)

87 In view of growing harms resulting from hate speech, disinformation, and the proliferation of other illegal content available on the internet, especially material related to pornography and child sexual abuse, there have been calls to amend Section 230 to give impetus to online service providers to keep and remove "slime" off their websites.[238] Most of the proposed amendments to Section 230 seek to define the meaning of the terms 'good faith' and 'otherwise objectionable' content pertaining to Section 230(c)(2)(A) and to clarify the interaction between Section 230(c)(1) and Section 230(c)(2) when it comes to providing immunity for removal decisions.

88 A number of proposals have been put forward to clarify the ambiguities in Section 230, particularly the lack of a statutory definition of good faith and objectionable content in Section 230(c)(2)(A). The proposals aim to depart from the subjective standard of the terms, by ensuring that removal decisions do not apply a selective enforcement of the policies[239] or are only undertaken when the provider or user has an objectively reasonable belief that the content is objectionable.[240] The proposals also highlight a gap

232 See *supra* note 48.

233 Candeub, *o.c.*

234 *Ibid.*

235 Klapper, *o.c.* See also Statement of Justice Thomas in *Malwarebytes v. Enigma*, *supra* note 87.

236 Meeran, *o.c.*

237 Daphne Keller, 'Who Do You Sue? State and Platform Hybrid Power Over Online Speech' (2019) 1902 Aegis Series Paper.

238 U.S. Sen. Ron Wyden in an interview with Esquire:*"I would like the big tech companies to do more to step up and deal with the slime that's on their platform. The companies are clearly capable of doing it when they think it helps their bottom line"*. (2019), <https://classic.esquire.com/article/2019/4/1/legislate-against-the-machine>, accessed 17 March 2024.

239 S.3983 ("Limiting Section 230 Immunity to Good Samaritans Act") proposed by Sen. Josh Howley in July 2020 <https://www.congress.gov/bill/116th-congress/senate-bill/3983/text?r=6&s=1.>, accessed 17 March 2024.

240 H.R. 3827 ('Protect Speech Act') introduced by US Rep.

between what Section 230(c)(1) is meant to apply to (i.e. claims for content that is left up)[241] and what in practice is used for (i.e. claims for content that is both left up and taken down).[242] A proposed bill would remove liability exemptions if the providers were aware of the illegal content or activity.[243]

89 There are also some reform efforts on the State level. For instance, a Texas law that forbids large service providers from removing or moderating content based on a user's viewpoint is awaiting review from the SCOTUS.[244] The law would violate online service providers' free speech rights under the First Amendment as it would force them to carry content that violates their content moderation policies. Although the law does not propose to amend Section 230, it may be in contradiction with the immunities afforded by it.[245]

90 Finally, suggestions for reforms have also come from academia to remove the 'good faith' covenant from Section (c)(2)(A) as it only "invites judicial confusion [...] only to reach the same result: a prevailing defendant".[246] Others, like Keats Citron and Wittes, have suggested that online service providers should be afforded immunity from liability if they could show that they have taken reasonable steps to prevent the illegal uses of their services.[247] Rustad and Koenig have recommended linking liability exemption to the lack of actual knowledge,[248] reflecting the common law distributor liability described in Cubby v. Compuserve.[249] A similar recommendation is that the test for liability exemption should be that the plaintiff first alleges that the defendant had actual knowledge of the illegality of content, after which the burden of proof switches to the defendant to show that it did not have knowledge and can invoke immunity under Section 230(c).[250] The test stems from the common law distributor liability described in Cubby, where a provider would be held liable for third-party content only if it knew or should have known about the unlawful content.[251]

91 The proposals to reform Section 230 are a consequence of the fact that the current provisions (or at least their interpretation by the Courts) affording immunity to online service providers no longer reflect the realities of how online service providers operate nowadays compared to when CDA was enacted more than twenty years ago. They also suggest a growing dissatisfaction with the online service providers' content moderation policies. Furthermore, it is evident for some that the broad immunity afforded to online service providers enabled them to act in bad faith contrary to the requirements of the Good Samaritan rule.

92 The approach to voluntary content moderation measures in Article 7 of the DSA diverges from Section 230(c)(2)(A) in that they are more explicit regarding the standards of acting in good faith and in a diligent manner, and with due regards to the fundamental rights of users. However, the proposals to amend Section 230(c)(2)(A) tend to align with the requirements of acting in good faith in the DSA. Specifically, the proposals seek to define objectionable content and the good faith covenant as acting in an objective and non-discriminatory manner, as it has been interpreted in a few cases by the US Courts. The proposed definitions would prevent removal decisions to be animated by pretextual, discriminatory or fraudulent motives, often inconsistent with their terms of service. Some recommendations to amend Section 230 would align the test of Section 230(c)(1) to that in Article 6 of the DSA which is based on whether the service provider

Jordan in June 2021.

241 CRS, *supra* note 92 (*"One conception of these two provisions is that Section 230(c)(1) applies to claims for content that is "left up,"*).See also Statement of Justice Thomas in *Malwarebytes v. Enigma*, *supra* note 87 (*"the statute suggests that if a company unknowingly leaves up illegal third-party content, it is protected from publisher liability by §230(c)(1)"*) and *supra* note 186 and the accompanying text.

242 CRS, *supra* note 92 (*"In practice, however, courts have also applied Section 230(c)(1) to "take down" claims"*). See also Statement of Justice Thomas in *Malwarebytes v. Enigma*, *supra* note 87 (*"This modest understanding is a far cry from what has prevailed in court [...] courts have relied on policy and purpose arguments to grant sweeping protection to Internet platforms"*).

243 PACT Act, S. 4066, 116th Cong. § 6 (2020).

244 Texas bill HB20, <https://capitol.texas.gov/BillLookup/History.aspx?LegSess=872&Bill=HB20.>, accessed 17 March 2024.

245 The Fifth Circuit did not consider whether the law is preempted by Section 230.

246 Goldman, *supra* note 113.

247 Danielle Keats Citron and Benjamin Wittes, 'The Internet Will Not Break: Denying Bad Samaritans Section 230 Immunity', (2017) 86 Fordham L. Rev. 401, <https://ir.lawnet.fordham.edu/flr/vol86/iss2/3>, accessed 17 March 2024.

248 Rustad and Koenig, *o.c.*

249 *Cubby, Inc. v. CompuServe Inc,.*

250 Meeran, *o.c.*

251 *Ibid.*

has actual knowledge of the illegality of content.[252]

93 The proposals advocate for more transparency when removing content by attaching certain requirements to objectionable content and good faith covenant and by requesting online service providers to be more transparent in their terms of service. The DSA already covers transparency obligations over voluntary content moderation, including for the use of automated tools.[253] Such transparency requirements would ensure that voluntary content moderation measures have met the requirements of objectivity, proportionality, and non-discrimination. Additionally, they would verify that the safeguards of limiting errors and avoiding unjustifiable removal of content have been implemented.

E. Conclusions

94 The purpose of this research paper was to provide a comparative analysis of online service provers' liability exemptions when undertaking good faith voluntary content moderation measures in the EU and the US. The objective was to determine to what extent the provisions in the two jurisdictions allow online service providers to keep their immunity, while at the same time achieving the objective of combating illegal and objectionable content online. The research paper has shown that under a narrow interpretation of Section 230(c)(1) CDA, both Section 230(c)(1) and Article 6(1)(b) of the DSA apply to online service providers that unknowingly host third-party illegal content. As for immunity for voluntary content moderation in Section 230(c)(2)(A) and Article 7 of the DSA the language of the statute suggests that both provisions apply to online service providers that remove or restrict access to illegal or objectionable content in good faith. In this respect, this research paper has shown that both provisions require (i) a nexus between liability exemption and the action of moderating content,[254] (ii) a good faith requirement which in both jurisdictions seems to be understood as being an objective standard, implicitly requiring a non-discriminatory approach, (iii) a decision to remove content that is either manifestly illegal or based on a reasonable belief that it is objectionable.

95 Under the DSA, the effectiveness of applying these standards will determine the online service providers' eligibility for liability exemptions. Whether taking a more explicit stance on these standards will effectively fight illegal content online and serve as the advocate for liability exemptions will become clearer once the jurisprudence on voluntary good faith moderation measures under DSA will mature. What we can learn from the US is that open terms such as 'good faith' or 'otherwise objectionable' content *"invites judges to introduce their own normative values into the consideration"*.[255] Given that good faith and diligence requirements in Article 7 of the DSA are standards, it is likely that the CJEU may still have to determine what acting in good faith and diligent manner requires in the circumstances of the specific case, similar to how it has been done in the US.

96 Under a broad interpretation of Section 230(c)(1), online service providers have enjoyed immunity even when they knew about the illegal activity, *"deliberately le[ft] up unambiguously unlawful content"*[256], and encouraged or facilitated illegal content.[257] By relying on Section 230(c)(1) to protect online service providers from any cause of action, the defendants prevailed in cases of sex trafficking of minors, illegal sale of guns, defective sale of products and even the encouragement of terrorist acts.[258] It also reduced their incentives to moderate content and fight illegal content online. This, in turn, is one of the reasons prompting calls for a reform of Section 230. The proposed reforms reveal not only the shortcomings in the Courts' interpretation of the statute, but they also suggest an alignment with some of the requirements attached to the good faith and diligence standards in the DSA, such as objectivity and non-discrimination.

97 The few cases dealt under Section 230(c)(2)(A) suggest that bad faith content moderation will lead to a loss of immunity, whereas the cases dealt under Section 230(c)(1) suggest that bad faith content moderation can still guarantee immunity depending on the Court of Appeal that examines the case. Since Section 230(c)(1) does not require a good

252 See Rustad and Koening, *o.c.*, proposing to reform Section 230 so that online service providers are liable only if they have actual knowledge *"and fail to expeditiously disable access to the posted illegal content"*. See also *supra* note 251.

253 DSA, Recital 66, Article 15.

254 See *supra* note 119.

255 Goldman, *supra* note 120.

256 Danielle Keats Citron and Marie Anne Franks, 'The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform' (2020), *2020*(3) University of Chicago Legal Forum, <https://chicagounbound.uchicago.edu/uclf/vol2020/iss1/3>, accessed 17 March 2024.

257 Klapper, *o.c.*, page 1258 and 1305 and referring to *Doe v. Backpage.com*, LLC, 817 F.3d 12, 22 (1st Cir. 2016), where the defendant prevailed under Section 230(c)(1) even though it facilitated illegal conduct.

258 Franks, *o.c.*

faith action and extends its protection to all types of action, defendants prevailed in Courts by invoking Section 230(c)(1) and thus bypassing the good faith standard required under Section 230(c)(2)(A). Such a broad interpretation diverges from the conditions for liability exemptions in the DSA which require that the online service providers do not have actual knowledge of the illegal content and when they do, they act promptly to remove it.

98 Interpreting Section 230(c)(1) to apply to any cause of action, including any removal decisions, allows service providers to bypass the good faith[259] standard in Section 230(c)(2)(A) or to benefit from immunity even when knowingly facilitating illegal activity online. But providing sweeping immunity for any type of actions,[260] under the pretext of promoting unfettered speech on the internet,[261] does not align with Congress's original intent when enacting Section 230 to protect minors from indecent material on the internet. It also reduces the incentives to moderate and fight illegal content online. The DSA aims to strike a balance between protecting various interests and fundamental rights, including users' freedom of expression and other rights enshrined in the EU Charter and maintaining the service providers' liability exemptions for voluntary content moderation measures. Whether this balance will effectively fight illegal content online will require further examination. The issue could be potentially clarified by the CJEU once it has the opportunity to examine a case on liability exemptions and voluntary content moderation measures under the DSA.

99 Liability exemptions for lack of *actual knowledge* has also been advocated as a measure to reform Section 230,[262] reflecting the common law distributor liability upheld in *Cubby*. Although such proposal may be controversial in many respects for the supporters of Section 230,[263] the knowledge standard is nevertheless used for addressing copyright infringements under the DMCA and has been the pillar of the EU liability regime for over two decades. While the notice and take-down regime in the EU is more effective at fighting illegal content online, the US system of shielding online service providers for failure to remove illegal content is more effective at protecting free speech. Choosing one regime over the other is a matter of policy choice.

259 *Murphy v. Twitter, Inc.*, 2021 WL 221489 (California Appeal Court, Jan. 22, 2021).

260 Statement of Justice Thomas in *Malwarebytes v. Enigma*, *supra* note 87.

261 See *supra* note 235.

262 See *supra* note 252.

263 Meeran, *o.c.*