

ALGORITHMIC ENFORCEMENT OF COPYRIGHT: APPROACHES TO TACKLING CHALLENGES POSED BY UPLOAD FILTERS

PHILIPP HOMAR^{*}

It is a widespread phenomenon that online platforms which enable users to upload copyright-protected content use software-based tools that automatically identify and restrict copyright-infringing user uploads based on algorithms – so-called upload filters. The European Union is currently at the forefront of the development to govern upload filters through legislation, both with regard to requiring platforms to filter infringing content (Art 17 of the Directive on Copyright in the Digital Single Market) and with regard to mitigating risks of automated filtering (especially through the newly adopted Digital Services Act). Against this background, this article analyses the European approach to algorithmic enforcement of copyright and compares it with the situation in the United States. After laying out the legal foundations from which the necessity to engage in algorithmic enforcement results, this article focuses particularly on legal approaches to avoiding the restriction of permissible content (“overblocking”).

I. INTRODUCTION

Recent years have brought a change in how technology shapes the dissemination and consumption of media content. While the widespread availability of digital devices and the emergence of online platforms that allow users to share content with a global audience initially increased the ability of users to access and share media content online, digital tools are now also being used to restrict this ability in relation to copyright-protected content. This development is epitomised by so-called upload filters: software-based tools that automatically detect and block copyright-protected content from being uploaded by users without authorisation of rightsholders based on algorithms. Although some online platforms have long been using automated filtering tools¹, they have recently gained significance within the EU. Art 17 of the Directive on Copyright in the Digital Single Market (“DSM-Dir”)² introduced a new regime of liability for certain online platforms that entails the use of upload filters. This was probably the most debated development in recent copyright history

^{*} Prof. Dr. iur., Chair of Intellectual Property at Johannes Kepler University Linz (Austria) and Vienna University of Economics and Business (Austria), University of Continuing Education Krems (Austria).

¹ Google, “How Content ID works” <<https://support.google.com/youtube/answer/2797370?hl=en>> [29 June 2023].

² Directive on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, European Parliament and Council Directive 2019/790, [2019] OJ L 130/92 [“DSM-Dir”].

that resulted from the discussion of whether revenues generated by platforms are being distributed fairly (“value gap”).³

The goal of this article is not to evaluate whether the shift towards algorithmic enforcement in copyright law is desirable or not; rather, it acknowledges that copyright law (especially European copyright law) has embraced the concept of algorithmic enforcement, and focuses on the legal framework that governs the use of filtering tools. In the following, two aspects are placed at the center of attention: First, it is analysed to what extent online platforms are required to use filtering tools to detect and block infringing uploads. Second, legal approaches to mitigating the risk of over-enforcing copyright are discussed. As the EU is currently at the vanguard of algorithmic enforcement, the analysis focuses on the European legal framework. Furthermore, commonalities and differences between the EU and the US are mapped. Against this background, the remainder of this article is organised as follows: Part II analyses the legal foundations that require platforms to use automated filtering tools to combat copyright infringement. Part III then focuses on the effects of filtering tools, particularly on the legal and societal risks of blocking permissible content (“overblocking”). In Part IV, approaches within US law and the European legal framework that provide safeguards against overblocking are identified and evaluated to determine how well they succeed in balancing the interests of rightsholders with the conflicting interests of users. It is also discussed whether these approaches, and especially those in the EU, successfully mitigate concerns raised around algorithmic enforcement. Part V concludes the article.

II. THE LEGAL FRAMEWORK THAT REQUIRES ALGORITHMIC ENFORCEMENT

Although algorithmic enforcement of copyright seems to be a global trend, the legal frameworks that govern the use of filtering software differ fundamentally.

A. *US: The Notice and Takedown System*

From the perspective of US copyright law, algorithmic enforcement is rooted in the “notice and takedown” (“NTD”) system provided in the Digital Millennium Copyright Act (“DMCA”).⁴ According to 17 U.S.C. § 512(c), service providers are shielded from liability for copyright infringements committed by their users (“safe harbour”) if they react expeditiously to notifications of infringements by removing, or disabling access to, infringing content.⁵ This system of enforcement can be

³ Giancarlo Frosio, “Reforming Intermediary Liability in the Platform Economy: A European Digital Single Market Strategy” (2017) 112 NWULR 19 at 26 <<https://ssrn.com/abstract=3009155>> [28 July 2023].

⁴ Digital Millennium Copyright Act, Pub L No 105-304, 112 Stat 2860, Online Copyright Infringement Liability Limitation Act, Title II [1998]; Maayan Perel & Niva Elkin-Koren, “Accountability in Algorithmic Enforcement” (2016) 19 Stanford Technology Law Review 473 at 477 <<https://law.stanford.edu/wp-content/uploads/2016/10/Accountability-in-Algorithmic-Copyright-Enforcement.pdf>> [28 July 2023] [“Perel & Elkin-Koren”].

⁵ 17 U.S.C. § 512(c)(1)(C).

described as a two-stage and two-person process.⁶ In the first stage, it is the responsibility of rightsholders to identify infringing uploads of their works and to notify a platform provider; in the second stage, it is the responsibility of the platform provider to review notifications and to act expeditiously by blocking access to or removing notified content if it infringes copyright. Although the legal framework outlines the NTD process in considerable detail,⁷ it does not explicitly require the individual steps to be handled by automated tools. In practice, however, both stages of the NTD process are largely automated. Rightsholders regularly use digital tools that detect online uses of their works and automatically send takedown notices to platforms. Platforms also handle these notices in a widely automated process in which algorithms implemented in filtering tools determine whether notified content is removed or stays online.⁸ Efficiency is not the only reason for the parties involved to use digital tools that automate the NTD process. For platform providers, further incentives to use automated tools are provided by the legal framework, as they face liability unless they respond expeditiously to notices by blocking or removing infringing content. The high number of infringement notices being submitted to platforms, the need to act expeditiously and the threat of liability are therefore significant drivers of automation. Many platforms also go beyond the algorithmic implementation of the NTD system and voluntarily apply automated tools to block infringing content not only after notifications (*ex post*), but also before it is uploaded to the platform (*ex ante*).⁹

B. EU: From Injunctions to Primary Liability

The requirement for platforms to use automated filtering tools is not contained explicitly in any provision of European copyright law, but arises from financial risks to which platform providers are exposed if they do not effectively detect and filter infringing uploads. These financial risks result from platform providers being subject to injunctions or even primary liability.

1. Injunctions against platforms as intermediaries

Since the early 2000s, European copyright law has provided rightsholders with the ability not only to file claims against users who upload copyright-infringing¹⁰

⁶ Martin Husovec, “The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown? Which Is Superior? And Why?” (2018) 42 Columbia Journal of Law & The Arts 53 at 64 <<https://doi.org/10.7916/jla.v42i1.2009>> [27 April 2023] [“Husovec”].

⁷ For the requirements of a notification see 17 U.S.C. § 512(c)(3) and for the counter-notice (“put back”) procedure 17 U.S.C. § 512(g)(2)-512(g)(3).

⁸ Husovec, *supra* note 6 at 60; Perel & Elkin-Koren, *supra* note 4 at 477; see also United States Copyright Office, Section 512 of Title 17 (Report of the register of copyrights) (Washington D.C.: US Copyright Office, 2020) 10 (available at: <<https://www.copyright.gov/policy/section512/section-512-full-report.pdf>> [8 May 2023]).

⁹ Perel & Elkin-Koren, *supra* note 4 at 504; Husovec, *supra* note 6 at 58.

¹⁰ Publicly available user uploads are infringing when they contain protected third-party content and are not privileged by a statutory exception, for instance, as quotation, caricature, parody, pastiche, or as incidental inclusion (Directive on the harmonisation of certain aspects of copyright and related rights

content to platforms, but also to obtain injunctions against platform providers in their capacity as intermediaries.¹¹ Although it is not entirely settled, these injunctions may be obtained against platform providers even when they are shielded from liability under the safe-harbour provision¹² for hosting services.¹³ Once an injunction has been filed against a platform provider, it is obliged to cease and desist from further infringements by blocking access to, or removing, copyright-infringing content (takedown) and also to prevent further infringements of the same kind (staydown).¹⁴ Violations of these obligations can be sanctioned with financial penalties.¹⁵ The first part of this obligation (takedown) corresponds to the NTD regime of US copyright law, but the second part, which aims to prevent re-appearance of the infringing content in the same or similar form (staydown), goes further. When injunctions result in staydown obligations, they effectively impose duties on platforms to monitor users' upload activities and to filter out future infringing uploads.¹⁶ However, the scope of this duty to prevent further infringements of the same kind depends on the specific circumstances and is employed to varying degrees by national courts in the EU. Several courts have interpreted this duty somewhat broadly, framing the duty to prevent future infringements as not merely preventing future uploads of the same infringing media file by the same user, but as including an obligation to prevent uploads of media files with similar infringing content by

in the information society, European Parliament and Council Directive 2001/29, Arts 5(3)(d), 5(3)(i), [2001] OJ L 167/10 ["InfoSoc-Dir"]; DSM-Dir, *supra* note 2 Art 17(7)) or by an authorisation of rightsholders (CJEU Joined Cases C-682/18 and C-683/18 *Peterson v. Google/YouTube and Elsevier v. Cyando* [2021] ECLI:EU:C:2021:503 at [75]).

- ¹¹ InfoSoc-Dir, *supra* note 10 Art 8(3) and Directive on the enforcement of intellectual property rights, European Parliament and Council Directive 2004/48, Art 11, [2004] OJ L 157/45 ["Enforcement-Dir"]; Orit Fischman-Afori, "Online Rulers as Hybrid Bodies: The Case of Infringing Content Monitoring" (2021) 23 U. Pa. J. Const. L. 351 at 368 <<https://scholarship.law.upenn.edu/jcl/vol23/iss2/3>> [12 March 2023] ["Fischman-Afori"].
- ¹² Digital Services Act (Regulation on a Single Market For Digital Services and amending Directive 2000/31/EC), European Parliament and Council Regulation 2022/2065, Art 6, [2022] OJ L 277/1 ["DSA"], formerly e-Commerce Directive (Directive on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market), European Parliament and Council Directive 2000/31, Art 14 [2000] OJ L 178/1 ["E-Commerce-Dir"].
- ¹³ DSA, *supra* note 12 Art 6(4); E-Commerce-Dir, *supra* note 12 Art 14(3); CJEU Case C-484/14 *Mc Fadden v. Sony Music Entertainment Germany* [2016] ECLI:EU:C:2016:689 at [79]; Martin Senftleben, "Breathing Space for Cloud-Based Business Models" (2013) 4(2) JIPITEC 87 at 95. But see Martin Husovec, *Injunctions Against Intermediaries in the European Union – Accountable But Not Liable?* (Cambridge: Cambridge University Press, 2017) at 57 et seq. Whether injunctions require a prior notice depends on national legislation. Basically, EU member states are allowed to provide injunctions against platform providers that require no prior notice (CJEU Joined Cases C-682/18 and C-683/18, *supra* note 10 at [130] and [143]). However, national legal regimes often require a prior notice (eg Austrian Copyright Act § 81(1a)).
- ¹⁴ German Federal Court of Justice Case I ZR 140/15 *YouTube II* [2022] GRUR 2022, 1308 at [119] (in German language). For trademark infringements on online marketplaces, see CJEU Case C-324/09 *L'Oréal v. eBay* [2011] ECLI:EU:C:2011:474 at [144]; German Federal Court of Justice Case I ZR 57/09 *Stiftparfüm* [2011] GRUR 2011, 1038 at [21] (in German language).
- ¹⁵ Michel M Walter & Dominik Goebel, *European Copyright Law – A Commentary* (Michel M Walter & Silke von Lewinski, eds) (Oxford University Press, 2010) at 13.11.9.
- ¹⁶ German Federal Court of Justice Case I ZR 53/17 *uploaded II* [2022] GRUR 2022, 1324 at [34] (in German language); Giancarlo Frosio, "Algorithmic Enforcement Online" in *Intellectual Property Law and Human Rights* (Paul L C Torremans, ed) (4th Ed, 2020) <<https://ssrn.com/abstract=3503419>> ["Frosio"].

any user.¹⁷ This broad judicial construction of staydown obligations resulting from injunctions is strengthened by recent case law concerning defamatory postings on social networks (“hate speech”).¹⁸

2. Primary liability of online content-sharing service providers (Art 17 DSM-Dir)

Despite the scope of obligations resulting from injunctive relief, a debate took place in the EU as rightsholders demanded that the liability of platform providers for copyright infringing uploads be increased. Rightsholders argued, in essence, that injunctions made it impossible for them to obtain appropriate remuneration from platform providers. Although some licence agreements could be reached also under these conditions,¹⁹ they led, according to the rightsholders, to the so-called value gap.²⁰ Ultimately, the lobbying by rightsholders was successful and, despite numerous concerns raised by representatives of users and platforms, resulted in a strengthening of platform providers’ liability. The epitome of this development was Art 17 of the DSM-Dir, with which the European legislature introduced a sector-specific liability concept that applies to platforms qualifying as online content-sharing service providers (“OCSSPs”).

(a) *Scope of application:* According to Art 2(6) of the DSM-Dir, an OCSSP is „a provider of an information society service of which the main or one of the main purposes is to store and give the public access to a large amount of copyright-protected works [...]”²¹ uploaded by its users, which it organises and promotes for profit-making purposes.”²² Because of this definition, platforms that do not fulfil the quantity criterion (*ie* small platforms) and/or the purpose criterion (*ie* non-commercial platforms) remain outside the scope of Art 17 of the DSM-Dir. The same applies to platforms that are specifically excluded (*eg* not-for-profit online encyclopedias,

¹⁷ See German Federal Court of Justice Case I ZR 140/15, *supra* note 14 at 1321.

¹⁸ Both the CJEU and the Austrian Supreme Court of Justice decided that duties of platforms to prevent future defamatory postings relate not only to further postings by the same user, but also to postings of other users and, furthermore, not only to postings with the same wording, but also to postings with the same meaning (CJEU Case C-18/18 *Glawischmig-Piesczek v. Facebook Ireland* [2019] ECLI:EU:C:2019:821; Austrian Supreme Court of Justice Case 4 Ob 36/20b *Lügen zu Nachrichten* [2020] ECLI:AT:OGH0002:2020:0040OB00036.20B.0330.000 and Case 6 Ob 195/19y *miese Volksverräterin IV* [2020] ECLI:AT:OGH0002:2020:0060OB00195.19Y.0915.000 [both in German language]). The duty to prevent postings with the same meaning is restricted to postings conveying messages that remain essentially unchanged when compared to the content which gave rise to the finding of illegality; furthermore, these postings must contain elements specified in the injunction that can be detected with automated search tools.

¹⁹ See Yafit Lev-Aretz, “Second Level Agreements” (2012) 45 *Akron L R* 137 at 155 et seq <<https://ssrn.com/abstract=2131864>> [7 February 2023].

²⁰ See European Commission, *Impact Assessment on the Modernisation of EU copyright rules SWD/2016/0301 final* [2016] at point 5.2.

²¹ The same applies to “*other protected subject matter*” which refers to subject matter protected by neighbouring rights, such as rights of performing artists or of record producers. Whenever the term “work” is hereafter used, it also includes subject matter protected by neighbouring rights unless stated otherwise.

²² DSM-Dir, *supra* note 2 Art 2(6).

not-for-profit educational and scientific repositories, open source software-developing and -sharing platforms, online marketplaces).²³ The liability of platforms that do not fall under Art 17 of the DSM-Dir is governed by Art 3 of the Copyright and Information Society Directive (“InfoSoc-Dir”).²⁴

(b) *Primary liability*: Because of Art 17(1) of the DSM-Dir, OCSSPs perform an act of making available copyright-protected works to the public.²⁵ Therefore, if users upload protected works and they do so without an authorisation of rightsholders and without the benefit of a statutory exception (eg the exception for quotations, parodies, caricature and pastiche or incidental inclusions)²⁶, an OCSSP, in addition to the user,²⁷ infringes the rightsholders’ making available right and is not shielded by the safe harbour provision for hosting services.²⁸ The consequence is that an OCSSP is subject to primary liability.²⁹ Thus, rightsholders may file injunctions that result in cease and desist obligations, and also financial claims³⁰ against an OCSSP once infringing uploads of their works have become publicly available. Art 17 of the DSM-Dir therefore increases platform providers’ liability and financial risks when copyright-protected works are uploaded by users. Of course, liability only occurs when user uploads infringe the copyright of third parties. Although some users can hold licences (eg creative commons licences) and some uses on platforms, especially transformative ones (eg reviews, tutorials, and memes), may indeed qualify as quotations or parodies, it is safe to assume that a vast number of uploads will not be covered by any of these privileges, especially as European copyright law, unlike those of the US and Singapore, does not contain a general fair use exception.³¹ In particular, uploads of entire works without modifications (eg when entire films or songs are uploaded, and this includes an amateur home video to which a song was added) will generally trigger primary liability.

(c) *Avoiding liability by filtering*: OCSSPs can avoid liability by concluding licence agreements with rightsholders whose works are uploaded. In this case, no copyright infringement occurs when users upload works covered by licence agreements.³² However, it is evident that OCSSPs do not know in advance what

²³ DSM-Dir, *supra* note 2 Art 2(6).

²⁴ See Part II.B.2 below.

²⁵ Although DSM-Dir, *supra* note 2 Art 17(1) is not entirely clear, this refers to the making available right as provided for in InfoSoc-Dir, *supra* note 10 Art 3 (*contra* Martin Husovec and João Pedro Quintais, “How to License Article 17? Exploring the Implementation Options for the New EU Rules on Content-Sharing Platforms under the Copyright in the Digital Single Market Directive” (2021) 70(4) GRUR Int. 2021, 325 <<https://ssrn.com/abstract=3463011>> [6 February 2023]).

²⁶ *Supra* note 10.

²⁷ The infringement committed by platform providers occurs in addition to the infringement committed by the users.

²⁸ DSM-Dir, *supra* note 2 Art 17(3) excludes the applicability of the safe-harbour provision.

²⁹ Christophe Geiger and Bern Justin Jütte, “Platform Liability Under Art. 17 of the Copyright in the Digital Single Market Directive, Automated Filtering and Fundamental Rights: An Impossible Match” GRUR Int. 2021, 517 at 530 (pre-print available at SSRN: <<https://ssrn.com/abstract=3776267>> [6 February 2023]).

³⁰ Especially claims for damages (Enforcement-Dir, *supra* note 11 Art 13).

³¹ InfoSoc-Dir, *supra* note 10 Art 5 (and DSM-Dir, *supra* note 2 Art 3 et seq and Art 17(7)) only provide statutory exceptions for specific uses.

³² The permissibility applies to both the platform providers and the uploading users, at least when they act on a non-commercial basis or when they do not generate significant revenues (DSM-Dir, *supra* note 2 Art 17(2)).

works will be uploaded, and, even if they did, it would be impossible for them to conclude licence agreements with *all* rightholders.³³ In particular, rightholders with small repertoires (*eg* independent music labels and independent artists) may be difficult to identify, and negotiating licence agreements with all of them would impose prohibitive transaction costs. Therefore, Art 17(4) of the DSM-Dir provides three criteria that must be met by an OCSSP in order to avoid liability. One of these criteria is that the provider has “*made, in accordance with high industry standards of professional diligence, best efforts to ensure the unavailability of specific works [...] for which the rightholders have provided [...] the relevant and necessary information*”.³⁴ These best efforts, which have to be made once an OCSSP has received information concerning works (*ie* reference files of films, videos or songs) that should not be made available, relate to the removal of, or the blocking of access to, infringing content that has already been uploaded (*ex post*) and to the prevention of future infringing uploads (*ex ante*).³⁵ As this duty to demonstrate best efforts is not triggered by receiving a notification of an infringement, but rather by receiving relevant information on works that should not be made available, two facts become evident. First, efforts must be made even before an infringement has occurred on a platform. Second, the efforts to prevent and to block or remove relate not to specific infringements committed by specific users, but to all uploads from all users that infringe the copyright in the work for which relevant information has been provided. This expands the duties resulting from the European system beyond the reaction-based duties resulting from NTD regimes, such as the one provided by US copyright law.³⁶

Although the provision does not determine which specific measures an OCSSP has to apply (“best efforts”) and the entire obligation is subject to proportionality,³⁷ it is evident that software-based filtering tools that automatically detect and block copyright-infringing uploads will typically be required.³⁸ In view of the large number of uploads, it is simply not conceivable that OCSSPs could fulfil this obligation in any other way, especially given the requirement to meet *high* industry standards of professional diligence. Because software-based filtering tools (*eg* YouTube’s Content ID) have been in use for several years, they undoubtedly constitute high industry standards and thus, ultimately, the implicit goal of Art 17(4)(b) of the DSM-Dir.

3. Primary liability of other platforms (Art 3 InfoSoc-Dir)

The liability regime provided by Art 17 of the DSM-Dir does not affect platforms that do not qualify as OCSSPs, for instance, small platforms and non-commercial

³³ This was also acknowledged by the CJEU (Case C-401/19 *Republic of Poland v. European Parliament and Council of the European Union* [2022] ECLI:EU:C:2022:297 at [48]).

³⁴ DSM-Dir, *supra* note 2 Art 17(4)(b).

³⁵ DSM-Dir, *supra* note 2 Recital 66: “to avoid the availability”; CJEU Case C-401/19, *supra* note 33 at [53].

³⁶ Under the US system, measures to prevent infringing content before it becomes available are voluntary. See Perel & Elkin-Koren, *supra* note 4 at 504; Husovec, *supra* note 6 at 58.

³⁷ DSM-Dir, *supra* note 2 Art 17(5).

³⁸ CJEU Case C-401/19, *supra* note 33 at [54]; Fischman-Afori, *supra* note 11 at 372. However, an exception applies to “start-ups” according to DSM-Dir, *supra* note 2 Art 17(6).

platforms.³⁹ However, the principle that providers of platforms not covered by Art 17 of the DSM-Directive continue to be subject only to injunctions has been reassessed in recent case law on the making available right as provided in Art 3 of the InfoSoc-Directive. In a dispute involving a video-sharing platform and a file-hosting platform, the Court of Justice of the EU (“CJEU”) clarified that the making available right of Art 3 of the InfoSoc-Directive does apply to a platform if its provider “*contributes, beyond merely making that platform available, to giving access to such content to the public in breach of copyright.*”⁴⁰ This can be described as a “platform plus” approach. If platform providers go beyond providing the platform as infrastructure and make further contributions to works becoming publicly available, the making available right applies to them. A relevant contribution exists, *inter alia*, when a platform provider knows, in a general sense, that users commit copyright infringements and “*refrains from putting in place the appropriate technological measures that can be expected from a reasonably diligent operator in its situation in order to counter credibly and effectively copyright infringements on that platform.*”⁴¹ This clearly paves the way for automated filtering tools becoming relevant also to platforms that do not fall under Art 17 of the DSM-Directive. The requirement to have general knowledge of copyright-infringing user activities is extensive⁴² and triggers the duty to take (appropriate) technical measures to counter copyright infringements on that platform.⁴³ These measures are, again, subject to proportionality;⁴⁴ however, the reference to the technical nature and the requirement to combat infringements credibly and effectively undoubtedly hint at which measures are actually meant by the court.⁴⁵

III. LEGAL AND SOCIETAL RISKS OF ALGORITHMIC ENFORCEMENT

It is evident that detecting and blocking user uploads by automated means has significant advantages over manual processes. For rightsholders, it is faster, more efficient and cheaper than intervening manually (*eg* by filing takedown notices or

³⁹ See also Part II.B.1.

⁴⁰ CJEU Joined Cases C-682/18 and C-683/18, *supra* note 10 at [102].

⁴¹ *Ibid* at [102].

⁴² It does not require specific knowledge of the infringing quality of a particular content (*eg* of a specific video).

⁴³ What remains debatable is whether, in such a general knowledge scenario, a platform provider is also excluded from the safe-harbour provision for hosting services as set out in DSA, *supra* note 12 Art 6. The CJEU is not entirely clear in this respect, as it indicated at one point that a platform provider that contributes beyond merely providing its platform is not covered by the safe-harbour provision, whereas it later concluded that specific knowledge (*ie* knowledge of specific infringements) is required for a platform provider to be excluded from the safe-harbour provision (CJEU Joined Cases C-682/18 and C-683/18, *supra* note 10 at [108] and [118]).

⁴⁴ This results from the reference to the appropriateness of the measures and the role model of a reasonably diligent platform provider.

⁴⁵ Following up on the CJEU decision, the German Federal Court of Justice decided that merely reactive measures (*eg* a button to report infringing videos) are not sufficient to credibly and effectively combat copyright infringements and that proactive technical measures that prevent copyright infringements even without prior notice from rightsholders are required (German Federal Court of Justice Case I ZR 140/15, *supra* note 14).

injunctions) in each individual case. For platform providers, although filtering tools incur significant costs, they also provide efficiency advantages, especially since the resources to review content manually would be prohibitive.

Although filtering tools have clear advantages, the question arises whether they are (yet) fully fit for purpose. Specifically, there is the risk of filtering algorithms detecting and blocking not only copyright-infringing content, but also content that is lawfully uploaded to platforms (“overblocking”). From the perspective of copyright, this concerns, for example, user uploads that do not contain protected third-party content and uploads of third-party content in which copyright has expired. Of course, uploading protected third-party content to platforms is also permissible if it falls under one of the exceptions provided by copyright law itself. Although these exceptions vary considerably across copyright regimes – not only across countries with different legal traditions, but also within the EU⁴⁶ – at least quotational uses are permissible under almost all of them.⁴⁷ Likewise, parodies are privileged in many copyright regimes, for example, as fair use⁴⁸ under US copyright law or under the exception⁴⁹ for parody, caricature or pastiche in EU member states.

When uploads to platforms are permissible on any of the above-mentioned grounds, any prevention or removal would not only mean that copyright were unjustifiably enforced, but also that fundamental rights might be impaired. From the perspective of fundamental rights in the EU, the freedom of expression⁵⁰ and the freedom of the arts⁵¹ can be affected; regarding platform providers, the freedom to conduct a business⁵² may be curtailed.⁵³ It is therefore essential that automated filtering tools strike a balance between the interests of rightsholders (that infringements are detected effectively and blocked)⁵⁴, platform providers (that filtering tools do not jeopardise legitimate business models) and users (that permissible uploads are not affected). Interference with these interests might not only result in violation of fundamental rights, but also pose risks to society, as restriction of content that falls under the freedom of expression may interfere with social discourse, political debate and, ultimately, democratic decision-making.

⁴⁶ Copyright exceptions are not comprehensively harmonised within the EU. InfoSoc-Di, *supra* note 10 Art 5 provides EU member states with a closed list of limitations or exceptions. However, only one exception (InfoSoc-Di, *supra* note 10 Art 5(1)) must be transposed, whereas all other exceptions are optional. Therefore, national copyright laws within the EU differ both in terms of which exceptions are transposed and (to some extent) in terms of the scope of the transposed exceptions (an overview can be found here: <<https://copyrightexceptions.eu/>> [8 February 2023]). The DSM-Di increased harmonisation, as it provides further limitations and exceptions that must be transposed by member states.

⁴⁷ Berne Convention for the Protection of Literary and Artistic Works (9 September 1886), 828 UNTS 221, Art 10 (entered into force 5 December 1887) (“Berne Convention”). EU member states are required to transpose an exception for quotational uses on OCSSPs (DSM-Di, *supra* note 2 Art 17(7)(a)).

⁴⁸ See Niva Elkin-Koren & Orit Fischman Afori, “Rulifying Fair Use” (2017) 59 *Ariz L. Rev* 161 at 184.

⁴⁹ DSM-Di, *supra* note 2 Art 17(7)(b), which obliges EU member states to transpose an exception or limitation that allows users to upload works on OCSSPs for the purpose of caricature, parody or pastiche.

⁵⁰ Charter of Fundamental Rights of the European Union (7 December 2000), OJ C 326/391, Art 11 (entered into force 1 December 2009) (“CFREU”).

⁵¹ CFREU, *supra* note 50 Art 13.

⁵² CFREU, *supra* note 50 Art 16.

⁵³ Geiger & Jütte, *supra* note 29 at 523 et seq.

⁵⁴ These interests also enjoy fundamental rights protection under the right to property (CFREU, *supra* note 50 Art 17(2)) and the right to an effective remedy (CFREU, *supra* note 50 Art 47).

IV. STRIKING A BALANCE: APPROACHES TO AVOIDING OVERBLOCKING

Within the copyright systems of the US and the EU, different approaches to mitigating the risk of permissible uploads being affected by copyright-enforcing measures can be identified.

A. Approaches in the US

In the NTD regime provided by US copyright law, three aspects that serve as safeguards against non-infringing content being taken down stand out in particular.

1. Requirements for a notification of infringement

In order to become effective, a notification of an infringement must contain several elements. These include information to permit the platform to contact the complaining party and a statement that the complaining party is authorised to act on behalf of the owner of an exclusive right that is allegedly infringed.⁵⁵ Furthermore, the complaining party must provide a statement that it “*has a good faith belief that use of the material in the manner complained of is not authorized by the copyright owner, its agent, or the law.*”⁵⁶ These formal requirements are designed to filter unjustified takedown requests. First, they are intended to prevent takedown requests from persons who are not authorised to demand the removal and thus ensure the availability of content whose rightsholders approve of its use.⁵⁷ Second, they shall ensure that only infringing content is subject to takedown requests, although the relevant factor here is not the infringement per se, but the good faith belief of rightsholders that content they notify does indeed infringe on their rights. How strictly this good faith criterion is handled is crucial to the question of how effective the protection against unjustified takedowns is. In *Lenz v. Universal Music Corp* (“*Lenz*”), the US Ninth Circuit Court of Appeals strengthened the good faith criterion, as it concluded that rightsholders must consider whether content is privileged as fair use before they send takedown notices to platforms.⁵⁸ However, it would be premature to celebrate the ruling from *Lenz*, because according to an earlier decision of the Ninth Circuit in *Rossi v. Motion Picture Association of America*, the good faith criterion construes a subjective standard rather than a standard of objective reasonableness, and this was not overruled in *Lenz*.⁵⁹ Therefore, only knowing misinterpretations of uses will eliminate rightsholders’ good faith. This undoubtedly weakens the protection of

⁵⁵ 17 U.S.C. § 512(c)(3)(A)(iv) and 512(c)(3)(A)(vi).

⁵⁶ 17 U.S.C. § 512(c)(3)(A)(v).

⁵⁷ However, protection against takedown requests from unauthorised persons also means that infringing content of which the rightsholders are unaware remains available.

⁵⁸ *Lenz v. Universal Music Corp.* 801 F.3d 1126 (9th Cir. 2015). Specifically, the court held that because fair use was a type of non-infringing use, it is “‘authorized by the law’ and a copyright holder must consider the existence of fair use before sending a takedown notification”, at 1133 [“*Lenz v. Universal Music Corp.* 2015”].

⁵⁹ *Rossi v. Motion Picture Association of America Inc.* 391 F.3d 1000 (9th Cir. 2004).

users against unjustified takedown notices, although it can nevertheless be assumed that it is not sufficient if rightsholders only consider the legitimacy of a use (*eg* as fair use) as mere lip service.⁶⁰

2. Counter-notification procedure

Another safeguard against the removal of non-infringing content is the “put-back procedure” provided in § 512(g)(2), which enables users to submit counter-notices if their content was taken down. According to this procedure, platform providers first need to take reasonable steps to notify users that their content was taken down, and then users can contest this decision by filing counter-notifications.⁶¹ This adds a third stage to the NTD system, at which it is the users’ responsibility to react to unjustified takedowns and thus ensure that lawful content is not impaired. In this procedure, the responsibility of the platform provider is to forward the counter-notice to the person who requested the content to be taken down, and ultimately to reverse the takedown if this person does not notify the platform of the initiation of a lawsuit.⁶²

3. Liability for unjustified takedowns

If platform providers or rightsholders violate their responsibilities in the NTD process and lawful content is consequently taken down, they face liability for damages.⁶³ However, a closer look reveals that liability risks are restricted. According to *Lenz*, rightsholders are liable if they submit takedown notifications without considering fair use or if they knowingly misinterpret non-infringing (*ie* fair) uses as infringing.⁶⁴ At the same time, rightsholders are not liable simply because they unknowingly make a mistake in the evaluation of the legitimacy.⁶⁵ Similarly, platform providers are not liable towards users for erroneous removals of their content that were made in good faith.⁶⁶ Given that the qualification of uses as fair is highly case-dependent and difficult to predict, this represents an important limitation, as it effectively restricts liability to clear-cut cases in which content that clearly falls under fair use is interfered with.⁶⁷

⁶⁰ See Matthias Leistner, European Copyright Licensing and Infringement Liability Under Art. 17 DSM-Directive Compared to Secondary Liability of Content Platforms in the U.S. – Can We Make the New European System a Global Opportunity Instead of a Local Challenge? (2020) 12:2 *Zeitschrift für Geistiges Eigentum/Intellectual Property Journal* 123 (at 47 <<https://ssrn.com/abstract=3572040>> [29 June 2023] [“Leistner”]).

⁶¹ 17 U.S.C. § 512(g)(2)(A)-512(g)(2)(C).

⁶² 17 U.S.C. § 512(g)(2)(B)-512(g)(2)(C).

⁶³ 17 U.S.C. § 512(f), § 512(g)(1)-512(g)(2).

⁶⁴ See *Lenz v. Universal Music Corp.* 2015, *supra* note 58.

⁶⁵ *Lenz v. Universal Music Corp.* 2015, *supra* note 58; *Rossi v. Motion Picture Association of America Inc.* 391 F.3d 1000 (9th Cir. 2004).

⁶⁶ 17 U.S.C. § 512(g)(2); Perel & Elkin-Koren, *supra* note 4 at 498.

⁶⁷ See Leistner, *supra* note 60 at 48 (in the pre-print): “rightsholders will have to cross-check whether fair use is overwhelmingly likely”.

B. Approaches in the EU

Introducing Art 17 of the DSM-Dir, the European legislature was aware that algorithm-based filtering tools can result in legitimate uploads being blocked. In the European legal framework, several approaches can therefore be identified that aim to mitigate the risk of overblocking. They can be classified into different categories: legal restrictions on filtering, procedural safeguards, transparency obligations and external audits. The following sections analyse the safeguards provided in the EU, compare and contrast them with the safeguards provided in the US and, finally, discuss whether the approaches in the EU can effectively address concerns that were raised around algorithmic copyright enforcement.

1. Legal restrictions on filtering

Platform providers' discretion regarding filtering users' uploads is limited by several legal provisions.

(a) *Restrictions in the European legal framework*: Art 17(7) and Art 17(9) of the DSM-Dir set limits to filtering by platform providers that qualify as OCSSPs. The provisions require that legitimate user uploads, including uploads covered by exceptions (*eg* as quotations or parodies), are not affected by platform providers' best efforts to filter according to Art 17 (4)(b) of the DSM-Dir. This prohibition of filtering non-infringing content is interpreted broadly in two respects.

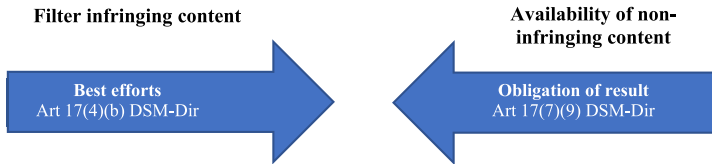
First, providing users with mechanisms to take *ex post* action against filtering (*eg* a complaint and redress mechanisms) is not sufficient. Rather, the filtering restriction must be considered *ex ante* when filtering algorithms are designed.⁶⁸ This is similar to the requirement that follows from *Lenz*, according to which rightsholders must consider the applicability of the fair use doctrine before they send takedown notices to platforms, but in the present context the obligation to consider the legitimacy of a use before it is restricted is that of platform providers. Requiring platforms to design filtering algorithms from the outset that they permit non-infringing content, including content that is privileged under exceptions (*eg* quotations, parodies), can be understood as a "balance by design" approach.⁶⁹ This approach, which makes it necessary to integrate legal requirements into the design of technical systems, is not novel, as it is already known from the European General Data Protection Regulation ("GDPR"), where it is reflected in the "privacy by design" requirement.⁷⁰

⁶⁸ AG opinion in the Case C-401/19 *Republic of Poland v. European Parliament and Council of the European Union* [2021] ECLI:EU:C:2021:613 at [170] et seq.

⁶⁹ This addresses demands put forward in the academic literature, see Maria Lillà Montagnani, "Virtues and Perils Of Algorithmic Enforcement and Content Regulation in The EU – A Toolkit For A Balanced Algorithmic Copyright Enforcement" (2020) 11 Case W. Res. J.L. Tech. & Internet 1 at 35 (<<https://scholarlycommons.law.case.edu/jolti/vol11/iss1/2>> [29 June 2023] ["Montagnani"]).

⁷⁰ General Data Protection Regulation (Regulation 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), European Parliament and Council Regulation 2016/679, Art 25, [2016] OJ L 119/1 ["GDPR"].

Second, platforms that qualify as OCSSPs must do more than demonstrate best efforts to comply with the prohibition of filtering non-infringing content: Art 17(7) and Art 17(9) of the DSM-Dir impose on OCSSPs an obligation to achieve the *result* that non-infringing content remains unaffected by their filtering tools.⁷¹ This creates a conflict, since OCSSPs must comply with two contradictory requirements. On the one hand they must demonstrate best efforts that infringing content is filtered according to Art 17(4)(b) of the DSM-Dir, and on the other hand they must achieve the result that non-infringing content is not filtered.



It is, of course, debatable whether the prohibition of overblocking contained in Art 17(7) and Art 17(9) of the DSM-Dir should be understood as an obligation of result, because a complete exclusion of filtering legitimate content (“false positives”) would only be given if no filters were applied at all.⁷² Certain impairments of copyright-permitted uses are therefore unavoidable and must consequently be accepted. A possible solution could take the form of a mitigated understanding, according to which the obligation to achieve the result is only violated if filtering tools do not sufficiently distinguish between copyright-infringing and copyright-permissible content.⁷³

(b) *Restrictions contained in national transpositions:* As Art 17 of the DSM-Dir requires transposition by EU member states, several of them took the opportunity to include further restrictions on filtering in their national transpositions. For example,⁷⁴ the Austrian and German legislatures introduced restrictions regarding the filtering of content that contains only “small snippets” of protected works.⁷⁵ Although these provisions differ from each other in detail, they both express the same assumption that the smaller the excerpts of other people’s works, the more likely it is that an exception (*eg* that for quotations or incidental inclusions) applies. Furthermore, another approach common to the Austrian and German legislatures is to restrict filtering of content that users have marked as permissible (“pre-flagging”).⁷⁶ Although details vary also in this respect, both national transpositions require OCSSPs to provide users with a tool to mark content as

⁷¹ CJEU Case C-401/19, *supra* note 33 at [78].

⁷² It is clear that this would contradict the intention of DSM-Dir, *supra* note 2 Art 17.

⁷³ See CJEU Case C-401/19, *supra* note 33 at [86].

⁷⁴ For a detailed analysis of national transpositions see Matthias Leistner, “The Implementation of Art. 17 DSM Directive in Germany – A Primer with Some Comparative Remarks” (2022) GRUR Int. 2022, 909.

⁷⁵ § 89b(3) of the Austrian Copyright Code; § 9(2) of the German Act on the Copyright Liability of Online Content-Sharing Service Providers.

⁷⁶ § 89b(4) of the Austrian Copyright Code; § 11 of the German Act on the Copyright Liability of Online Content-Sharing Service Providers.

copyright-compliant if they are of the opinion that it is privileged under an exception (eg as quotation, parody or incidental inclusion) or by an authorisation of rightsholders.

The main result of these additional filter restrictions provided at the national level is that uploads of small snippets or pre-flagged content must basically pass automated filtering tools.⁷⁷ Consequently, if this content becomes available, OCSSPs do not violate their best efforts and do not face primary liability even if these uses should indeed be infringing. It is instead the responsibility of rightsholders to file a substantiated notice according to Art 17(4)(c) of the DSM-Directive in which they eliminate the doubts regarding the permissibility.

(c) *Art 14(4) DSA*: The previously mentioned restrictions on filtering relate to the requirement that platform providers which qualify as OCSSPs must demonstrate best efforts to filter infringing content in order to avoid primary liability according to Art 17(4)(b) of the DSM-Directive. Therefore, they can be understood as a prohibition on permissible content being filtered on the grounds of (alleged) copyright infringement. What these restrictions do not affect is the discretion of platform providers to filter content for reasons other than copyright, for instance, content that violates their terms of service.⁷⁸ However, this freedom is now likely to be restricted by Art 14 of the Digital Services Act (“DSA”), which obliges providers of online platforms⁷⁹, when they impose restrictions on information provided by users for the purpose of content moderation⁸⁰ (including algorithmic decision-making), to act in a diligent, objective and proportionate manner with due regard to rights and legitimate interests of all parties involved, including users’ fundamental rights (eg freedom of expression).⁸¹ Furthermore, very large online platforms⁸² are required to diligently identify, analyse and assess any systemic risks – for example, to the fundamental right of freedom of expression – that arise from algorithmic filtering systems and to mitigate these risks by adapting their algorithms.⁸³ Although the restrictions on content management defined in the DSA contain only vague criteria, it can be assumed that they will tighten the standards for permissible filtering.

⁷⁷ According to the Austrian transposition, small snippets can nevertheless be filtered in exceptional cases if their availability would significantly impair economic exploitation of the work; pre-flagged content must be filtered if its flagging as permissible is immediately recognisable as being abusive.

⁷⁸ For the German transposition see Louisa Specht-Riemenschneider, „§ 9 UrhDaG“ in Thomas Dreier & Gernot Schulze, eds, *Urheberrechtsgesetz* (Munich: Beck 2022) [6] (in German language).

⁷⁹ According to DSA, *supra* note 12 Art 3(i), this refers to hosting services that, at the request of a recipient of the service, store information and disseminate it to the public, unless that activity is a minor and purely ancillary feature of another service or a minor functionality of the principal service and, for objective and technical reasons, cannot be used without that other service, and the integration of the feature or functionality into the other service is not a means of circumventing the applicability of the DSA.

⁸⁰ Content moderation is defined in a broad sense that does not only refer to (automated) filtering, but also to other restrictions, such as demonetisation and termination or suspension of a user’s account (DSA, *supra* note 12 Art 3(t)).

⁸¹ DSA, *supra* note 12 Art 14(1) and (4).

⁸² According to DSA, *supra* note 12 Art 33(1), this refers to online platforms which have an average number of monthly active recipients of the service in the EU equal to or higher than 45 million, and which are designated as very large online platforms.

⁸³ DSA, *supra* note 12 Art 34(1)(b), (2)(a)(b) and 35(1)(c).

2. Procedural safeguards

The European framework also provides procedural safeguards that mitigate the risk of automated filtering tools blocking content that is permissible under copyright law.

(a) *Internal complaint and redress mechanisms*: Art 17(9) of the DSM-Dir requires OCSSPs to provide users with effective and expeditious complaint and redress mechanisms in the event of disputes over filtering. This is reminiscent of the counter-notice procedure provided by US copyright law,⁸⁴ but Art 17(9) of the DSM-Dir does not define the handling process in as much detail. It only states that complaints submitted by users shall be processed without undue delay and that, when decisions are made in favour of (maintaining) filtering, they must be subject to human review. Art 17(9) of the DSM-Dir is, however, not the only relevant provision in the European framework. Art 20 of the DSA extends the obligation to provide internal complaint-handling procedures, as all platforms⁸⁵ (not only OCSSPs) are required to provide them and as users are entitled to file complaints not only when their content is filtered, but also when their access to the platform is restricted or terminated.⁸⁶ Furthermore, the provision adds specifications to the complaint-handling procedure:⁸⁷ it must be easily accessible, user-friendly and free of charge; it must be possible for users to file sufficiently precise and adequately substantiated complaints, not only in cases in which content is filtered because of (alleged) copyright infringement, but also when it is filtered because of violations of a platform's terms and conditions.⁸⁸ Contrary to the situation in the US, the duty of the platform provider in this procedure is not defined as one of mediation between the parties, but rather as one of decision-making.⁸⁹ Specifically, a platform provider must act in a timely, non-discriminatory, diligent and non-arbitrary manner, and must ultimately reverse the filtering decision without undue delay if a complaint contains sufficient grounds for the provider to consider that the use is permissible.⁹⁰ Importantly, both Art 17(9) of the DSM-Dir and Art 20(6) of the DSA require such decisions regarding complaints to be subject to human review.⁹¹ Therefore, at this point at the latest, fully automated filtering procedures must be supplemented with human decisions that evaluate not only the content itself, but also the reasons that users have put forward in their complaints.

⁸⁴ 17 U.S.C. §512(g)(2).

⁸⁵ See footnote 79. Platforms that qualify as small and medium-sized enterprises are excluded (DSA, *supra* note 12 Art 19).

⁸⁶ DSA, *supra* note 12 Art 20(1).

⁸⁷ Note that the relation between the DSA and the DSM-Dir is not entirely clear. Art 2(4)(b) of the DSA states that the DSA “*is without prejudice to*” European copyright law. However, it can be assumed that requirements provided in the DSA are to be understood as concretisations and supplements to the requirements of Art 17 DSM. See João Pedro Quintais & Sebastian Felix Schwemer, “The Interplay between the Digital Services Act and Sector Regulation: How Special Is Copyright?” (2022) 13 European Journal of Risk Regulation 191 at 213.

⁸⁸ DSA, *supra* note 12 Art 20(1) and (3).

⁸⁹ See Ruth Janal, „Friendly Fire? Das Urheberrechts-Diensteanbieter-Gesetz und sein Verhältnis zum künftigen Digital Services Act“ (2022) GRUR 211 at 219 (in German language).

⁹⁰ DSA, *supra* note 12 Art 20(4).

⁹¹ DSM-Dir, *supra* note 2 Art 17(9) only requires human review if the challenged decision to filter content is maintained.

(b) *Out-of-court settlement procedures*: Art 17(9) of the DSM-Dir also requires that users of OCSSPs can avail themselves of out-of-court settlement procedures for the impartial settlement of filtering disputes. This entitlement to challenge filtering decisions not only within internal complaint and redress procedures but also before an external authority is taken up by Art 21 of the DSA, which provides further specifications of the settlement procedure (eg regarding accessibility⁹² and the bearing of the costs⁹³) and the settlement body, but also clarifies that the settlement body shall not have the power to impose a binding settlement of the dispute on the parties. The non-binding nature of these settlements, of course, weakens the effectiveness of the settlement procedures. But as these procedures must not prejudice the rights of users to have recourse to efficient judicial remedies,⁹⁴ it becomes evident that they offer users an additional forum to challenge filtering decisions and do not prevent access to the courts of law.

3. *Transparency obligations*

A prominent approach to addressing platform-related challenges that is currently pursued in the EU is to foster transparency. Several information obligations demonstrate that transparency is also sought with regard to automated filtering mechanisms.⁹⁵ Specifically, the information that must be provided can be classified as follows.

(a) *Permissibility of uses*: In their published terms and conditions, platform providers which qualify as OCSSPs must inform their users that they are allowed to use protected works in accordance with applicable exceptions (eg the quotation and parody exceptions).⁹⁶ This obligation aims to increase users' awareness that uses such as quotations and parodies are permissible under copyright law and that they can demand redress if filtering algorithms wrongfully qualify permissible uses as copyright infringements.

(b) *Functioning of the filtering mechanisms*: Furthermore, OCSSPs must provide adequate information on the functioning of their filtering mechanisms.⁹⁷ This information is important because filtering algorithms are opaque, and without disclosure of the filtering parameters it is impossible for users to understand why their content is filtered. It is remarkable that the DSM-Dir requires this transparency only vis-à-vis rightsholders, whereas individual users – who are, after all, the main beneficiaries – need not necessarily be provided with this information. The

⁹² DSA, *supra* note 12 Art 21(1): easily accessible within the online interface, clear and user-friendly.

⁹³ In a nutshell, if the dispute is decided in favour of users, platform providers must bear all fees charged by the out-of-court dispute settlement body and reimburse users for any other reasonable expenses incurred in relation to the dispute settlement. If the dispute is settled in favour of the platform, users are not required to reimburse the platform providers for any fees or other expenses unless the users manifestly acted in bad faith (Art 21(5) DSA).

⁹⁴ DSM-Dir, *supra* note 2 Art 17(9); DSA, *supra* note 12 Art 21(1).

⁹⁵ DSM-Dir, *supra* note 2 Art 17(8), (9) and (10).

⁹⁶ DSM-Dir, *supra* note 2 Art 17(9).

⁹⁷ DSM-Dir, *supra* note 2 Art 17(8) and (10).

directive requires only that „users’ organisations“ (ie institutions, which are not further specified) can access this information for the purpose of stakeholder dialogues. However, these shortcomings of the DSM-Dir have been (at least partly) eradicated. The Austrian legislature went beyond the requirements of Art 17 of the DSM-Dir and obliges OCSSPs to provide information on the functioning of their filtering tools also to individual users.⁹⁸ At the European level, Art 14(1) of the DSA now requires all platforms⁹⁹ to have publicly available terms and conditions that include (clear, plain, intelligible, user-friendly and unambiguous) information on any restrictions they impose for the purpose of content moderation. Furthermore, providers of very large online platforms¹⁰⁰ must explain their algorithmic systems vis-à-vis the Digital Service Coordinator or the European Commission.¹⁰¹

(c) *Reasons for filtering*: Platform providers¹⁰² are further required to give users, after their content has been filtered,¹⁰³ a clear and specific statement of reasons. According to Art 17 of the DSA, they must inform users, *inter alia*,

- about the facts and circumstances¹⁰⁴ of the filtering decision and, if strictly necessary, the identity of the person who has notified the (alleged) infringement;¹⁰⁵
- whether the decision to filter was taken by automated means;¹⁰⁶
- about the legal grounds relied on if content was filtered because of alleged infringement and explanations as to why it was considered to be infringing.¹⁰⁷

(d) *Transparency reports*: Finally, according to Art 15 of the DSA, platforms¹⁰⁸ must publish comprehensive transparency reports on any content moderation annually and in a machine-readable format. These reports must provide, *inter alia*, transparency regarding the frequency of and reasons for content filtering. Furthermore, they must provide a qualitative description of the filtering measures, inform about the precise purposes of filtering, and indicate the accuracy and possible rate of error of filtering measures and the safeguards applied. For providers of very large online platforms,¹⁰⁹ Art 42 of the DSA further strengthens this transparency obligation.

⁹⁸ § 89b(2) of the Austrian Copyright Code.

⁹⁹ See footnote 79.

¹⁰⁰ According to DSA, *supra* note 12 Art 33(1). See footnote 82.

¹⁰¹ DSA, *supra* note 12 Art 40(3).

¹⁰² See footnote 79.

¹⁰³ This also applies to cases in which user accounts are suspended or terminated (DSA, *supra* note 12 Art 17(1)(d)). Furthermore, a statement of reasons is required when users file a complaint against a decision to filter under the internal complaint and redress mechanism and the platform provider maintains the decision filter (DSA, *supra* note 12 Art 20(5)).

¹⁰⁴ This includes information on whether the decision was taken pursuant to a notice or based on investigations initiated voluntarily by platforms.

¹⁰⁵ DSA, *supra* note 12 Art 17(3)(b).

¹⁰⁶ DSA, *supra* note 12 Art 17(3)(c).

¹⁰⁷ DSA, *supra* note 12 Art 17(3)(d).

¹⁰⁸ See footnote 79.

¹⁰⁹ See footnote 82.

4. External audits

According to Art 37 of the DSA, very large online platforms¹¹⁰ are subject to independent audits carried out by an external organisation at least once a year. This audit evaluates, *inter alia*, whether a platform is complying with its duties regarding content management, specifically if they act in a diligent, objective and proportionate manner with due regard to rights and legitimate interests of all parties involved and mitigate systematic risks related to their algorithmic systems.¹¹¹

C. Discussion

The development towards algorithmic enforcement of copyright has given rise to numerous concerns. In the legislative process, users feared that Art 17 of the DSM-Dir would lead to widespread filtering and ultimately to censorship. The concerns expressed in the academic field were related, in essence, to the increased involvement of platforms as private actors in the process of copyright enforcement, the inability of algorithmic filtering tools to sufficiently distinguish between infringing and non-infringing uses, the lack of accountability, transparency and contestability, and to users being deprived of their right to a fair judicial process.¹¹²

1. The involvement of platforms in the process of copyright enforcement

A main point of criticism is that in the process of algorithmic enforcement, law enforcement tasks are exercised by platforms as private actors, which leads to a privatisation of a public function, and that all enforcement functions (*ie* detection, prosecution, adjudication and execution) are concentrated in the hands of platform providers.¹¹³ Indeed, when platforms comply with takedown request submitted by rightsholders under the NTD system of the US DMCA or when content is blocked by upload filters that are used to comply with Art 17 of the DSM-Dir, no court has made a decision on whether restricted content was infringing copyright. However, it should be acknowledged that neither system denies users the ability to take legal action against platform providers if they think that permissible uses (*eg* quotations, parodies) have been wrongfully restricted.¹¹⁴ From a structural point of view, platforms and filtering algorithms do not replace, but precede human decision-making by judges. This admittedly reduces the probability of courts deciding on the legitimacy of user uploads, as many users will not take legal action against restrictions.

¹¹⁰ See footnote 82.

¹¹¹ DSA, *supra* note 12 Art 14 and 35.

¹¹² Perel & Elkin-Koren, *supra* note 4 at 478 et seq; Fischman-Afori, *supra* note 11 at 373; Montagnani, *supra* note 69 at 24 et seq; Frosio, *supra* note 16.

¹¹³ Perel & Elkin-Koren, *supra* note 4 at 481; Montagnani, *supra* note 69 at 24; Frosio, *supra* note 16.

¹¹⁴ DSM-Dir, *supra* note 2 Art 17(9). In the EU, it is actually only in the course of the shift towards algorithmic enforcement that the discussion on whether copyright exceptions can be understood as conferring “rights” on users has gained momentum, see *eg* CJEU Case C-401/19, *supra* note 33 at [87].

These effects, however, are mitigated by legal provisions, such as Art 17(7) and Art 17(9) of the DSM-Dir, that prohibit the filtering of permissible uses, especially when filtering of small snippets and content that users have marked as permissible (“pre-flagging”) is restricted, as is the case under the Austrian and German transposition of Art 17 of the DSM-Dir. Furthermore, recent case law demonstrates that the inclusion of platforms as non-judicial actors in the enforcement process is in line with the European legal framework. The CJEU has recently decided that, when a person requests a search-engine operator to delete inaccurate search result, it is sufficient for the applicant to submit relevant and sufficient evidence that establishes the manifest inaccuracy of the search result. In this case, de-referencing is not subject to the condition that the question of the accuracy of the search result has been resolved in a judicial decision.¹¹⁵ Therefore, it must be acknowledged that, also in the field of data protection, an operator of a search engine is burdened with the evaluation of the accuracy of search results before a court has ruled on it. If this has to be accepted in the area of data protection, it is not apparent why it should be unacceptable in the area of copyright law. With regard to the US, the fact that the US copyright office stresses that the NTD regime gives rightsholders an “*extra-judicial method for addressing [copyright] infringement*” indicates that the involvement of platform providers as non-judicial actors in the enforcement process can also be reconciled with the US legal system.¹¹⁶

Apart from the legal permissibility of involving platforms in the enforcement process, it should be considered that platforms also engage in content moderation for other reasons. Platform providers suggest content on starting pages, rank it in lists of search results or indexes, and remove content that either violates their terms of service or appears less profitable. In other words, there is no unmoderated content in a platform environment, and the question is only whether (alleged) copyright infringement is a factor in moderation. Finally, the question remains whether the nature of platform providers as profit-oriented entities should solely be seen as a factor that gives rise to fears of fostering over-enforcement.¹¹⁷ Admittedly, it must be assumed that profit-oriented businesses try to avoid liability risks and, therefore, are incentivised to restrict content that raises doubts with regard to copyright infringement. These incentives are amplified when platforms are exposed to primary liability which provides significant financial risks. At the same time, however, it is evident that platform providers – and rightsholders¹¹⁸ – also have incentives not to filter content, as its availability generates revenues. It is therefore possible that the profit orientation of platforms also provides certain incentives for content not to be filtered. Such “stay-up” incentives have been ignored so far, but should be taken into consideration in order to take full account of the realities.

¹¹⁵ CJEU Case C-460/20 *TU and RE v. Google* [2022] ECLI:EU:C:2022:962 at [68] et seq.

¹¹⁶ United States Copyright Office, Section 512 of Title 17, *supra* note 8 at 1.

¹¹⁷ The profit orientation of platform providers is usually brought forward in a critical context, see Perel & Elkin-Koren, *supra* note 4 at 482; Fischman-Afori, *supra* note 11 at 373.

¹¹⁸ From the perspective of rightsholders, the main goal of DSM-Dir Art 17 was not to prevent their works from being used on platforms, but to improve financial participation.

2. Changes in the principles of the enforcement process

Academia has criticised that using algorithms to detect and block infringing content changes the “default” of copyright policy, as content detected by algorithms remains unavailable unless explicitly authorised by the copyright owner, whereas otherwise content is publicly available unless proven infringing.¹¹⁹ In this respect, the European system goes even further than the NTD system of the US. Firstly, as platforms must use upload filters to comply with Art 17(4)(b) of the DSM-Dir,¹²⁰ rightsholders need not initiate the enforcement process in every single case by submitting a takedown request. Rather, a single initiative (*ie* submitting a reference file of a work) is sufficient to enforce copyright regarding all infringing content that is presently available or is uploaded to the platform by any user in the future.¹²¹ This can be described as an extension of enforcement from “one to one” to “one to many”. Secondly, from the moment a reference file is submitted to a platform, the task of detecting infringements is no longer one of rightsholders, but shifts to one of platform providers.¹²²

In my opinion, however, taking these effects solely as arguments against algorithmic enforcement is not entirely convincing. It is hard to envisage how a system could work in which decisions to restrict content were made exclusively by human judges and it is the sole responsibility of platforms to enforce these decisions. In such a system, rightsholders would have to initiate court proceedings every time an infringing upload of their works became available on a platform – which comes close to a Sisyphean task, especially considering that infringing content can easily be reuploaded after it has been taken down. Therefore, efficiency and, from a European perspective, also the fundamental right to property¹²³ and an effective remedy¹²⁴ favour algorithmic enforcement and the proactive involvement of platforms because they are often the sole institutions that can effectively provide redress against copyright infringements.

With regard to the permissibility of algorithms from the perspective of US law, the Ninth Circuit’s original decision in *Lenz* also demonstrated a positive approach towards algorithmisation, as it referred to computer algorithms appearing “to be a valid and good faith middle ground for processing a plethora of content while still meeting the DMCA’s requirements to somehow consider fair use.”¹²⁵ However, this wording is not present in the amended opinion.¹²⁶ Even when a critical view of the algorithmisation of copyright enforcement is maintained, the question remains of

¹¹⁹ Perel & Elkin-Koren, *supra* note 4 at 492.

¹²⁰ The same holds true if platforms are subject to staydown obligations resulting from injunctions.

¹²¹ However, the fact that platforms do not have to demonstrate best efforts to filter infringing content before rightsholders have submitted reference files demonstrates that under DSM-Dir Art 17, the principle of initiation of enforcement by rightsholders is not abandoned entirely. This was one of the reasons why the CJEU concluded that DSM-Dir Art 17 does not unduly restrict the fundamental right to freedom of expression (Case C-401/19, *supra* note 33 at [89]).

¹²² This also holds true when platforms are subject to staydown obligations resulting from injunctions.

¹²³ CFREU, *supra* note 50 Art 17(2).

¹²⁴ CFREU, *supra* note 50 Art 47.

¹²⁵ *Lenz v. Universal Music Corp.* 2015, *supra* note 58 at 1135.

¹²⁶ *Lenz v. Universal Music Corp.* 815 F.3d 1145 (9th Cir. 2016).

whether a system that restricts platform providers' obligations to comply with take-down requests actually leads to a fair allocation of revenues between platforms and rightsholders, particularly if this must be done manually. After all, as any initiative to remove infringing content must come from rightsholders, all revenues remain with the platform if rightsholders do not detect infringing content. Furthermore, even if rightsholders detect infringing uploads and file takedown notices, any revenues generated by the platform provider in the period between upload and takedown remain entirely with the platform provider.¹²⁷

3. *Inability of filtering algorithms to distinguish between copyright-infringing and non-infringing content*

Major concerns have been raised that automated filtering systems result in unjustified overblocking because of the inability of algorithms to make context-based decisions and to accurately identify legitimate uses, such as quotations and parodies.¹²⁸ These concerns cannot be dismissed because distinguishing between copyright-infringing and non-infringing uses does require a case-by-case analysis that considers the context. This does not only apply to copyright regimes with fair use or fair dealing provisions, but also to European copyright law, because the statutory exceptions for quotations and parodies also depend on the specific circumstances.¹²⁹ For example, a quotation must have a valid quotation purpose, such as criticism or review, and its use must be in accordance with fair practice and required by the specific purpose;¹³⁰ a parody must constitute an expression of humour or mockery.¹³¹ It is evident that these context-related factors (criticism, review, fair practice, humour, mockery) cannot be integrated into algorithms, at least not according to the current state of technology.¹³² Maybe some individual factors such as the shortness of an excerpt, additions and alterations by users can be construed that can be processed by filtering algorithms and serve as indications of permissible uses such as quotations and parodies. It has to be admitted, though, that none of these factors (alone) is a robust standard for assessing the permissibility of a use.

Despite these technical deficiencies, the question arises whether algorithms inevitably lead to overly strict filtering. It seems more appropriate to also take into account the legal framework that determines the design of the filters and the risks

¹²⁷ For an in-depth analysis see Philipp Homar, "Wertschöpfung vs. Partizipation: Die Vergütung der Nutzung auf User-Upload-Plattformen" in Lena Maute & Mark-Oliver Mackenrodt, eds, *Recht als Infrastruktur für Innovation* (Baden-Baden: Nomos, 2019) 225 at 233 et seq. (in German language).

¹²⁸ Montagnani, *supra* note 69 at 21; Geiger & Jütte, *supra* note 29 at 533; Pamela Samuelson, "Pushing Back on Stricter Copyright ISP Liability Rules" (2021) 27 Mich. Tech. L. Rev. 299 at 317. Samuelson, Testimony to the Senate Judiciary Committee, Subcommittee on Intellectual Property, of March 10, 2020, pp. 7 et seq.

¹²⁹ Montagnani, *supra* note 69 at 21.

¹³⁰ InfoSoc-Directive, *supra* note 10 Art 5(3)(d).

¹³¹ CJEU Case C-201/13 *Deckmyn and Vrijheidsfonds v. Vandersteen et al* [2014] ECLI:EU:C:2014:2132 at [33]. In this decision, the CJEU also held that a parody "must be regarded as an autonomous concept of EU law and interpreted uniformly throughout the European Union", *ibid* at [15] et seq.

¹³² Whether these shortcomings can be overcome by future developments in the field of machine learning remains to be seen.

that platforms face if their filters are too strict. Regarding the situation in the US, the liability of a platform for unjustified takedowns does not seem to be excessive, as 17 U.S.C. § 512(g)(1) imposes no liability if platform providers remove content in good faith.¹³³ This suggests that fears of an overly strict approach are not entirely unfounded in the US. However, a different picture emerges in the EU, as platform providers' discretion to restrict user content is increasingly limited. In particular Art 17(7) and Art 17(9) of the DSM-Directive counteract excessively designed filtering algorithms, as any design that leads to copyright-compliant content (*eg* quotations, parodies) being filtered because of alleged copyright infringement ("false positives") would in fact violate the obligation of result that follows from these provisions. This standard is stricter than its counterpart in the US, as good-faith belief that content is infringing is not sufficient. Nevertheless, there are two caveats. The first is that the European framework leaves it unclear what the consequences are if the prohibition of filtering permissible content as provided in Art 17(7) and Art 17(9) of the DSM-Directive is violated.¹³⁴ However, national transpositions of Art 17 of the DSM-Directive do contain sanctions. According to the Austrian transposition, for example, fines of up to EUR 1 million can be imposed by a supervisory authority if platforms filter non-infringing content systematically and to a significant extent. In Germany, registered associations may seek injunctive relief against OCSSPs that filter non-infringing content. Therefore, the prohibition that upload filters restrict permissible content is not sanctionless, but it remains to be seen, of course, whether these national approaches will be effective. The second caveat is that legal restrictions to filtering leave it unclear how filtering algorithms can (and must) be designed to be sufficiently balanced. Pending further clarification by the European legislature or the CJEU, however, it could be workable to design filtering algorithms in such a way that they distinguish between manifestly infringing content and ambiguous content, which refers to content for which it is unclear whether it is infringing.¹³⁵ Filtering manifestly infringing content (*eg* user uploads that contain entire films, videos or musical compositions of third parties without any changes or additions created by the uploading users) would be permitted because of Art 17(4)(b) DSM-Directive and also required to demonstrate best efforts. Although it cannot be excluded that content of this category may be permissible as quotation or parody, it is relatively unlikely that the relevant exceptions¹³⁶ apply or that users have obtained an authorisation. To the contrary, in order to meet the obligation to achieve the result that permissible uses are not interfered with (Art 17(7) and Art 17(9) of the DSM-Directive), it seems appropriate that filtering algorithms are designed such that ambiguous content is generally allowed to pass. The category of ambiguous content should predominantly include transformative uses, in which third-party content is substantially adapted or is contained only in part (*eg* remixes, cover versions, memes or reviews),

¹³³ Perel & Elkin-Koren, *supra* note 4 at 501.

¹³⁴ According to DSA, *supra* note 12 Art 52(2), violations of DSA Art 14(4), *ie* when platforms do not act in a diligent, objective and proportionate manner and violate users' rights to freedom of expression, are sanctioned with fines of up to 6% of the annual worldwide turnover of a platform.

¹³⁵ See AG opinion in the case C-401/19, *supra* note 68 at [202].

¹³⁶ This is supported, for quotations, by the three-steps test of InfoSoc-Directive, *supra* note 10 Art 5(5) (CJEU Case C-516/17 *Spiegel Online v. Volker Beck* [2019] ECLI:EU:C:2019:625 at [67]) and, for parodies, by the requirement that they have to be *noticeably different* from the original work (CJEU Case C-201/13, *supra* note 131 at [33]).

as it cannot be ruled out that this type of content falls under an exception (*eg* as parody or quotation). How this differentiation can be implemented technically must, of course, be examined in more detail. A promising initial approach may be for filtering algorithms to differentiate according to the rate of match detected between an upload and the reference files provided by rightsholders. When filters are activated only in case of high match rates, a broad range of transformative content is likely to remain unaffected by filtering measures. This range becomes even wider when national transpositions, such as those in Austria and Germany, provide additional filtering restrictions for small snippets and content that has been flagged by users as permissible. These broad limitations undoubtedly weaken the often-stated robustness of filtering algorithms. Despite these considerations it is, of course, still possible that filters provide inaccurate results, as content with a high match rate may, in individual cases, be privileged as quotation or parody or covered by an authorisation. However, a differentiation according to the match rate may be a starting point to go in the right direction. What remains to be discussed in this context is whether it is desirable to legally determine match rates (or other decision factors) according to which filtering systems must be triggered or whether certifications for filtering systems are a desirable solution. For platforms, this could have the advantage of increasing legal certainty, as such solutions could include an exclusion of liability to rightsholders (for false negatives) and users (for false positives) if filters meet the specified requirements. Further approaches to aligning the interests of rightsholders and those of users may be to combine algorithmic procedures with human intervention,¹³⁷ for instance, in the form of human reviews in particular ambiguous cases and to fine-tune the complaint and redress procedure.

4. Accountability, transparency and contestability

Considerable criticism focuses on platform providers' lack of accountability when they automatically filter permissible content and, furthermore, the lack of contestability, especially because filtering algorithms are not transparent, which keeps users from understanding and contesting filtering decisions.¹³⁸ Against this background, Perel & Elkin-Koren argued that filing complaints against filtering decisions and reverse engineering the code of filtering algorithms should be encouraged, that information about the functioning of filtering mechanisms (*eg* criteria according to which algorithms determine infringing uploads, such as quantitative thresholds) should be published, that private "watchdogs" should monitor content restrictions, and that platforms should be subject to external regulatory inspection.¹³⁹ Similarly, Montagnani advocated for the disclosure of information on the functioning of filtering algorithms, a right to explanation when content is filtered, a right to demand that filtering algorithms be audited, and an obligation for platforms to perform an

¹³⁷ See *eg* Leistner, *supra* note 60 at 50, who suggested a coupling of algorithmic tools with inputs from human agents.

¹³⁸ See Perel & Elkin-Koren, *supra* note 4 at 513 et seq.; Montagnani, *supra* note 69 at 25 et seq; Fischman-Afori, *supra* note 11 at 373; Frosio, *supra* note 16.

¹³⁹ Perel & Elkin-Koren, *supra* note 4 at 525 et seq.

impact assessment when they develop and deploy their filtering algorithms.¹⁴⁰ More fundamentally, Fischman-Afori demanded that public law standards be imposed on platforms; the consequence of these standards would be that platforms (i) are subject to obligations to set clear and transparent rules of operation, (ii) must secure equality between all users, (iii) must give reasoned decisions, and that (iv) users can file an appeal against decisions to an objective (quasi-judicial) body.¹⁴¹ With regard to the NTD system in the US, there has been widespread criticism that these demands for accountability, transparency and contestability are insufficiently met.¹⁴² However, as described above,¹⁴³ in the European framework several approaches can be identified that aim to enhance accountability, transparency and contestability and that go beyond the standard offered by the NTD regime in the US.

As far as *transparency* is concerned, it is particularly noteworthy that European law requires platforms to disclose information on the functioning of filtering mechanisms to rightsholders¹⁴⁴ and users.¹⁴⁵ Of course, it can be questioned whether the goal to inform users will be achieved when platforms are only required to provide information in their terms and conditions.¹⁴⁶ It would certainly increase awareness if, although not required by European law, the information were also provided together with notifications sent to users when their content is filtered. Demands that have not (yet) been met are those for rights to audit and to reverse-engineer the code of filtering algorithms.¹⁴⁷ This gap, however, seems to have been closed for very large online platforms, as the DSA provides that external auditors, the Digital Services Coordinator and the European Commission can request access to data necessary to monitor and assess compliance with the provisions on content moderation.¹⁴⁸ As the code of the algorithms undoubtedly qualifies as “data”, it is possible that these provisions may result in a right for institutional actors to inspect the code of filtering algorithms, at least if platforms fail to sufficiently explain their functioning by other means. Another important aspect in terms of transparency is that European law requires platforms to provide users with reasons why their content was filtered.¹⁴⁹ This is decisive because, depending on the reason, a bifurcated regime of filtering restrictions applies. When OCSSPs filter content on grounds of (alleged) copyright infringement, any false positives are actually in conflict with the obligation to achieve the result that permissible content is not interfered with as provided in Art 17(7) and Art 17(9) of the DSM-Dir. Only filtering carried out for other reasons than copyright infringement (*eg* violations of the terms of service) is subject to the less strict content moderation standards of Art 14(4) of the DSA.¹⁵⁰

¹⁴⁰ Montagnani, *supra* note 69 at 33 et seq.

¹⁴¹ Fischman-Afori, *supra* note 11 at 395 et seq.

¹⁴² See *eg* Perel & Elkin-Koren, *supra* note 4 at 497 et seq.

¹⁴³ See chapter IV.B.

¹⁴⁴ DSM-Dir, *supra* note 2 Art 17(8).

¹⁴⁵ DSA, *supra* note 12 Art 14(1).

¹⁴⁶ DSA, *supra* note 12 Art 14(1).

¹⁴⁷ See Montagnani, *supra* note 69 at 41; Perel & Elkin-Koren, *supra* note 4 at 526.

¹⁴⁸ DSA, *supra* note 12 Art 37(2), Art 40(1).

¹⁴⁹ DSA, *supra* note 12 Art 17.

¹⁵⁰ In my opinion, this also applies to the filtering of content that is protected by copyright, but see João Pedro Quintais, Giovanni De Gregorio & João C. Magalhães, “How platforms govern users’ copyright-protected content: Exploring the power of private ordering and its implications” (2023) 48

Concerning *contestability*, both the European and the American system provide an *ex post* procedure that enables users to contest content restrictions. Within the US framework, if a rightsholder fails to initiate a lawsuit within a certain period of time, the complaint procedure will be decided in favour of the user.¹⁵¹ This feature of the US framework is currently not provided in the EU and could be considered in the future. However, two aspects are likely to strengthen contestability in the EU when compared to the situation in the US. First, in addition to complaint and redress procedures, which are handled by the platforms, the European framework stipulates out-of-court settlement procedures. Users are thus provided with an additional extra-judicial forum to challenge filtering decisions before an external and independent authority. Second, there is a specific focus on providing users with reasons for content restrictions. How effective these settlement procedures and reasonings will be in practice remains, of course, to be seen. A key factor will be the level of detail at which platforms are required to provide reasons for content restrictions to users.¹⁵² To foster contestability, it would be advantageous if it were not sufficient for platforms to indicate that content was filtered because it was considered to be copyright-infringing, but if they also had to specify why the content was considered not to be privileged as quotation or parody. The legal framework provides an indication that platforms must provide such explanations.¹⁵³ Of course, the automated procedure makes it unreasonable to demand a substantive explanation in each individual case. However, if the aim to foster contestability is taken seriously, the goal should be that platforms also disclose information about why filtering algorithms have detected and blocked content. For example, when a video is filtered, information that would be both reasonable for a platform to provide and helpful to users could include: what specific copyright-protected work was detected in the video, the match rate, the exact parts of the video in which matches were detected and an explanation of why, according to the decision-making structure of the filtering algorithm, no permissibility was assumed. Should the reasons provided to users not be sufficient to guarantee contestability, it may be necessary to consider refinements of the legal requirements for reasonings in the future.

A key *accountability* argument is that platforms are not required to sufficiently justify their choices about what content they restrict and that they cannot be held accountable for failures and wrongdoings.¹⁵⁴ Although this criticism is valid and important, it should be noted that platforms¹⁵⁵ must be granted some freedom to restrict user content – even if it does not infringe copyright – in the case of violations

Computer L & Security Rev (available at: <<https://doi.org/10.1016/j.clsr.2023.105792>> [15 March 2023]). If platforms argue that content was restricted because of violations of their terms of service, DSA Art 17(3)(e) requires them to provide a reference to the specific contractual ground that was violated and to provide an explanation as to why the content was considered to be incompatible with that ground.

¹⁵¹ 17 U.S.C. § 512(g)(2)(C).

¹⁵² The effectiveness of the counter-notice procedure in the US is doubtful because few users actually file counter-notices and platforms often do not provide sufficient information that allows users to understand the reason why their content was restricted and to respond substantively, see Perel & Elkin-Koren, *supra* note 4 at 507 et seq.

¹⁵³ DSA, *supra* note 12 Art 17(3)(d).

¹⁵⁴ See Perel & Elkin-Koren, *supra* note 4 at 481; Montagnani, *supra* note 69 at 25.

¹⁵⁵ Except otherwise required by antitrust law or special must-carry obligations.

of their terms of service or their “content policy” without facing sanctions.¹⁵⁶ For example, it must be possible for a platform such as TikTok to reject videos that exceed the maximum duration allowed by the platform, even if the video qualifies as a parody or quotation. However, the European framework arguably narrows platform providers’ discretion regarding content policy and moderation and thus increases the scope within which platforms can indeed be held responsible for content (mis-)moderation. The main indications of this are filtering and content moderation restrictions as provided in Art 17(7) and Art 17(9) of the DSM-Directive and Art 14(4) of the DSA and the related financial penalties that can be imposed when these restrictions are violated. Especially the requirement resulting from Art 14(4) of the DSA, according to which any content moderation must be carried out in a diligent, objective and proportionate manner with due regard to fundamental rights, is nothing less than a step towards imposing public law standards and fundamental rights on platforms as private actors.¹⁵⁷

V. CONCLUSION

In this article I have argued that the phenomenon of algorithmic enforcement of copyright law is present in both the US and the EU. There are, however, fundamental differences between the legal frameworks, which I compared and contrasted both regarding the requirement to use algorithms to detect and combat copyright infringements and regarding the protection of user interests. In comparison, the European copyright framework creates much stronger incentives for platforms to use automated filtering tools. These incentives follow from platforms being subject to staydown obligations resulting from injunctions, and they are amplified when platforms are subject to primary liability as is the case under Art 17 of the DSM-Directive. Against this background, the European copyright framework serves as a catalyst for algorithmic enforcement of copyright.

The numerous concerns around algorithmic enforcement raised by stakeholders and academics must, of course, be taken seriously. However, despite the shortcomings of algorithms in detecting infringing uses and distinguishing them from permissible uses, it is far from certain that broad-scale overblocking is an inevitable consequence of the developments in the EU. Since the DSM-Directive and national transpositions have been in force for some time, it must be acknowledged that Art 17 of the DSM-Directive has – at least so far – not turned out to be the feared “censorship machine”. Furthermore, several approaches can be identified in the EU, especially provided by the DSA, that tackle remaining concerns. In general, these safeguards seem more comprehensive in the EU than in the US. Of course, whether these safeguards will be effective in practice and succeed in making automated filtering decisions sufficiently explainable and contestable remains to be seen, but it is undeniable that the newly introduced obligations of platforms to provide information and reasons to users demonstrate an awareness of the problem and provide

¹⁵⁶ This falls under the freedom to conduct a business according to CFREU, *supra* note 50 Art 16.

¹⁵⁷ This addresses demands that were raised in the academic literature, see Fischman-Afari, *supra* note 11 at 394 et seq.

a foundation to work with. I therefore conclude that the shift towards algorithmic enforcement within the EU is not a one-sided, but rather a nuanced development. Computer scientists, representatives of stakeholders (users, platforms) and legal scholars should take it as an opportunity to investigate how algorithms can be designed to meet the legal requirements, how they can be implemented such that they balance the interests of rightsholders, users and platform providers, and what legal provisions are required to govern this process.