

Platform regulation, content moderation, and AI-based filtering tools

Some reflections from the European Union

by **María Barral Martínez***

Abstract: Online platforms have voluntarily relied on screening tools for content moderation purposes for quite some time now. They do so to deal with the problems of scale and the speed content is shared online. Currently, the efforts of online platforms to fight illegal and harmful content are continuously focusing on innovative AI-based solutions for better performance of their content moderation systems. At the same time, in the EU, new rules on content moderation are entering the arena.

These rules may require a more active role of online intermediaries to detect and remove illegal content in their sites. This begs the question whether we are moving towards a filtering obligation in disguise on online intermediaries. If that is the case, are AI-based filtering systems fit to avoid blocking lawful content? What safeguards should be taken at regulatory level to ensure the protection of fundamental rights of online users?

Keywords: content moderation; artificial intelligence; filtering tools; platform regulation; DSA; EU law

© 2023 María Barral Martínez

Everybody may disseminate this article by electronic means and make it available for download under the terms and conditions of the Digital Peer Publishing Licence (DPPL). A copy of the license text may be obtained at <http://nbn-resolving.de/urn:nbn:de:0009-dppl-v3-en8>.

Recommended citation: María Barral Martínez, Platform regulation, content moderation, and AI-based filtering tools: some reflections from the European Union, 14 (2023) JIPITEC 211 para 1.

A. Introduction

1 The vast amount of digital content being uploaded and posted by users of online platforms—such as Meta, Twitter, or YouTube—is leading these companies to invest in better technologies to efficiently track and block illegal and harmful content. Until now, this has been a self-governance voluntary effort from online platforms.¹ However, in the EU, a wave of new regulatory instruments to tackle online illegal content may put service providers between a rock and a hard space.

2 Are we moving towards a *de facto* obligation on online platforms to use filtering systems in the EU? If so, are AI-based filtering systems fit to avoid blocking lawful content? What safeguards should be taken at regulatory level? In light of the EU current legal developments, this paper analyses the technological limitations and legal challenges arising from the use of AI based filtering tools in content moderation. Despite the progress made by Digital Services Act Regulation setting up transparency and accountability requirements for online platforms, there are still a few issues that deserve regulatory attention. The paper is divided as follows: the second part provides background on content moderation and algorithmic screening tools. Part C analyses the EU legal landscape impacting content moderation from current rules to future measures. In part D, the article explores the technological concerns of AI based filtering tools in an EU context-specific assessment. Finally, part E takes stock on the implications of imposing

* María Barral Martínez, Legal Counsel, LuxTust S.A., LL.M International and EU Law University of Amsterdam.

1 The term online platform is used in a broad sense to capture the different categories of internet intermediaries under the scope of analysis of the present article.

filters on online intermediaries and calls for further regulatory responses.

B. Facts and technology

- 3 Next to the traditional hashing, watermarking, and fingerprinting technologies for automated content recognition (ACR)², online service providers like Meta³ or YouTube⁴ are relying on new artificial intelligence (AI) enhanced solutions to deploy content moderation screening tools in a more efficient manner. Content moderation is the organized practice of screening user-generated content (UGC) posted to Internet sites, social media, and other online outlets, to determine the appropriateness of the content for a given site, locality, or jurisdiction⁵. In broad terms, content can be illegal, lawful but harmful—the so-called “lawful but awful” content—or go against the terms of use or community guidelines of the online service provider.
- 4 While moderation has traditionally been a job for humans, for reasons of scale and costs, artificial intelligence tools have been developed to help with the task. Algorithmic content moderation techniques aim at identifying, matching, predicting some piece of content on the basis of its exact properties or general features.⁶ Within this context, companies usually use matching or predictive models.⁷

Matching algorithms require a manual process of collating and curating individual examples of the content to be matched. Classification algorithms predict the likelihood that a previously unseen piece of content violates a rule.⁸ When a piece of content is a match or is classified as content that violates a rule, the content can be flagged for review, deleted, or prevented from going online.⁹

- 5 Last year, YouTube released its first Copyright Transparency report providing some insight in their platform copyright enforcement actions.¹⁰ In Meta's latest community standards enforcement report, the social media online platform highlighted the better performance in detecting harmful content thanks to proactive detection technologies based on AI.¹¹
- 6 Figures speak by themselves: YouTube processed 729.3 million copyright actions in the first quarter of 2021¹², Meta has acted against 905,000 pieces of content related to terrorism only over the last quarter of 2021, and Twitter removed in the first half of 2021 5.9 million pieces for violating Twitter rules.¹³ According to the World Economic forum, by 2025 the amount of data created globally by humans each day will reach 463 exabytes.¹⁴ Against this background, reliance, and investment on these technologies to detect illegal and harmful content seems the way forward to tackle such a massive amount of online content.

2 European Union Intellectual Property Office Automated content recognition: discussion paper. Phase 1, Existing technologies and their impact on IP, 2020 p. 5 <<https://data.europa.eu/doi/10.2814/52085>>.

3 <https://ai.facebook.com/blog/the-shift-to-generalized-ai-to-better-identify-violating-content>

4 Julia Alexander, “Youtube can now warn creators about copyright issues before videos are posted” *The Verge* (17 March 2021) <<https://www.theverge.com/2021/3/17/22311111/youtube-copyright-issues>> accessed 08 April 2022

5 See definition at Roberts, S.T. (2022). Content Moderation. In: Schintler, L.A., McNeely, C.L. (eds) *Encyclopaedia of Big Data*. Springer, Cham. <https://doi-org.proxy.bnl.lu/10.1007/978-3-319-32010-6_44>.

6 Robert Gorwa et al., Algorithmic content moderation: Technical and political challenges in the automation of platform governance, *Big Data & Society* 2020, p.3. <<https://doi.org/10.1177/2053951719897945>>.

7 Nafia Chowdhury, Daphne Keller, Automated Content Moderation: A Primer, Stanford Cyber Policy Center, 2022, p.2. <[FSI | Cyber - Automated Content Moderation: A Primer \(stanford.edu\)](https://www.stanford.edu/cyber-policy-center/automated-content-moderation)>.

8 Ibid, p.2.

9 Gorwa (n 6) p.6.

10 YouTube Copyright Transparency Report H1 2021 <[YouTube Copyright Transparency Report H1 2021 \(storage.googleapis.com\)](https://www.youtube.com/transparencyreport)>.

11 Guy Rosen “Community Standards Enforcement Report, Fourth Quarter 2021” Meta news room (1st March 2021) <<https://about.fb.com/news/2022/03/community-standards-enforcement-report-q4-2021/>> accessed 8 March 2022.

12 Paul Keller “Youtube copyright transparency report: Overblocking is real” (Kluwer Copyright blog 9 December 2021) <[YouTube Copyright Transparency Report: Overblocking is real - Kluwer Copyright Blog \(kluweriplaw.com\)](https://www.kluweriplaw.com/youtube-copyright-transparency-report-overblocking-is-real)> accessed 8 March 2022.

13 Twitter Transparency Report published in January 2022 available at Rules Enforcement - Twitter Transparency Center. accessed 8 March 2022.

14 Rem Darbinyan “The growing role of AI in content moderation” *Forbes* (14 June 2022) <https://www.forbes.com/sites/forbestechcouncil/2022/06/14/the-growing-role-of-ai-in-content-moderation/> accessed 10 August 2022.

C. The EU legal landscape

7 Before delving into the challenges of automated filters applied to the content moderation scene in the EU, it is helpful to briefly go through the rules on illegal content online. In the EU, illegal content online is subject to two layers of regulation: at EU level, a horizontal framework and sectoral regulation for specific types of content, and then Member State national laws. Until now, the horizontal rules were set by the e-Commerce Directive, but soon the Digital Services Act (DSA)¹⁵ will be the central piece of legislation. Sectoral rules are for example the Audiovisual Media Services Directive¹⁶, the Directive on Copyright in a Digital Single Market (DSM Directive)¹⁷ or the Terrorism Online Content Regulation (TERREG).¹⁸

I. The current horizontal framework

8 Articles 14 and 15 are the e-Commerce Directive key provisions for intermediaries' liability and content monitoring.¹⁹ Pursuant to Article 14, intermediaries of online services are exempt from liability for content stored in their services by its users, subject to not being aware of illegal activity or information in their services, or if made aware, for example, through an injunction ordered by a Court, to expeditiously remove or to disable access to the content. Article 15 prohibits Member states to impose a general obligation on providers [...] to

monitor information which they transmit or store, or actively seek facts or circumstances indicating illegal activity. In the case *Poland v Parliament*²⁰, Advocate General (AG) Saugmansgaard Øe regarded the prohibition enshrined in Article 15 as a general principle of law governing the internet.²¹

9 What constitutes general monitoring against specific monitoring has not been determined by the e-Commerce Directive.²² The European Court of Justice (ECJ) explored the subject in judgments like *L'Oreal vs Ebay*²³ and *Scarlet Extended v SABAM et al.*²⁴, and provided some sort of guidance on what kind of content screening is allowed under Article 15 in *SABAM vs Netlog*²⁵ and *Glawischnig-Piesczek*²⁶.

10 In *SABAM vs Netlog*, SABAM—a Belgium private collective rights management organisation—sought through an injunction against Netlog, that the latter install a filtering system at their own cost to prevent copyright infringements of their repertoire. The ECJ found that preventive monitoring not compatible with Article 15.²⁷ The deployment of such a system, would require the social media company Netlog to carry an active monitoring of almost all the data stored relating to all of its service users.²⁸ In this case, the obligation to monitor was broad and too burdensome for Netlog, and it would be at odds with Netlog's freedom to conduct a business and its users right to personal data and freedom of information.²⁹

15 Regulation (EU) 2022/2065 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 DSA).

16 Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of Audiovisual media services (*Audiovisual Media Services Directive*).

17 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 [2019] OJ L 130/92. (*DSM Directive*).

18 Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online (*TERREG*).

19 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, 2000 O.J. (L 178) 1-16 (*e-Commerce Directive*).

20 C-401/19 Republic Poland v European Parliament and of the Council of the European Union [2022] ECLI:EU:C:2022:297 (hereinafter *Poland v Parliament*).

21 C-401/19 Republic Poland v European Parliament and of the Council of the European Union [2022] ECLI:EU:C:2021:613, Opinion AG Saugmansgaard Øe, point 106.

22 Folkert Wilman, *The responsibility of online intermediaries for illegal user content in the EU and the US* (Edward Elgar Publishing Limited 2020).

23 C-324/09, *L'Oreal v Ebay*, [2011] ECLI:EU:C:2011:474

24 C-70/10 *Scarlet Extended SA v SABAM* [2011] ECLI:EU:C:2011:771.

25 C-360/10, *Sabam v Netlog* [2012] ECLI:EU:C:2012:85 (*SABAM v Netlog*)

26 C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Ltd* [2019] ECLI:EU:C:2019:821 (*Glawischnig-Piesczek*)

27 *Ibid* 25 C-360/10, para 38.

28 *Ibid*.

29 *Ibid* paras 47-48.

- 11 In *Glawischnig-Piesczek*, an Austrian Court was concerned with whether an interim injunction against a host provider (Facebook), to remove a post previously declared defamatory could also extend to other posts of identical or equivalent content. Here, the Court held that the measure did not impose a general obligation to monitor within the meaning of Article 15. However, the national court order for removal of identical or equivalent defamatory content should contain “specific elements” to identify the content—targeted monitoring one could say—and in any event, it should not require an independent assessment of the content by the host provider because it will make use of automated tools.

II. New EU rules striving for a safer online environment in the Digital Single Market

- 12 The DSA seeks to contribute to the proper functioning of the internal market by harmonising the rules for intermediary services, such as social media networks or marketplaces, to tackle the spread of illegal content, address online disinformation, and other societal risks.
- 13 Articles 7 and 8 are of special interest: Article 7 shields against liability those intermediary services which in good faith and diligently [...] take measures aimed at detecting, identifying, and removing, or disabling of access to illegal content or take the necessary measures to comply with the requirements of national law, in compliance with Union law, including the requirements set out in this Regulation.
- 14 Article 8 contains the prohibition on general monitoring and active fact-finding, replicating the wording of Article 15 e-Commerce Directive. It is worth mentioning that throughout the legislative process, the European Parliament (EP) made an amendment to Article 8 by clarifying that there is no general obligation to screen information providers transmit and store *neither the jure nor the facto through automated or non-automated means*.³⁰ In addition, the EP also introduced a new limb to Article 8 stating providers of intermediary services should not be obligated to use automated tools for content

moderation [...]. Both amendments, however, did not make it to the final version of the text just approved at time of writing. Article 8 now reads as follows: “no general obligation to monitor the information which providers of intermediary services transmit or store, nor actively to seek facts or circumstances indicating illegal activity shall be imposed on those providers”.³¹

III. Current sectoral measures

- 15 In recent years, on the online content sector-specific front, several legal instruments have been passed and others are now in the pipeline of the EU legislature. These measures can target specific types of online service providers or particular categories of illegal content harmonised under EU law.³²
- 16 Under the Audiovisual Media Services Directive, Article 28b requires Member States to ensure video-sharing platforms providers (VSPs) take appropriate measures against illegal and harmful content. These measures, however, should not lead to any ex-ante control measure or upload-filtering of content contrary to Article 15 of the e-Commerce Directive.
- 17 The DSM Directive ignited a heated debate around its Article 17. The lengthy provision on the use of protected works by online content-sharing services providers (OCSSPs), sets out a specific liability regime for OCSSPs departing from the principle under Article 14 of the e-Commerce Directive.³³ OCSSPs can be liable for the content uploaded by its users to their services when such content infringes copyright-protected works. To escape liability for acts of communication to the public and make available to the public copyright-protected works, OCSSPs shall obtain licenses for these works or make best efforts to obtain them. In the event of no licensing agreements, OCSSPs are subject to the obligation to prevent the availability of those works in their services and to the take down and stay down of that content.³⁴
- 18 Poland challenged the legality of Article 17 before the ECJ.³⁵ It argued the obligations arising from Article

30 Amendments adopted by the European Parliament on 20 January 2022 on the proposal for a regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC(COM(2020)0825 – C9-0418/2020 – 2020/0361(COD))1, amendment 139-140 Article 7 of the proposal corresponding to Article 8 of the final version. Emphasis added.

31 DSA Regulation.

32 De Streel, A. et al. (2020) p.15.

33 See Recital 65 DSM Directive.

34 Article 17(4) letters (a), (b),(c) DSM Directive.

35 Ibid (n 20) Case C-401/19.

17.4³⁶ on OCSSPS implicitly require the use of filtering technologies to monitor content uploaded by users to prevent the infringement of copyright. In the view of the Polish government, deploying automatic filters is a serious interference on the users right to freedom of expression and information. The Court dismissed Poland's action and reasoned that even though Article 17.4 liability regime indeed imposes a limitation on the exercise of freedom of expression and information of users, Article 17 provides appropriate safeguards to preserve the essence of that right as guaranteed by Article 11 of the Charter of Fundamental Rights (CFR). Furthermore, the ECJ agreed that some filtering will be needed to comply with the mandates of Article 17.³⁷ Yet, as long as these filters do not screen and block lawful content when uploaded by users, their use is compatible with Article 11 CFR.

- 19 Since June 7, 2022, the TERREG is in force. Hosting service providers are obligated to remove or disable access to terrorist content at least within one hour of receipt of a removal order from a competent authority of any Member State.³⁸ Pursuant to Article 5.8, hosting service providers, when implementing specific measures³⁹ to address the dissemination of terrorist content in their services, are under no obligation to use automated tools. However, Recital 25 clarifies that providers should have recourse to automated tools if they consider them appropriate and necessary to address the dissemination of terrorist content online. When using automated means, providers should take appropriate measures through human oversight and verification and ensure accuracy to avoid blocking or removing content that is not terrorist related.⁴⁰

IV. Future sectoral measures

- 20 More controversial is the new proposal for a Regulation fighting child sexual abuse published in early May of 2022.⁴¹ The Regulation seeks to harmonize the requirements imposed on online services providers removing the divergences from Member States rules to prevent and combat

child sexual abuse.⁴² It complements the general framework of the DSA and among others, it introduces an obligation on providers to detect, report, remove and block child sexual abuse material (CSAM). Apart from the privacy and mass surveillance concerns voiced⁴³, the Article 10 mandate is of interest. Online service providers shall execute detection orders by national authorities by installing and operating technologies—AI systems—to detect the dissemination of CSAM, favouring systems which have been vetted by a new coordination authority, the EU Centre on Child Sexual Abuse.⁴⁴ Although the proposal explains these orders will be specific and targeted, it is not yet clear how the screening would be performed and if the current tools are effective. Some commentators warned there are no technologies available that can safely scan people's messages or discern what is abusive from what is not.⁴⁵

D. AI-based filtering for content moderation: technological concerns

- 21 Online platforms are filters only in the way that trawler fishing boats “filter” the ocean: they do not monitor what goes into the ocean, they can only sift through small parts at a time, and they cannot guarantee that they are catching everything, or that they are not filtering out what should stay.⁴⁶

36 Ibid para 24.

37 Ibid para 54.

38 Article 3(3) TERREG

39 Article 5 TERREG.

40 Article 5(3) and Recital 24 TERREG.

41 Proposal for a regulation to prevent and combat child sexual abuse (COM (2022) 209 2022/0155 (COD) (CSAM proposal)

42 Both providers of hosting services and providers of interpersonal communication services.

43 James Vicent “New Eu rules would require chat apps to scan private messages for child abuse” (The Verge 11 May 2022) <<https://www.theverge.com/2022/5/11/23066683/eu-child-abuse-grooming-scanning-messaging-apps-break-encryption-fears>> and Mathieu Pollet “children first, privacy second” (Euroactiv 270502022) <<https://www.euractiv.com/section/digital/podcast/csam-proposal-children-first-privacy-second/>>.

44 Ibid 28 Article 10 CSAM proposal.

45 Mathieu Pollet “CSAM proposal: children first, privacy second?” *Euroactiv* (27 May 2022) <<https://www.euractiv.com/section/digital/podcast/csam-proposal-children-first-privacy-second/>> (accessed 09 September 2022) and <https://edri.org/our-work/private-and-secure-communications-put-at-risk-by-european-commissions-latest-proposal/>.

46 Gillespie Tarleton, *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media* (Yale University Press 2018) p.87.

22 The use of filtering solutions can lead to over-blocking patterns by online service providers. In other words, lawful content which should in principle be allowed online, can risk being caught by a filter, be flagged or removed. Some authors have already signalled the limits of automated systems and the challenges posed by false positives.⁴⁷ Others argued there will always be a need for human intervention for the content to be appropriately screened.⁴⁸ This is partly because filtering technologies have a problem with content contextualization, they are able to detect certain content but not infringing content *per se*.⁴⁹ As a result, a piece of content that can be illegal in certain circumstances, may not be if used in a different context.⁵⁰

I. Training the algorithm: context, human bias, and accuracy challenges

23 For the efficient deployment of a filtering system in content moderation, the premise is that the system will work with clear and defined parameters of what constitutes illegal or harmful content. The first challenge in this respect is to define the nature of the content and work backwards—i.e., why a post can be labelled as hate speech, or what is hate speech for that matter. In AI terms, this would consist in training the model with data sets to teach the system to recognize on its own the targeted illegal content. In this process, the quality of the data fed to the system will be key. Automatic detection can assess only what it can know and what can be represented as data, but limited to the data it has.⁵¹ In addition, algorithms can be subject to human bias during the AI training process. Human bias can take place

in both machine-supervised learning and non-supervised learning processes. In the former, human intervention is needed to evaluate data examples and select the appropriate labels or to evaluate an automatically applied labels.⁵² In the later, hidden biases could arise from the dataset itself.⁵³

24 Further, considering these technologies are context insensitive and unable to make subjective decisions⁵⁴ there are certain bars at technical level. Although some context can be incorporated in a tool, historical, political, and cultural context are more difficult for an AI system to be trained to detect.⁵⁵ As Spoerri points out, the state-of-the-art of filtering technologies is quite limited as tools are only capable of matching content, but it is not yet possible to determine whether the use of a file—be it music, text or image—constitutes an infringement.⁵⁶ Despite this situation, the EU legislator seems to assume that online service providers can employ intelligent filters that identify infringing content while enabling the upload and making available of lawful content.⁵⁷

25 Against this background, the ECJ in C-401/19 warned that where a filtering system does not adequately distinguish between lawful and unlawful content, leading to blocking of lawful content, the system is not compatible with the right to freedom of expression in Article 11 CFR.⁵⁸

26 Yet, there is no infallible filtering system able to make such a clear distinction.⁵⁹ For that reason, the focus should be put on the accuracy of these tools. Accuracy in this context can be defined as the rate

47 Christophe Geiger and Bernd Justin Jütte, Platform Liability Under Article 17 of the Copyright in the Digital Single Market Directive, Automated Filtering and Fundamental Rights: An Impossible Match, PIJIP/TLS Research Paper Series no. 64, 2021 p.36.

48 Sarah T Roberts, Behind the screen: Content moderation in the shadows of social media. (Yale University Press 2019) p. 35.

49 Ibid 21 AG Saugmansgaard Øe Opinion in *Poland v Parliament* at point 148.

50 Giovanni Sartor, Andrea Loreggia, The impact of algorithms for online content filtering or moderation: upload filters. European Parliament, Directorate-General for Internal Policies of the Union, (2020) p.46. <<https://data.europa.eu/doi/10.2861/824506p>>.

51 Gillespie, (n 46) p.105.

52 Emma Llansó et Al., Artificial intelligence, content moderation, and Freedom of expression, Working, Transatlantic working group, paper series, 2020, p.8 <[doi:https://doi.org/10.1177/2053951720920686](https://doi.org/10.1177/2053951720920686)>.

53 Althaf Marsoof, Andrés Luco, Harry Tan & Shafiq Joty, Content-filtering AI systems—limitations, challenges and regulatory approaches, Information & Communications Technology Law, 2022 p.16.

54 Geiger and Jutte, (n 47) p.36 and Santa Clara Principles 2.0 Open Consultation Report (accessed 24 August 2022)

55 Ibid 52 Llansó.

56 Thomas Spoerri, On Upload-Filters and other Competitive Advantages for Big Tech Companies under Article 17 of the Directive on Copyright in the Digital Single Market, 10, JIPI-TEC, 2019 pp 173-186, p.182 at 35.

57 Ibid Geiger and Jutte (no 47) p.36.

58 Ibid (n 20) *Poland v Parliament* para 86.

59 Spoerri (note 56) p.182 at 34.

at which the tool's evaluation of content matches a human's evaluation of the same content.⁶⁰ The results can be divided into four categories: true positives, true negatives, false positives, and false negatives.⁶¹ This begs for the question how many false positives or false negatives are acceptable to not fall in over-blocking patterns threatening users' rights. Perhaps, certain standards should be set for the development and use of filtering technologies within this sphere, and improvements of filtering tools should focus on bringing these mistakes within an acceptable range.⁶² What is acceptable would depend on analysing the content and harm at stake. Trade-offs in this regard are unavoidable⁶³—a balance between leaving false negatives online and blocking lawful content should be achieved. At any rate, predictability of the systems should be guaranteed as well as mechanisms to correct the potential mistakes.

- 27 In the EU, another layer of complexity exists; the regulation of illegal content categories is not entirely harmonized at Union level⁶⁴ so the same type of content may be considered illegal, legal but harmful, or legal and not harmful across the 27 Member States.⁶⁵ This is reflected in the broad definition of illegal content enshrined in the DSA⁶⁶ “illegal content means any information, which, in itself or in relation to an activity, including the sale of products or the provision of services, is not in compliance with Union law or the law of any Member State, irrespective of the precise subject matter or nature of that law”.

60 Emma Llansó, No amount of “AI” in content moderation will solve filtering’s prior-restraint problem. *Big Data & Society*, 7(1) 2020, p.4. <doi:https://doi.org/10.1177/2053951720920686p4>.

61 Sartor and Loreggia (n 50) p. 45.

62 Llansó, (n 60) p 4.

63 Federal Trade Commission Report to Congress: Combatting Online Harms Through Innovation, June 2022 available at [Combatting Online Harms Through Innovation](#); Federal Trade Commission Report to Congress ([ftc.gov](#)) p. 41.

64 De Streef, A. et al., *Online Platforms’ Moderation of Illegal Content Online*, Study for the committee on Internal Market and Consumer Protection, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2020. <[https://www.europarl.europa.eu/Reg-Data/etudes/STUD/2020/652718/IPOL_STU\(2020\)652718_EN.pdf](https://www.europarl.europa.eu/Reg-Data/etudes/STUD/2020/652718/IPOL_STU(2020)652718_EN.pdf)> p.16.

65 See *Ibid* De Steel, p.16. and for example *The German law on Hate Speech, NetzDG (2017)* < *BMJ | Netzwerkdurchsetzungsgesetz*>.

66 Article (3)(g) DSA.

- 28 Therefore, online service providers deploying filtering systems need to consider that for those categories of illegal content not harmonized at EU level, tailor-made filters for specific jurisdictions within the EU need to be implemented. For larger online services providers operating worldwide, having to deal with different degrees of requirements across the globe or even with conflicting rules is already the case. This forces them to operate their compliance content policy and enforcement programmes based on global rules and adjust them through a risk-level approach. One should question then, if the risk of non-compliance comports fines as those established by the DSA⁶⁷, online service providers would not be prone to self-censorship/over-removal for the sake of compliance, paying little heed to the fundamental rights of its users.

II. The need for human review

- 29 The difficulties of screening tools to consider language and social/cultural context evidences the gap between the capabilities of a human and that of machines. The high rate of false positives and the removal of lawful content resulting from automated screening emphasize the added value of human moderation. It is for this reason that the inclusion of human review at some stage of the moderation chain should be a requirement to safeguard users’ fundamental rights. Typically, human review can take place when content is reported by a user and a decision needs to be made by the online service provider on the content flagged. Similarly, filters can serve to flag content by the platform own initiative and subsequently be reviewed by a moderator. Moreover, although most online platforms follow a “publish-then-filter approach”⁶⁸ human review can happen either before the content is online or after it is published.⁶⁹

- 30 If human content moderators are excluded entirely from the screening process, it will be the automated system deciding which content stays online or is taken down. Still, reviewing every piece of content caught by a filter as a potential infraction of the law or from the online service provider TOUs, would defeat the purpose of using content screening systems by the online service providers, rendering the content moderation exercise not feasible.

67 See Article 52.3 and 74 DSA.

68 *ibid* 51 p.75.

69 Roberts (n 48) 33.

- 31 To that extent, the ELI principles on automated decision making⁷⁰ proposed an ex-post content human review after a decision has been taken by automated means and challenged by the user. To put it simply, users who posted/uploaded a piece of content which was afterwards blocked or removed should have access to a redress mechanism to challenge the decision requiring human review. In that sense, human review guarantees full compliance with applicable law without relinquishing the benefits of automation.⁷¹ However, such an approach does not resolve the issue of users being at the mercy of online platforms and their internal dispute settlement mechanisms at first instance, forcing them to rely on them regardless the content disputed was lawful from the outset. An issue that fuels the debate on the role of these platforms acting as delegated enforcers of public powers vis-à-vis online users' freedom of expression and due process rights.⁷²
- 32 But what is EU law position on human review? Although the outcomes of non-compliance with a human review requirement remain to be seen, references to human review or human intervention can be found in different EU legal acts.⁷³ As an illustration, under the General Data Protection Regulation (GDPR)⁷⁴, data subjects have the right not to be subject to a decision based solely on automated processing without human intervention.⁷⁵
- 33 Some of the above-mentioned new rules introduce provisions on the inclusion of human review when using filtering technologies. The TERREG mandates hosting service providers to include human oversight and **verification safeguards** when using technological measures to protect its services against the dissemination to the public of terrorist content, to ensure accuracy and to avoid the removal of material that is not terrorist content.⁷⁶ In the same fashion, the proposal on CSAM establishes that providers should ensure regular human oversight and where necessary, **intervention**, to ensure technologies operate in a sufficiently reliable manner. Even more, when detecting potential errors and potential solicitation of children.⁷⁷ In the case of the DSA Regulation, Article 20(6) requires that decisions on removal/block of allegedly illegal content or content incompatible with the platform T&Cs, are reviewed by qualified staff and are not solely taken on the basis of automated means.⁷⁸ Therefore, the human review requirement comes within the context of the complaint handling system. This is, ex post, when the online service provider receives a complaint by a user of the platform.
- 34 Lastly, the ECJ had the opportunity to provide some precisions on human review while using automated filters through *Glawischnig-Piesczek* and *Poland v Parliament*. In the first judgement, the Court was of the view that a hosting service provider is not under the obligation to include human review—"an independent assessment" in the Court's words—when using automated filtering technologies to comply with a removal order by a national court. In the second case, the Court held in similar terms that OCSSPs are not obliged to conduct an independent assessment of the content uploaded by their users to prevent the uploading or making available to the public of copyright-protected works, in the light of the information provided by the rightsholders and of any exceptions and limitations to copyright.⁷⁹
- 35 This approach, however, seems difficult to conjugate with the rules discussed above and creates different standards for human review depending on the type of content at stake and on the subject requesting the removal. Even if human review in such situations

70 De las Heras Ballell, Teresa, *ELI Innovation Paper on Guiding Principles for Automated Decision-Making in the EU*, European Law Institute (2022). <ELI Innovation Paper on Guiding Principles for Automated Decision-Making in the EU by European Law Institute, TERESA RODRIGUEZ DE LAS HERAS BALLELL:: SSRN>.

71 Ibid.

72 For a more detailed discussion see Martin Husovec, *Ir Responsible Legislature? Speech Risks under the EU's Rules on Delegated Digital Enforcement* (2021) <<https://dx.doi.org/10.2139/ssrn.3784149>> and Víctor Javier Vázquez Alonso, *The «private» censorship of large digital corporations and the emerging system of freedom of expression*, *Teoría y Derecho*, no 32, Tirant, (2022) pp 108-129.

73 Codagnone, C. et Al., *Identification and assessment of existing and draft EU legislation in the digital field*, Study for the special committee on Artificial Intelligence in a Digital Age (AIDA), Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg (2022) p.61.

74 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC O.J. L 119 1-88. (GDPR).

75 Article 22 and to Recital 71 GDPR.

76 Article 5(3) (d) *in fine* TERREG.

77 Article 10(4) (c) and Recital 28 CSAM proposal.

78 Article 20(6) and Recital 45 DSA.

79 Ibid 20 case Poland v Parliament para 90.

is not required by law, in all likelihood AI based filters will match/block ambiguous content or lawful content when searching for the objectionable content. This is particularly relevant in the case of the copyright exceptions and the rights of content creators to use protected works without the prior authorisation of rightsholders. Essentially, this implies that someone whose lawful content was removed due to a filter mistake would have to go through the complaint handling system of the provider to challenge the removal. Only then human review would be required per Article 17(9) DSM. With this set up by default, the balance tilts towards the right to intellectual property vis-à-vis the freedom of expression and creation of users, placing a heavy burden on non-professional creators and user-generated content.⁸⁰

- 36 Although online service providers could of course still decide to rely on human review⁸¹ for those cases, even platforms that use filters and human review are incentivized to remove legal “grey area” content.⁸²

E. AI based filters for content moderation are here to stay, so what is next?

I. Towards a filtering obligation on online intermediaries?

- 37 For reasons of scalability, speed, and cost-efficiency, online service providers will keep relying on AI based filtering solutions on voluntarily basis to tackle illegal or harmful content.⁸³ Thus, the key question is no longer whether to rely on AI based systems to screen content, but whether automated content screening is turning into an obligation in disguise for online platforms and if so, how could it be articulated with online intermediaries’ liability

exemption and their fundamental rights, namely, their freedom to conduct a business.

- 38 Looking at Article 7 DSA and other sector-specific rules, one observes a trend of the EU legislator to require a more active role of online service providers to tackle illegal content.⁸⁴ In fact, the use of automated filtering systems seems a *de facto* must for online intermediaries to escape liability, especially in cases where short removal time is required. Recital 26 of the DSA sheds light on the scope of Article 7 DSA voluntary measures of providers to conduct investigations and actions for the detection, identification, removal or disable access to illegal content. It clarifies the requirements of conducting such activities “in good faith” and “in a diligent manner” by also stating that if the provider uses automated tools for those purposes, it should take reasonable measures to ensure the technology is sufficiently reliable to limit to the maximum extent possible the rate of errors. There are still some open questions, how the error rate could be measured, if a threshold should be established, or what happens if the filter fails to detect illegal content despite the intermediary voluntary actions. Would this “bad” filter engage the liability of a provider conducting voluntary measures to fight illegal content? A reading of Recital 22 tells us that to benefit from the exemption of liability, the provider, upon obtaining actual knowledge or awareness of illegal content, needs to act expeditiously to remove or to disable access to that content, and that knowledge of the illegal nature of the content can be obtained through own-initiative investigations.

- 39 In this regard, the case *Delfi AS v. Estonia*⁸⁵ of the European Court of Human Rights (ECtHR) is quite insightful. This Court found the liability imposed by the Estonian Supreme Court to an online news portal operator for defamatory comments made in their site by third parties, did not violate the applicant freedom of expression. The Court noted that Delfi’s filtering system in question failed to detect the harmful comment and left it online for some weeks.⁸⁶ This amounted to not having taken reasonable measures to remove the comments without delay. In such a circumstance, the Court found the liability on the online news operator to be a proportionate restriction on the applicant’s right to freedom of expression.

- 40 Looking at the case from another perspective, one could also ask if requiring an online service provider to deploy a filtering system to look for

80 Giancarlo Frosio and Sunimal Mendis, *Monitoring and Filtering European Reform or Global Trend?* in Giancarlo Frosio (ed), *The Oxford Handbook of Online Intermediary Liability* (Oxford University Press 2020) p.28.

81 Daphne Keller, *Facebook Filters, Fundamental Rights, and the CJEU’s Glawischning-Piesczek Ruling*, *GRUR International*, June 2020 Volume 69, Issue 6, pp 616–623 p.621.

82 *Ibid.*

83 See for example, Meta transparency statement on how Meta prioritises content for review at: < <https://transparency.fb.com/en-gb/policies/improving/prioritizing-content-review/>>(accessed 15 July 2022)

84 *Ibid* (n 50) p.59.

85 *Delfi A S v. Estonia* App no. 64569/09 (ECtHR, 16 06 2015).

86 *Ibid* para. 156 - 159.

illegal content or manifestly illegal content is a proportionate restriction to its freedom to conduct a business under Article 16 of the CFR. The answer would depend among other things, on the size of the internet intermediary and its resources. Thus, a liability obligation of that sort would create differences between the market players. As Frosio and Geiger flagged, the economic impact of enforcing filtering and monitoring obligations on online service providers has been discussed in the case-law of the ECJ, in particular, in *Netlog v. SABAM* where the Court held that imposing a monitoring obligation on Netlog to screen all works uploaded would burden the online service provider with the requirement of installing at its own expense filtering technologies.⁸⁷

II. What regulation?

- 41 The DSA renders online services providers accountable through algorithmic transparency and reporting obligations including disclosure obligations on metrics for notices processed by automated means, any use of automated means for the purpose of content moderation, and information on the type of content moderation engaged by providers of online services.⁸⁸ Furthermore, the Regulation provides for procedural measures for users to dispute content blocking or removals of information labelled as illegal content or against the ToUs of the platform.⁸⁹
- 42 While these measures purport a robust layer of protection for online users' fundamental rights, the fact remains that AI based filtering tools are far from being perfect and the technical challenges of these tools cannot be resolved only with transparency obligations on online intermediaries. There are issues that are yet to be addressed: the algorithmic fairness and human bias on data sets, accuracy standards since for certain type of content, a higher rate of accuracy may be easier to achieve⁹⁰, although that would not be the case when context is an intrinsic factor for determining illegality of a piece of content. In addition, it should not be assumed that online intermediaries are best placed to assess the legality or illegality of content, and they should not be seen as neutral when making such decisions.⁹¹ With that in mind, there will always be "grey zone" cases which would require human judgement rather than an AI based system taking a decision.⁹² It is particularly in those situations where online intermediaries may feel compelled to over-block to not risk liability.
- 43 If despite the flaws, we take AI based filters as a "necessary evil" for content moderation, then, closer regulatory scrutiny should be paid to their design, implementation, and the consequences of their use on public speech and the fundamental rights of online users but also the role and responsibilities of online intermediaries. To that end, data sets to build automated AI systems should be documented and traceable.⁹³ Guidelines on content moderation automated systems, including accuracy and error thresholds could be adopted to complement the DSA. Human review should continue to be an essential component of moderating with filters. Some authors postulate that automated decision-making processes should be subject to the "human-in-command" principle, namely, human intervention to supervise the overall activity of the AI system, its impact, and the ability to decide when and how to use the system.⁹⁴ By the same token, there should be clear accountability, liability, and redress mechanisms to deal with potential harm resulting from using applications, automated decision-making and machine learning tools.⁹⁵ Other propositions advocate for AI ethical principles specific for content filtering⁹⁶ which could form part of a regulatory framework to ensure compliance and enforceability. Such a framework should include the setting up of
-
- 87 Giancarlo Frosio, and Christophe Geiger, Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime. *European Law Journal* (forthcoming 2022) p.30 <<http://dx.doi.org/10.2139/ssrn.3747756>>.
- 88 Article 15, 24, and the provisions addressed to very large platforms (VLOPs) read in conjunction with Recital 39.
- 89 Articles 20 & 21 DSA.
- 90 Sartor and Loreggia (n 50) p.56.
- 91 Guidance note on content moderation, adopted by the Steering Committee for Media and Information Society (CDMSI) at its 19th plenary meeting, Council of Europe 19-21 May 2021 p.13.
- 92 Ibid (n 90) p.57.
- 93 André Tambiama Madiega, EU guidelines on ethics in artificial intelligence: Context and implementation, European Parliamentary Research Service, (2019) p.4. < EU guidelines on ethics in artificial intelligence: Context and implementation (europa.eu)>.
- 94 Frosio and Geiger (n 87) p.43.
- 95 Ibid. Moreover, with the new proposal for an AI liability Directive, it is to be seen how this liability regime could apply to the use AI tools in the context of content moderation. See Proposal for a Directive of the European Parliament and the Council on adapting civil liability rules to artificial intelligence. <IMMC.COM%282022%29496%20final.ENG.xhtml-ml.1_EN_ACT_part1_v10.docx (europa.eu)>
- 96 Marsoof et al. (n 53) p.22.

mandatory certification standards for testing AI systems to ensure they meet minimal safety and accuracy requirements.⁹⁷ This is in line with the approach of the EU Artificial intelligence Act (AIA).⁹⁸ The proposal sets a horizontal legal framework for the development, placement on the market, and use of AI applications in the Union, based on a risk-based approach. Under the current form of the proposal, there is no specific reference to AI systems for content moderation purposes. However, it has been argued that the AIA could be the right place to regulate the use of upload filters.⁹⁹

- 44 As a closing remark, one should also not lose sight of technological innovation to question if other options to automated tools managed by online platforms are possible. In its report on combating online harms through innovation, the American Federal Trade Commission listed user tools in its recommendations to tackle harmful content.¹⁰⁰ These tools could help users to control what content they see on the internet, shifting the content moderation effort from private platforms towards users. This is the idea of the so-called middleware for content moderation services. Middleware in this context, is a software program that rides on top of an existing internet or social media platform such as Google, Facebook or Twitter and can modify the presentation of underlying data.¹⁰¹ Middleware can be understood as a layer between the user and the online platform. By relying on this software, users could control their experience in a relevant platform but at the same time have the option to interact with other users of the online platform.¹⁰² The development of these

tools is still at a very early stage, and for now aimed mainly to the specifics of legal but harmful content. Nevertheless, there is room to consider if similar AI initiatives could provide effective alternatives to automated filters for fighting illegal content.

F. Conclusion

- 45 AI based filtering tools have become an integral part of content moderation. Although these technologies have been used until now as voluntary measure to fight illegal and harmful content, the new EU regulatory framework may be implicitly requiring online platforms to rely on them to escape liability. The EU legislator should not ignore the technical developments in the field and the current practices in content moderation carried out by online intermediaries, albeit regulatory efforts must ensure that the benefits of deploying and using these technologies to fight illegal and harmful content are not hampering the fundamental rights of online users. Accordingly, further guidance on human intervention and what entails human review should be provided. AI-based filtering tools should be designed, developed, and deployed when they meet certain safety and quality performance criteria. Accuracy standards and error rate thresholds must be established to ensure predictability and most importantly, a clear role responsibility and a reparation framework should be established to enable online users to seek redress from harms arising from the malfunction of filters.

97 Ibid p.28.

98 Proposal for a Regulation laying down harmonised rules on artificial intelligence, COM/21/206 FINAL (2021/0106(COD))

99 See for example Martin Husovec, *Euroactiv* “Internet filters do not infringe freedom of expression if they work well. But will they?” (2 May 2022) <Internet filters do not infringe freedom of expression if they work well. But will they? – EURACTIV.com> accessed 14 July 2022.

100 Federal Trade Commission Report to Congress: Combatting Online Harms Through Innovation, June 2022 <Combatting Online Harms Through Innovation; Federal Trade Commission Report to Congress (ftc.gov)>.

101 Katharine Miller “Radical proposal, Middleware could give consumers choices over what they see online” Stanford University (20 October 2021) <<https://hai.stanford.edu/news/radical-proposal-middleware-could-give-consumers-choices-over-what-they-see-online>> accessed 24 August 2022.

102 Daphne Keller (Blogpost The University of Chicago Law Review Online. 28 07 2022) Lawful but Awful? Control over Le-

gal Speech by Platforms, Governments, and Internet Users – The University of Chicago Law Review Online (uchicago.edu).

ALLEA Statement on Open Access Publication Under “Big Deals” and the new Copyright Rules

The increasing costs of publication under the Gold Open Access model and “Big Deals”

by ALLEA*

© 2023 ALLEA

Everybody may disseminate this article by electronic means and make it available for download under the terms and conditions of the Digital Peer Publishing Licence (DPPL). A copy of the license text may be obtained at <http://nbn-resolving.de/urn:nbn:de:0009-dppl-v3-en8>.

Recommended citation: ALLEA, Statement on Open Access Publication Under “Big Deals” and the new Copyright Rules, 17 (2023) JIPITEC 222 para 1.

The European Federation of Academies of Sciences and Humanities (ALLEA) has for many years supported the move away from proprietary models of scientific publishing towards Open Access (OA).¹ OA publication of publicly funded scientific research bears the triple promise of (1) fostering access to published research and knowledge by researchers, and the general public, all over the world; (2) recognising that outputs derived from publicly funded research are essentially a public good; and (3) reducing the mounting costs of accessing published research for universities and other academic institutions.

ALLEA, therefore, welcomes recent studies showing that OA publication in scientific journals is on the rise.² An

* The European Federation of Academies of Sciences and Humanities (ALLEA).

1 See, for example: <https://allea.org/portfolio-item/allea-response-to-plan-s/>; <https://allea.org/portfolio-item/ethical-aspects-of-open-access-a-windy-road/>; <https://allea.org/portfolio-item/allea-statement-on-enhancement-of-open-access-to-scientific-publications-in-europe/>

2 Zhang, L., Wei, Y., Huang, Y. et al. “Should open access lead to closed research? The trends towards paying to perform research”. *Scientometrics* (2022): <https://doi.org/10.1007/>

important driver of this development is the so-called “Big Deals”; “read and publish agreements” that have been negotiated in recent years between (consortia of) research libraries, institutions, and universities on the one hand, and scientific publishers on the other. These agreements, also known as “transformative agreements”, have replaced the subscription deals that were previously agreed between research libraries and publishers, and which provided for large bundles of subscriptions to proprietary journals to be made available electronically to libraries and their affiliated researchers.³

The new generation of deals is “transformative” in that they additionally allow for OA publication under the “Gold” standard of (usually a finite number of) research articles by institution-affiliated researchers in return for payment of substantial “article processing charges” (APCs)³ that allow publishers to recoup their investment in OA publication.

s11192-022-04407-5

3 European University Association “2019 Big Deals Survey Report - An Updated Mapping of Major Scholarly Publishing Contracts in Europe” (2019): <https://eua.eu/downloads/content/2019%20big%20deals%20report%20v2.pdf>

As a recent study demonstrates, commercial publishers currently derive more than two billion USD annually from Author Processing Charges (APCs).² Despite gradually decreasing subscription revenues, the commercial publishers have managed to embrace the Gold OA model without compromising their total revenues and enormous profit margins. Evidently, Gold OA publishing has become a new, highly profitable business model in and of itself,² in addition to the subscription model which has remained partially intact. Incorporating Gold OA publication into all-encompassing read and publish agreements has thus allowed the major commercial publishers to effectively consolidate and enhance their already dominant position in the field of scholarly publishing,⁴ solidifying their role as the gatekeepers of publicly funded research.⁵

While the rising number of Gold OA publications facilitated by these deals is to be applauded, they do not deliver on the triple promise of OA. In particular, they have not led to a reduction in the exorbitant costs to the academic community incurred in the process of research publication. While the downstream costs of journal subscriptions are gradually falling, the upstream costs of publication, made up of the APCs, have risen sharply.

Concomitantly, the imposition of APCs has created new, and sometimes insurmountable, barriers to publication for researchers that are not affiliated to a contracting institution.⁶ In addition, as already underlined in previous ALLEA Statements,^{6,7} the Gold OA model creates a disadvantage for those coming from less wealthy countries and institutions, under-funded researchers in the social sciences and humanities, and early career researchers, among others. For these academics, OA of published research comes at the expense of closure of first-tier publication fora.

In addition, ALLEA is concerned that the conditions of the “Big Deals” that drive these developments do not adequately reflect the rules on copyright law in the

European Union (EU) and fail to fairly value the creative and research endeavours of researchers and their institutions, as well as their investment and efforts over time to generate research results and publications to the benefit of the public.

A. The new copyright rules relevant to “Big Deals”

Under the law of copyright, the authors of works of science are the copyright owners of their published articles. Unless these rights are contractually assigned or licensed, it is for the authors, and the institutions that employ them (not for the publishers), to determine the conditions under which their works are to be published, reproduced, and otherwise used (including by way of OA).

In current practice, authors are expected to assign or exclusively license their copyright to publishers. Under the new rules of the 2019 Directive on Copyright in the Digital Single Market, which have been recently implemented in most EU Member States, authors that license or assign their rights “for the exploitation of their works” are entitled to receive appropriate and proportionate remuneration,^{8a} except where they have granted OA licences.^{8b} Ordinary publishing contracts between authors and publishers on which the “Big Deals” largely rely, however, rarely, if ever, provide for such remuneration. To the contrary, researchers or their institutions are expected to remunerate the publishers through APCs for having their scientific research published.

In addition, various EU Directives allow Member States to provide for limitations and exceptions to copyright for the purpose of scientific research. For example, EU law allows Member States to exempt the reproduction and making available of works “for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author’s name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved”.^{9a} While not all Member States have implemented this provision, and modalities of implementation vary,^{5,10} downloading and sharing of

4 Frontiers. “It is not transformation if nothing changes” (2022): https://blog.frontiersin.org/wp-content/uploads/2022/06/Frontiers_transformative_agreements_whitepaper_2022.pdf

5 European Commission - DG for Research and Innovation “Study on EU copyright and related rights and access to and reuse of scientific publications, including open access” (2022): <https://op.europa.eu/en/publication-detail/-/publication/884062d5-1145-11ed-8fa0-01aa75ed71a1/>

6 ALLEA “Statement on Equity in Open Access” (2021): <https://allea.org/portfolio-item/equity-in-open-access/>

7 ALLEA “Statement on Enhancement of Open Access to Scientific Publications in Europe” (2013): <https://allea.org/portfolio-item/allea-statement-on-enhancement-of-open-access-to-scientific-publications-in-europe/>

8 (a) Art. 18, (b) Recital 74 and (c) Art.3 of the Directive on Copyright in the Digital Single Market, 2019/790: <https://eur-lex.europa.eu/eli/dir/2019/790/oj>

9 (a) Art. 5(3)a, (b) Art. 5(3)d and (c) Art. 5(3)n of the Information Society Directive, 2001/29/EC: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32001L0029>

10 Knowledge Rights 21 “A Position Statement from Knowledge Rights 21 on Secondary Publishing Rights” (2022): <https://www.knowledgerights21.org/wp-content/up->

articles for the purpose of conducting or producing scientific research is permitted without authorisation in many EU Member States. Where such limitations and exceptions exist, publishers that have acquired the copyrights may not subject the downloading of articles by researchers to licensing conditions and payment of licence fees, as the European Court of Justice (ECJ) has clarified in its case law.¹¹

Other relevant limitations and exceptions in EU law permit the use of “quotations for purposes such as criticism or review”^{9b} and the making available of articles on dedicated terminals in library networks.^{5,9c} Moreover, the 2019 Directive on Copyright in the Digital Single Market requires EU Member States to grant non-profit research institutions broad freedoms to reproduce works for the purpose of “text and data mining”.^{8c} Accordingly, publishers may not restrict or condition text and data mining from scientific journals to which the researchers have lawful access.

Additionally, an increasing number of Member States (e.g., Germany, Netherlands, Austria, France, and Belgium) have introduced special rules permitting researchers to reproduce and make available published articles in non-profit repositories, regardless of having transferred their rights to publishers.¹² These so-called Secondary Publication Rights allow authors of scientific works that are the product of fully or partially publicly funded research to provide Open Access to their articles, following the expiry of a variable embargo period set by national legislation or good practices.⁵ National rules vary as well in respect of the version of the article that is subject to the Secondary Publication Right. While some countries limit the right to the Author Accepted Manuscript, the law in other countries seems to extend the right to the printed version, the so-called Version of Record. In all countries, the right is limited to articles; entire monographs and other scholarly books are therefore excluded.⁵

While ALLEA applauds the introduction of these new rights, we believe that, with the accelerated pace of scientific output and the need to adequately respond to today’s societal challenges, embargo periods are unnecessary impediments to the timely dissemination of publicly funded research. Today, as recently underlined

loads/2022/10/Secondary-Publishing-Rights-Position-Paper.pdf

- 11 Court of Justice of the European Union: Judgement of 27 June 2013, C-457/11 (VG Wort), ECLI:EU:C:2013:426. <https://ipcuria.eu/case?reference=C-457/11>
- 12 ALLEA “Supplementary Statement on Enhancement of Open Access to Scientific Publications in Europe” (2015): <https://allea.org/portfolio-item/supplementary-statement-on-enhancement-of-open-access-to-scientific-publications-in-europe/>

in a Guidance of the White House Office of Science and Technology Policy (OSTP) of 25 August 2022,¹³ research that is made “widely available to other researchers and the public (...) can save lives, provide policymakers with the tools to make critical decisions, and drive more equitable outcomes”, and therefore “there should be no delay or barrier” for the research outcomes to be made available to the public which has funded this research. ALLEA agrees, and therefore favours copyright rules that allow for OA publication of (partially) publicly funded research with immediate access and no embargo.

While EU and national copyright laws provide for a variety of rules intended to facilitate the free use and sharing of scientific works, without the need to compensate copyright holders, the current “Big Deals” do not generally factor in these statutory free uses.

Admittedly, the value added to the scientific article during its journey from submission to final publication is the result of a review and editing process that deserves financial reward. However, much if not most of this work (such as peer-reviewing and journal editing) is outsourced by the publishers to members of the academic community directly affiliated to institutions that are also parties to the agreements. To better judge the added value provided by publishers, there is a need for greater transparency on the pricing of journal publishing services and fees, and developments like the cOAlition S “Journal Comparison Service” are to be welcomed.¹⁴

All in all, it is difficult to see why an overall licensing agreement between research institutions representing the authors of thousands of publicly funded works, allowing affiliated researchers to publish and access the products of their own research or their fellow researchers’ endeavours, would justify payment of “read and publish” fees in the order of magnitude of the present “Big Deals”.

B. Recommendations

1. Negotiate future deals considering national and EU copyright law.

Now that the first generation of “Big Deals” is soon to expire, ALLEA recommends that research institutes and affiliated authors reconsider the terms of these agreements. In particular, ALLEA advises negotiators on the part of the research community to better leverage the

- 13 OSTP Issues Guidance to Make Federally Funded Research Freely Available Without Delay: <https://www.whitehouse.gov/ostp/news-updates/2022/08/25/ostp-issues-guidance-to-make-federally-funded-research-freely-available-without-delay/>
- 14 cOAlition S - Journal Comparison Service: <https://www.coalition-s.org/journal-comparison-service/>

rights and limitations accorded to authors and research institutions under national and EU copyright law, in order to further enhance the possibilities of (immediate) OA publication and substantially reduce the costs of APCs and journal subscriptions. ALLEA is concerned that if researchers perceive present and future “Big Deals” as vehicles that further strengthen and enrich the scientific publishing oligopoly, their willingness to permit OA publishing will dissipate.

2. Move away from the current rights assignment models.

Future “Big Deals” should pave the way for a future of scientific publishing where publicly funded research is freely available from multiple competing platforms, whether operated for profit or not-for-profit, including platforms operated by the research community itself.¹⁵ Therefore, future deals with scientific publishers should depart from the rights assignment model that still prevails today. Rather than forcing authors to individually negotiate with publishers, universities, and other research institutions might consider reserving certain rights in employee-produced publications to themselves, for example, by way of (collectively bargained) labour agreements. In addition, funding organisations should ensure that all researchers participating in the research they fund commit to publishing the research outcomes under an OA model that does not impose APCs or embargos.

3. Harmonise EU legislation to allow publication of post-print versions without embargo.

ALLEA recommends that national legislatures follow the example of an increasing number of European states in providing for Secondary Publication Rights that give researchers the right to make the post-print version (i.e., the Version of Record) of articles that are the product of fully and partially publicly funded research available in public repositories without embargo. Authors of scholarly books, scholarly book chapters, and edited research books should also be encouraged to publish their work in OA where reasonably possible. Ideally, such Secondary Publication Rights should be harmonised and made mandatory at the EU level.^{5,10} In doing so, the EU would set an important step towards operationalizing the 2018 European Commission Recommendation, which advised that all scientific publications resulting from publicly funded research be available OA by 2020,¹⁶ while refraining from creating new barriers for authors.

15 For example, Latin America has demonstrated for many years that an OA system based around federated institutional repositories works very well, and inspiration should be drawn from initiatives like Redalyc and SciELO.

16 European Commission Recommendation on access to and preservation of scientific information, 2018/790. <https://eur-lex.europa.eu/eli/reco/2018/790/oj>

Additionally, with a view to international collaborations that go beyond the EU, further efforts should be made to harmonise Secondary Publication Rights globally.

4. Develop a sustainable non-profit publishing ecosystem.

Finally, ALLEA recommends that research institutions and funding organisations prioritise the development of a sustainable non-profit publishing ecosystem that allows for OA of scientific publications without imposing undue financial barriers to publication, and that prevents scarce financial resources from being syphoned off by the private sector.³ The development of community-driven journals that charge no fees to authors and readers (Diamond OA) are an important contribution to a more equitable publishing landscape and an enrichment in bibliodiversity. ALLEA therefore welcomes and supports the Action Plan for Diamond Open Access that was published in March 2022.¹⁷

About ALLEA

ALLEA is the European Federation of Academies of Sciences and Humanities, representing more than 50 academies from over 40 countries in Europe. Since its foundation in 1994, ALLEA speaks out on behalf of its members on the European and international stages, promotes science as a global public good, and facilitates scientific collaboration across borders and disciplines. Learn more: www.allea.org

About this Statement

This ALLEA statement has been prepared by ALLEA’s Permanent Working Group on IPR, with Prof P. Bernt Hugenholtz as principal author. Through its Working and Expert Groups, ALLEA provides input on behalf of European academies to pressing societal, scientific, and science-policy debates and their underlying legislations. With its work, ALLEA seeks to ensure that science and research in Europe can excel and serve the interests of society. Read more about the ALLEA Permanent Working Group on IPR and its members: <https://allea.org/intellectual-property-rights/>

17 Science Europe, cOAlition S, OPERAS, and the French National Research Agency (ANR). “Action Plan for Diamond Open Access” (2022): <https://www.scienceurope.org/media/t3jgyo3u/202203-diamond-oa-action-plan.pdf>